volume 29, number 2

summer–fall 2021

journal of slavic linguistics

Journal of the Slavic Linguistics Society

Slavic Linguistics Society

The Slavic Linguistics Society is an organization devoted to the systematic and scholarly study of the Slavic languages. Membership in the Society is open to all linguists regardless of field, theory, framework, or school. Members are entitled to present papers at annual meetings and receive the *Journal of Slavic Linguistics*, which is the official journal of the Slavic Linguistics Society. Individuals receive a subscription to *JSL* with their annual membership in the Slavic Linguistics Society. To join, go to http://www.slaviclinguistics.org.

SLS is governed by an Executive Board consisting of six members elected at the annual meeting for three-year staggered terms. The Executive Board negotiates arrangements for the annual meetings with host institutions. Meetings alternate between North America and Europe.

Executive Board

Guidelines for Contributors

JSL is intended to address issues in the description and analysis of Slavic languages of general interest to linguists, regardless of theoretical orientation. Papers dealing with any aspect of synchronic or diachronic Slavic phonetics, phonology, morphology, syntax, semantics, or pragmatics will be considered, as long as they raise substantive problems of broad theoretical concern or propose significant descriptive generalizations. Comparative studies and formal analyses are especially encouraged. Submissions ranging from full-length articles to shorter remarks and replies will be considered, along with substantive reviews.

Manuscripts should be submitted through the journal's submission website http://ojs.ung.si/index.php/JSL in .pdf format. All submissions should follow the submission guidelines published on the journal website. Papers should be anonymous and include a one-paragraph abstract. A computer file (preferably .doc[x]) will be requested before publication; we ask that you use only Unicode fonts.

Visit our World Wide Web page: http://ojs.ung.si/index.php/JSL

The *Journal of Slavic Linguistics* is published twice annually. Each volume consists of two issues, totaling approximately 360 pages: Winter–Spring and Summer–Fall. Library and institutional subscriptions are available from Slavica Publishers at the following rate:

Institutions $60.00

Individuals should join the Slavic Linguistic Society (http://www.slaviclinguistics.org), which carries with it an annual subscription to *JSL*.

Airmail supplement for institutions outside of North America: $11.00/year.

A limited number of back issues are available for $24.00 apiece.

Indexing and Abstracting: *American Bibliography of Slavic and East European Studies, ERIH (European Reference Index for the Humanities and Social Sciences), Humanities International Index, IBZ (Internationale Bibliographie der Geistes- und Sozialwissenschaftlichen Zeitschriftenliteratur), MLA International Bibliography (Modern Language Association), OCLC ArticleFirst, and Web of Science Emerging Sources Citation Index*

This issue was published in March 2022.

ISSN: 1068-2090

Technical Editor: Jordan Hussey-Andersen

Cover design: Nikki Life

Slavica Publishers
Indiana University
1430 N. Willis Drive
Bloomington, Indiana USA
47404-2146

[Tel.] 1-812-856-4186
[Fax] 1-812-856-4187
[Toll-free] 1-877-SLAVICA
slavica@indiana.edu
http://www.slavica.com

journal of slavic linguistics

volume 29, number 2

summer–fall 2021

Special Issue

Exploring the Impersonal Domain: Empirical Observations from Slavic

Edited by

Katrin Schlund and Peter Kosta

articles

From the Editors

This is the second issue of volume 29. It is a special issue entitled *Exploring the impersonal domain: Empirical observations from Slavic* and guest-edited by Katrin Schlund and Peter Kosta.

With the completion of issue 29.1, Jordan Hussey-Andersen took over from Renata Uzzell as JSL managing editor; we thank Renata for her service, and we welcome Jordan to the team. We also thank Frank Gladney for continued help with language editing.

While this issue was in production, SLS issued a "Position Statement on the Russian Invasion of Ukraine". The text is published as part of the front matter in this issue. This is an official statement of the Slavic Linguistics Society, unrelated to this special issue, and need not reflect the views of the contributors to this issue.

We welcome new submissions through our website: http://ojs.ung.si/index.php/JSL.

Franc Marušič and Rok Žaucer
University of Nova Gorica
franc.marušič@ung.si & rok.zaucer@ung.si

On February 24, 2022, at 05:55 Moscow time, after several weeks of military preparations along the eastern Ukrainian border, Vladimir Putin, President of Russia, announced the initiation of what he referred to as a "military operation" in the Donbas region of Ukraine. Ukrainian President Volodymyr Zelensky swiftly enacted martial law and ordered a military response against the ingress of Russian troops. The combat continues to escalate; though mainly focused in the east, conflict has been reported in the major Ukrainian cities of Kyiv, Kharkiv, and Odesa. [British Broadcasting Corporation. (updated 24 February 2022). Ukraine conflict: what we know about the invasion. https://www.bbc.com/news/world-europe-60504334]

In response to these events, the Slavic Linguistics Society issued the following statement:

Position Statement on the Russian Invasion of Ukraine

We, the Slavic Linguistics Society, are an international organization dedicated to scholarship of the Slavic languages and focused on encouraging research across a broad variety of domains in the field of Slavic Linguistics. A cornerstone of our organization and one of its founding principles is that, in contrast to other groups, the Slavic Linguistics Society maintains a panoptic approach to Slavic linguistic scholarship. We embrace research in all subfields, from various theoretical and analytical perspectives, and addressing any and all of the languages across the kaleidoscopic spectrum of Slavic. Truly, the single unifying feature of our multifaceted and diverse membership is that fundamentally we are all Slavists.

As Slavists, we are placed unequivocally within the sphere of Slavic culture and life, and therefore inevitably, politics. As such, we are not only in a position to address the ongoing situation in Ukraine, but we are under a clear ethical obligation to do so.

The Slavic Linguistics Society stands firmly in solidarity with Ukraine. We recognize Ukraine's linguistic, cultural, and political autonomy, and its consequent rights to self-determination and self-governance without the interference or intervention of outside entities. We consider the current military intervention led by President Putin to be a transparent transgression of those rights, and we therefore condemn it.

We acknowledge and support the large and growing body of dissenting Russian and Belorusian citizens and nationals, who by their

oppositionto President Putin's actions and to the complicity of their governments put themselves at great personal risk.

Furthermore, the Slavic Linguistics Society is morally supportive of our members, colleagues, friends, and associates who are currently in Ukraine and who find themselves in the midst of battle. We stand in steadfast solidarity with them.

This is the formal position of the Slavic Linguistics Society, which may differ from that of individual members and affiliates.

Drafted on February 24, 2022 (PT)
First Updated on February 26, 2022 (PT)
By the Slavic Linguistics Society Executive Board

Introduction

Katrin Schlund and Peter Kosta

Impersonal constructions have always intrigued syntacticians because they run counter to the traditional definition of a sentence as including a nominative subject and an agreeing predicate. Therefore, as Siewierska (2008b: 115) puts it, "[t]he notion of impersonality is a broad and disparate one". The Slavic languages, as is well known, are particularly rich in impersonal constructions, which is why their analysis has long been a center of interest.

Research about impersonals in Slavic began with the advent of the first handbooks and grammars dealing with syntax at the turn of the twentieth century (above all, Miklosich 1883; Jagić 1899; Potebnja 1899; Peškovskij 1914; Vondrák [1906] 1928; not to forget Havránek's 1928, 1937 fundamental works). The first specific studies of impersonals, including monographs, appeared in the 1950s (e.g., Fodor 1957; Galkina-Fedoruk 1958; Micklesen 1968; Doros 1975; Wolińska 1978). During the last third of the twentieth century, generative accounts have taken up a growing share of the literature, with two related but distinguishable points of focus. Accounts with the first type of focus seek to integrate impersonal structures into a broader typology of diathesis (e.g., Růžička 1986; Kosta 2021). Other generativist studies have analyzed impersonals against the background of syntactic unaccusativity (e.g., Harves 2002; Szucsich 2007; Lavine and Franks 2008; Lavine 2010, 2014).

The last two decennia have seen a peak in interest in impersonal constructions, with an emphasis on comparative studies and typology, both within and outside of Slavic linguistics. One pioneering effort regarding Slavic linguistics is the overview of impersonal structures provided by Mrazek 1990. The growing interest in impersonality also appears in anthologies, some with and some without the consideration of Slavic languages (e.g., Siewierska 2008a; Kor Chahine 2013; Redder 2012; Herbeck, Pöll, and Wolfsgruber 2019).

One of the most influential recent typological accounts is the functionally based outline given in Malchukov and Ogawa 2011. With reference to Siewierska 2008b, Malchukov and Ogawa include impersonals in the domain of agent-defocusing devices (other such constructions are passives or de-causatives). Given that impersonal constructions lack a full-fledged subject not only in terms of formal (structural, behavioral) but also functional (that is,

Journal of Slavic Linguistics 29(2): 115–21, 2021.

semantic and pragmatic) criteria (Malchukov and Ogawa 2011: 22), the authors distinguish impersonals with respect to the semantic-pragmatic subject property they mostly lack as Agentivity impersonals (A-impersonals), Reference impersonals (R-impersonals), and Topicality impersonals (T-impersonals).[1] Typical examples of A-impersonals are weather impersonals (e.g., Russian *gremit* 'it thunders'), impersonals denoting physical and emotional states (e.g., Bosnian/Croatian/Serbian *dosadno mi je* 'I am bored'; lit. 'it is boring to me'), and modal impersonals (e.g., Bulgarian *trjaba da* 'it is necessary to').[2] R-impersonals involve a human agent; crucially, the referential status of this agent is decreased. A typical instance of R-impersonals in Slavic is 3PL impersonals, traditionally referred to as *neopredelënno-ličnye predloženija* 'indefinite-personal sentences' in Russian. T-impersonals are not very frequent in Slavic, because they signal non-topicality of the subject referent. As is well known, Slavic languages make use of word order to signal non-topicality of the subject referent by putting the subject constituent in post-verbal position. Therefore, Slavic languages are not in need of specialized T-impersonals.[3] Some existential constructions, however, may also be classified as T-impersonals in Slavic. Cases in point are existential constructions with the verb 'have' in Polish and Bosnian/Croatian/Serbian, or the Russian reflexive existential verb *imet'sja*.

There is also a growing body of work suggesting typologies of impersonals, either for individual languages (e.g., Kibort 2008 for Polish; Babby 2010 and Schlund 2018 for Russian) or for subsets of impersonal constructions across languages (e.g., Siewerska and Papastathi 2011; Gast and van der Auwera 2013).

A new strand of research in impersonality seeks to assess how particular impersonal constructions are actually used—that is, how they function in discourse. Zinken's 2016 study of (impersonal) requesting strategies in Polish and English and Mazzitelli's 2019 analysis of Lithuanian reference impersonals are pioneering studies in this regard.

This Special Issue takes up the empirical, typological, and discursive trend of analyzing impersonal constructions. The contributions by Anastasia Bauer, as well as Maria Katarzyna Prenner and Daniel Bunčić, are concerned with the empirical analysis of R-impersonals. Whereas Prenner and Bunčić focus on three types of R-impersonals in Polish, Bauer offers a comparative study of various linguistic strategies used in six Slavic languages to render arbitrary human reference. In a careful quantitative study, Bauer analyzes more than

[1] Note that most instances of impersonal constructions are mixed types, typically with one factor predominating over the others.

[2] Interestingly, these three types of A-impersonals form the oldest layer of impersonal constructions in Indo-European languages (Bauer 2000: 96f.).

[3] An illustrative case of a T-impersonal is the "presentational inversion construction" (Creissels 2019: 6, 11) in French.

5,300 examples and singles out 18 strategies used in the Slavic translations of the German impersonal pronoun *man*.[4] This allows her to reveal differences in the use of these strategies across languages and groups of languages. Bauer also pays due diligence to potential caveats with regard to her study design. She shows, for instance, that the language of the original is a crucial predictor of the translation strategy used in the Slavic target language.

Jasmina Grković-Major investigates when and why certain types of Proto-Slavic A-impersonals evolved into personal constructions in contemporary Slavic. She focuses on impersonal constructions with accusative and dative experiencers, some of which have developed into middle and personal constructions in contemporary Slavic languages. These constructions were built from *ē*-statives (infinitives in *-ěti*) denoting sensations, emotions, perception, and cognition. Carefully evaluating data from historical stages of Slavic languages, Grković-Major shows that the transformation of these impersonals into canonical (personal) constructions with nominative-accusative alignment included primarily *ē*-statives whose experiencers were marked with the agentivity features of volition and control. Such experiencers occurred in impersonals denoting emotions because emotions imply a conscious, human participant and not merely an animate participant. Constructions denoting negative bodily sensations, in turn, imply a lesser degree of volition and control on the part of the primary participant, which is why most of these constructions have retained their status as impersonals up to the present day. Grković-Major points to the parallelism in markedness in semantic and morpho-syntactic terms manifest in this latter construction type.

Maria Katarzyna Prenner and Daniel Bunčić investigate the factors motivating the variation between three "quasi-synonymous" Polish R-impersonals. The three constructions are the *-no/-to* construction, the reflexive impersonal, and the 3pl impersonal. All three constructions include a demoted, arbitrary human participant with reduced referentiality. The authors extract predictions about the use of the *-no/-to* construction, the reflexive impersonal, and the 3pl impersonal from the available literature and test them in an explorative corpus study and with an acceptability judgment test among native speakers. Investigating the morphological and contextual variables, Prenner and Bunčić draw a detailed picture of the variables underlying the choice between the three constructions in contemporary Polish, including register, tense, generic vs. specific reading, and, for the first time, also the category of aspect. The multifactorial analysis provides insight into the complex interplay of these variables, makes it possible to estimate their relative weight, and points to potential additional factors.

[4] The study is conducted on the data collected in the ParaSol corpus, a parallel corpus of Slavic and other languages (von Waldenfels and Meyer 2006–).

Katrin Schlund examines the notorious issue of quantified subjects (QSs) and agreement in Polish. The referential status of QSs is typically reduced, particularly with QSs denoting numbers higher than five or unspecific quantifiers. Therefore, QS constructions are associated with R-impersonals, but the subject properties of agentivity and topicality are often also reduced. Comparing the strict rules of agreement resolution with QSs in Polish with the pragmatically and semantically motivated variation observable in Russian and, to a minor extent, Bosnian/Croatian/Serbian, Schlund asks why there is no semantically and pragmatically determined variation in the agreement resolution in QSs in contemporary Polish. Data from historical texts and previous analyses show that the resolution of QSs historically has likewise tended to mark QSs with strong subject properties (such as the animacy, strong agentivity, referentiality, and topicality of the QS) with semantic (that is, plural) agreement, and weak subject properties with grammatical (that is, singular) agreement. Finally, focusing on the deviant behavior of Polish virile QSs with paucal numbers 2–4, Schlund takes the side of the "accusative hypothesis", assuming that these oblique forms (*dwóch*, *trzech*, and *czterech*) are accusatives, and points out that they developed later than the regular nominative virile forms of *dwaj*, *trzej*, and *czterej*. A short corpus analysis suggests that in contemporary Polish the now-vanishing nominative virile forms are still preferred over the accusative forms precisely in contexts of increased referentiality and for pure naming. From this perspective, the nominative and accusative forms of Polish virile QSs can be interpreted as instances of differential subject marking.

The papers gathered in this volume look back at a "joint history" of presentations and discussions at various occasions, including conferences, exam colloquia at the University of Cologne, and numerous lunch and coffee breaks, which are, we hope, pleasantly remembered not only by ourselves but also by our colleagues. Jasmina Grković-Major, Maria Katarzyna Prenner, and Daniel Bunčić participated in a panel on impersonal constructions at the Conference on Explanation and Prediction (CEP) held in February 2019 at the University of Heidelberg.[5] In September 2019, Katrin Schlund chaired a panel about impersonal constructions at the Congress of the German Association of Slavists in Trier, Germany, in which Anastasia Bauer, Daniel Bunčić, and Maria Katarzyna Prenner presented the development of their research.[6]

All contributions have undergone a thorough double-blind reviewing procedure. We would therefore like to thank the anonymous reviewers for their careful reading and precious advice. Our sincere thanks also go to the

[5] The conference was funded by the German Research Foundation and organized by the editors of this volume (cf. Kosta and Schlund 2021).

[6] We regret that Aleš Půda (Heidelberg), who gave an inspiring talk about impersonal reflexives in Russian and Czech on this occasion, could not publish in this volume.

editors of the *Journal of Slavic Linguistics*, Franc Marušič and Rok Žaucer, for their patience and assistance in the preparation of this volume.

References

Babby, Leonard. (2010) "Prolegomenon to any future typology of impersonal constructions". Donna B. Gerdts, John C. Moore, and Maria Polinsky, eds. *Hypothesis A/Hypothesis B. Linguistic explorations in honor of David M. Perlmutter*. Cambridge: MIT Press, 19–40.

Bauer, Brigitte. (2000) *Archaic syntax in Indo-European: The spread of transitivity in Latin and French*. Berlin and New York: Mouton de Gruyter. [Trends in Linguistics. Studies and Monographs, 125.]

Creissels, Denis. (2019) "Impersonal and related constructions: A typological approach". Available at: https://youngconfspb.com/application/files/2615/7480/4139/Creissels-Impers.Constr.pdf.

Doros, Aleksander. (1975) *Werbalne konstrukcje bezosobowe w języku rosyjskim i polskim na tle innych języków słowiańkich*. Wrocław, Warsaw, Cracow, and Gdańsk: Wydawnictwo polskiej akademii nauk. [Prace komisji słowianoznawstwa, 32.]

Fodor, István. (1957) "Subjektlose Sätze im modernen Russischen". *Studia Slavica Academiae Scientiarum Hungaricae* 3: 149–206.

Galkina-Fedoruk, Evdokija M. (1958) *Bezličnye predloženija v sovremennom russkom jazyke*. Moscow: Izdat. Mosk. Univ.

Gast, Volker and Johan van der Auwera. (2013) "Towards a distributional typology of human impersonal pronouns, based on data from European languages". Dik Bakker and Martin Haspelmath, eds. *Languages across boundaries: Studies in memory of Anna Siewierska*. Berlin: de Gruyter Mouton, 119–58.

Kor Chahine, Irina, ed. (2013) *Current studies in Slavic linguistics*. Amsterdam: Benjamins, 1–22. [Studies in Language Companion Series, 146.]

Harves, Stephanie A. (2002) *Unaccusative syntax in Russian* (Order No. 3067146). Ph.D. dissertation, Princeton University. [ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection, 305496700.] Available at: https://search.proquest.com/docview/305496700?accountid=11359.

Havránek, Bohuslav. (1928) *Genera verbi v slovanských jazycích*. Vol. 1. Prague: KČSN.

————. (1937) *Genera verbi v slovanských jazycích*. Vol. 2. Prague: KČSN.

Herbeck, Peter, Bernhard Pöll, and Anne Wolfsgruber, eds. (2019) *Semantic and syntactic aspects of impersonality*. Hamburg: Buske. [Linguistische Berichte, 26.]

Jagić, Vatroslav. (1899) *Beiträge zur slavischen Syntax*. Vienna: Gerold.

Kibort, Anna. (2008) "Impersonals in Polish: An LFG perspective". *Transactions of the Philological Society* 106(2): 246–89.

Kosta, Peter. (2021) "The Unified Voice Phase Hypothesis and the nature of parameterized functional heads in a phase-label-driven model of Radical Minimalism with special reference to Slavic languages". Peter Kosta and Katrin Schlund, eds. *Keynotes from the International Conference on Explanation and Prediction in Linguistics (CEP): Formalist and Functionalist Approaches. Heidelberg, February 13th and 14th, 2019*. Berlin, Bern, and Brussels: Peter Lang, 149–211. [Potsdam Linguistic Investigations, 33.]

Kosta, Peter and Katrin Schlund, eds. (2021) *Keynotes from the International Conference on Explanation and Prediction in Linguistics (CEP): Formalist and Functionalist Approaches. Heidelberg, February 13th and 14th, 2019*. Berlin, Bern, and Brussels: Peter Lang. [Potsdam Linguistic Investigations, 33.]

Lavine, James E. (2010) "Case and events in transitive impersonals". *Journal of Slavic linguistics* 18(1): 101–30.

————. (2014) "Anti-Burzio predicates: From Russian and Ukrainian to Icelandic". *Journal of the Moscow State University for the Humanities. Philology series* 2: 91–106. Available at: http://www.facstaff.bucknell.edu/jlavine/publications/JE_Lavine_AntiBurzio_Predicates.pdf. Last accessed 12 May 2018.

Lavine, James and Steven Franks. (2008) "On accusative first". Andrei Antonenko, John F. Bailyn, and Christina Y. Bethin, eds. *Formal Approaches to Slavic Linguistics: The Stony Brook Meeting*. Ann Arbor: Michigan Slavic Publications, 231–47.

Malchukov, Andrej and Akio Ogawa. (2011) "Towards a typology of impersonal constructions: A semantic map approach". Andrej Malchukov and Anna Siewierska, eds. *Impersonal constructions: A cross-linguistic perspective*. Amsterdam: Benjamins, 19–56. [Studies in Language Companion Series, 124.]

Mazzitelli, Lidia Federica. (2019) "Referential and pragmatic-discourse properties of Lithuanian reference impersonals: 2SG-IMP, 3-IMP and ma/ta-IMP". *Kalbotyra 72*: 32–57.

Micklesen, Lew R. (1968) "Impersonal sentences in Russian". Henry Kučera and William E. Harkins, eds. *American contributions to the 6th International Congress of Slavists, Prague 1968, August 7–13*. Vol. 1. *Linguistic contributions*. The Hague and Paris: Mouton, 275–331. [Slavistic Printings and Reprintings, 80.]

Miklosich, Franz. (1883) *Subjectlose Sätze*. Vienna: Braunmüller.

Mrazek, Roman. (1990) *Sravnitel'nyj sintaksis slavjanskix literaturnyx jazykov*. Vol. 1. *Isxodnye struktury prostogo predloženija*. Brno: Univ. J. E. Purkyně v Brně. [Opera Universitatis Purkynianae Brunensis Facultas Philosophica/ Spisi Univerzity J. E. Purkyně v Brně Filozofická Fakulta, 289.]

Peškovskij, A. M. (1914/1956) *Russkij sintaksis v naučnom osveščenii*. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izdatel'stvo ministerstva prosveščenija RSFSR.

Potebnja, Aleksandr A. (1899/1968) *Iz zapisok po russkoj grammatike.* Vol. 3. *Ob izmenenii značenija zamenax suščestvitel'nogo*. Moscow: Prosveščenie.

Redder, Angelika, ed. (2012) *„Unpersönliche Konstruktionen": Prädikationsformen funktional und sprachübergreifend betrachtet*. Munich: Iudicium.

Růžička, Rudolf. (1986) "Typologie der Diathese slavischer Sprachen in parametrischen Variationen". *Die Welt der Slawen* 31: 225-274.

Schlund, Katrin. (2018) "A unifying approach to Russian impersonal constructions". *Zeitschrift für Slawistik* 63(1): 120–68.

Siewierska, Anna, ed. (2008a) "Impersonal constructions in grammatical theory" [Special issue]. *Transactions of the Philological Society* 106(2).

————. (2008b) "Introduction: Impersonalization from a subject-centered vs. agent-centered perspective". *Transactions of the Philological Society* 106(2): 115–37.

Siewerska, Anna and Maria Papastathi. (2011) "Towards a typology of third person plural impersonals". *Linguistics* 49(3): 575–610.

Szucsich, Luka. (2007) "Nothing wrong with finite T: Non-agreeing accusative impersonal sentences". Magda Golędzinowska, Ulyana Savchenko, and Richard Compton, eds. *Formal Approaches to Slavic Linguistics 15: The Toronto Meeting 2006.* Ann Arbor: Michigan Slavic Publications, 417–35.

von Waldenfels, Ruprecht and Roland Meyer. (2006–) *ParaSol, a corpus of Slavic and other languages*. Available at: http://parasolcorpus.org/.

Vondrák, Václav. (1906/1928) O. Grünenthal, ed. *Vergleichende slavische Grammatik.* Vol. 2. *Formenlehre und Syntax*. Göttingen: Vandenhoeck and Ruprecht.

Wolińska, Olga. (1978) *Konstrukcje bezmianownikowe we współczesnej polszczyiźnie*. Katowice: Uniwersytet Śląski.

Zinken, Jörg. (2016) *Requesting responsibility: The morality of grammar in Polish and English family interaction*. Oxford: Oxford University Press. [Foundations of Human Interaction.]

Katrin Schlund
Department of Slavic Studies
Martin Luther University Halle-Wittenberg
Halle, Germany
katrin.schlund@slavistik.uni-halle.de

Peter Kosta
Department of Slavic Studies
University of Potsdam
Potsdam, Germany
peter.kosta@uni-potsdam.de

Impersonalization in Slavic: A Corpus-Based Study of Impersonalization Strategies in Six Slavic Languages*

Anastasia Bauer

Abstract: This paper gives a comprehensive overview of how impersonalization is expressed in Slavic. It presents the results of a comparative corpus study, outlining all possible strategies for expressing impersonalization in six Slavic languages (Russian, Ukrainian, Bulgarian, Croatian, Czech, and Polish), using German *man* as a filter. This paper shows on the basis of a random sample of over 5,000 translated sentences which impersonalization means Slavic languages use to express propositional content expressed by the pronoun *man* in German. Additionally, this pilot study answers two questions: (1) How do Slavic languages differ in the distribution of these impersonalization strategies? and (2) Are there major translation effects? The main findings are an outline of a cross-Slavic set of impersonalization strategies that reveals significant differences between the Slavic languages in the distribution of *man*-equivalents and a highly significant impact of the source language on the choice of the impersonalization strategy in translation.

1. Introduction

This paper reports the findings from the first corpus-based contrastive study of how impersonalization is expressed in Slavic. A large and varied group of

* An earlier version of this paper was presented at the 13th Deutscher Slavistentag in Trier, 24–26 September 2019. The present paper has profited from the discussion, questions, and comments during the panel, which was organized by Katrin Schlund. I thank the editors, Katrin Schlund and Peter Kosta, as well as my colleagues Daniel Bunčić, Lidia Federica Mazzitelli, and Anna-Maria Sonnemann for their suggestions and valuable feedback on the earlier versions of this paper. Special thanks are owed to Lidia Federica Mazzitelli, Tobias-Alexander Herrmann, Maria Katarzyna (Nini) Prenner, Fenja Erdmann, Iliana Chekova, Insa Kind, Vilen Adamskyi, Dragana Grbić, Jelena Chulum, Aleksandar Borkovac, and Giuseppe Russo for their assistance with the annotation of the language examples in the corpus. Advice given by Maximilian Hörl and Dominic Schmitz with regard to statistical analysis in RStudio has been a great help. I am also grateful to Volker Gast and two anonymous reviewers for *JSL* for a careful reading of the text and many invaluable comments and suggestions that helped in improving this paper.

constructions used in various Slavic languages as means of impersonalization will be presented in this paper. Impersonalization largely comprises human impersonal constructions like German *man* (e.g., *Man sollte wegen des Coronavirus zu Hause bleiben* 'One should stay at home because of the coronavirus'). Although a number of studies have been devoted to impersonal constructions in the Slavic languages (Guiraud-Weber and Kor Chahine 2013), we are still lacking distributional and contrastive investigations of such constructions in Slavic from a typological point of view. Drawing on recent theoretical work by Gast and van der Auwera (2013), who researched the distributional typology of impersonal pronouns, and using data from a parallel corpus, this study identifies the distribution and frequency of impersonalization strategies used across six selected Slavic languages representing all three Slavic language subfamilies (West, East, and South Slavic). The data in this study presents inter- and intra-group variations in the use of impersonalization strategies. The goal is to reveal the main differences between Slavic languages and to show what might have an impact on the choice of an impersonalization strategy.

The main purpose of this corpus-based study is to outline the strategies for expressing impersonalization in six Slavic languages (Russian, Ukrainian, Bulgarian, Croatian, Czech, and Polish) (see Section 3). While many studies concentrate on the inventory of impersonal pronouns only (e.g., Gast and van der Auwera 2013; van Olmen and Breed 2018), this paper goes beyond impersonal pronouns and describes other means of expressing impersonal reference in Slavic that have received little to no attention in the literature on impersonals. This study provides a cross-Slavic distribution of impersonalization strategies and can serve as a basis for further investigations of selected constructions and for establishing a detailed typology of impersonal constructions in the Slavic languages.

The paper is structured as follows: Section 1.1 explains how the notion of impersonalization is understood in this study and summarizes relevant research on impersonal constructions in Germanic and Slavic languages. Section 1.2 presents the latest research on the typology of human impersonal pronouns. Section 2 introduces the research questions (2.1), provides a descriptive overview of the data (2.2), makes some remarks about the corpus, and explains the methods used for the present study (2.3). In Section 3, I present the findings and outline the 18 strategies for expressing impersonalization in six Slavic languages. A cross-Slavic comparison of these impersonalization strategies is presented in Section 4. Section 5 reveals inevitable translation effects in the choice of impersonalization strategy. Section 6 offers a discussion of the data, and Section 7 summarizes and concludes the study.

1.1. Impersonalization

Impersonal reference and the linguistic means of expressing it in the world's spoken languages[1] have received a lot of attention in the literature of late (e.g. Cabredo Hofherr 2017; Siewierska 2008; Malchukov and Siewierska 2011; Zobel 2012; Guiraud-Weber and Kor Chahine 2013). There are a large number of typological studies on impersonal pronouns in Germanic and Romance languages (Siewierska and Papastathi 2011; van der Auwera et al. 2012; Gast and van der Auwera 2013; van Olmen and Breed 2018) and a considerable amount of research on specific constructions in individual Slavic languages (Padučeva 2012; Nikitina 2011; Guiraud-Weber and Kor Chahine 2013; Schlund 2018a; Bunčić 2019) or contrastive studies between Slavic (mostly Russian) and non-Slavic languages (Dušková 1973; Anochina 1981; Berger 1991; Rudolf 2014; Lavine 2017). Some studies present an in-depth analysis of one or two impersonal constructions in a few Slavic languages, such as a detailed comparison of reflexive impersonals in Polish and Slovenian by Rivero and Milojevic Sheppard (2003); a contrastive study of the syntactic properties of adversative impersonals in Russian with *-no/-to* impersonals in Polish by Szucsich (2007); an investigation of *-no/-to* constructions, reflexive impersonals, and third-person plural impersonals in Polish and Russian by Prenner (forthcoming), and in Polish and Serbo-Croatian by Bunčić (2018); passive constructions and the third-plural impersonals in Russian and Bulgarian by Ivanova and Gradinarova (2015); reflexive impersonals in Slovenian and Russian by Uhlik and Žele (2018); and a study on accusative impersonals (also adversity impersonals, elemental constructions) in Russian and other Slavic languages by Schlund (2020). Yet very little attention has so far been paid to contrastive research of a wide range of impersonal constructions between a sizable number of Slavic languages.

No cover term for constructions expressing impersonality is commonly accepted in linguistics. Moreover, the notion of impersonality is extremely broad and is not uniformly interpreted by linguists. While some researchers interpret impersonalization in semantic terms, others adopt a morphological or syntactic perspective (see Siewierska 2008 for more details). From a semantic perspective, impersonality is divided into two notions depending on human agentivity (see also Malchukov and Ogawa 2011). In the first sense, constructions are considered impersonal when they depict events or situa-

[1] The investigation of impersonal reference has largely been constrained to the auditory modality. Sign languages, as languages in the visual-gestural modality, have not yet been well investigated with regard to impersonal reference. One exception is a special issue of *Sign Language & Linguistics* (2018) including a study on impersonal reference in Russian Sign Language (Kimmelman 2018).

tions brought about by an unspecific *non-human* agent,[2] as in (1). In the second sense, constructions are also regarded as impersonal when they depict events or situations brought about by an unspecific *human* agent (Siewierska 2008), as in (2). In this article, I concentrate on devices expressing impersonality in the latter sense of the term.

(1) Dorogu zasypalo peskom. (Russian)
street showered$_{3SG.N}$ sand
'The street got strewed with sand.'
Lit. 'It strewed the street with sand.'

(2) Dorogu zasypali peskom.
street showered$_{3PL}$ sand
'People/Someone strewed the street with sand.' (Mel'čuk 1974: 350)[3]

The term "impersonal" for constructions as in (2), which depict events brought about by a non-specific human agent, has been criticized by some linguists, since these constructions are never impersonal in the strict sense (Kitagawa and Lehrer 1990; Padučeva 2012; Plungjan 2016). Instead, they may generalize over individuals or may be vague and refer to a specific group of individuals who cannot or should not be identified by the speaker. In Russian grammars and Russian theoretical research, such constructions are therefore approached at the sentential level and are traditionally referred to as *odnosostavnye predloženija* 'one-argument sentences', which are in turn subdivided into *neopredelenno-ličnye predloženija* 'indeterminate-personal sentences', as in (2), and *obobščenno-ličnye predloženija* 'generalized-personal sentences', as in (3a–b) (Šaxmatov 2001; Švedova 1980; Valgina 2003; Padučeva 2012).

(3) a. Cypljat po oseni sčitajut. (Russian)
chickens in fall count$_{3PL}$
'Do not count your chickens before they hatch.'
Lit. 'Chickens are counted in the fall.'

b. Na vsex ne ugodiš'.
on everyone not please$_{2SG}$
'You cannot please everyone.' (Valgina 2003: 160)

[2] Such constructions are termed "Elemental Constructions" (EC) as a translation of the Russian expression *stixijnaja konstrukcija* (see Schlund 2018b for a thorough investigation of EC in contemporary Russian).

[3] All sentences originally available in Cyrillic in the corpus or literature (e.g. in Russian, Ukrainian, or Bulgarian) were transliterated here by the author.

However, the above-mentioned Russian classification does not seem to be very satisfactory either. Firstly, as already mentioned by Padučeva (2012: 27), both subtypes (indeterminate-personal as well as generalized-personal sentences), according to Russian grammars (Vinogradov 1954; Švedova 1980), encompass the same syntactic constructions—i.e., the 3rd-person plural form of the verb, as in (2) and (3a). As for generalized-personal sentences, they do not form a single class syntactically; they cover the 3rd- and 1st-person plural, as well as the 1st- and 2nd-person singular forms of the verb (Švedova 1980). Secondly, as already mentioned by Bunčić (2018, 2020), the term "indeterminate-personal" is unfortunate and even misleading because these constructions are, on the one hand, not necessarily indefinite as the Russian term implies (cf. Berger 1991: 72; Gast and van der Auwera (2013: 26) distinguish between definite and indefinite reference). On the other hand, it is difficult to refer to these constructions as personal, since the subject is not fully referential and not overt.

As none of the above labels appear to be perfect and this paper aims to describe a wider range of constructions, I will use the notion of impersonalization as it is defined by Gast and van der Auwera (2013). They define it as "the process of filling an argument position of a predicate with a variable ranging over a set of human participants without establishing a referential link to any entity from the universe of discourse" (2013: 136). Using the German impersonal pronoun *man*, which epitomizes impersonalization as defined above (in Section 2.3), various Slavic impersonalization strategies will be described. Thus, the impersonalization strategies in this study are translation strategies for German *man.* These strategies in Slavic include the impersonal uses of 3PL, 2SG, and 1PL; the impersonal passive; the *-no/-to* construction; reflexive impersonals; generic nouns such as Czech *člověk* or Ukrainian *ljudina;* bare infinitives; modals; adverbial and participle constructions; as well as some other minor devices to be discussed in Section 3. Consider the following German sentence in (4):

(4) An jeder Straßenecke riecht man es.
on each street.corner smell_{3SG} IMPS it_{ACC}
'You can smell it at every corner.'[4]

This study will show a set of possible structures, termed impersonalization strategies, that are used in Slavic as equivalents of German *man,* as shown in (4) above. It is important to note that some strategies—such as, for example, bare infinitives, modals, or 1PL constructions—have clearly not been referred to as impersonal in the literature, but they can also be used in the Slavic languages for argument backgrounding in some impersonal contexts.

[4] All examples are taken from the ParaSol corpus (see Section 3.2) unless stated otherwise.

1.2. Human Impersonal Pronoun Typology

Siewierska and Papastathi (2011) establish a typology of 3PL impersonals on the basis of data from ten European languages (including Russian and Polish). They show that the languages under study differ significantly in their usage of 3PL impersonals. Following up on work done by Siewierska (2008), Gast and van der Auwera (2013) investigate a wide range of impersonal pronouns in several European languages (including Russian and Bulgarian) on the basis of corpus data and determine the factors which are relevant to their distribution cross-linguistically. Gast and van der Auwera (2013) propose that the contexts in which human impersonal pronouns are used can be classified according to two major groups of parameters: state of affairs and quantification (see Figure 1 below).

Not all of these hierarchically ordered feature combinations appear to be possible in the languages they investigate. The authors present a connectivity map for human impersonal pronouns. The various contexts (from Figure 1) form a semantic map in the shape of a ring, as shown in Figure 2. The most crucial point for the present study is the fact that only one strategy, according to Gast and van der Auwera (2013: 30), the Germanic human impersonal pronoun *man*/*men*, or French *on*, can cover the entire map. That means this pronoun can be used in all types of contexts (as for example in node 1, *Man klopft an der Tür* 'They're knocking on the door', or in node 5, *Man lebt nur einmal* 'You only live once').

All other impersonal pronouns are restricted with regard to the number of contexts they are used in and can thus cover only a subset of connected regions on this map. Thus, the English indefinite pronoun *someone* or the Russian modal/infinitive construction such as *možno govorit'* 'one can talk' can only cover two nodes (1 and 7); English and Russian 3PL impersonals cover the regions 1 to 4; the English 2SG as well as Bulgarian *čovek* are used in contexts 5 to 7.

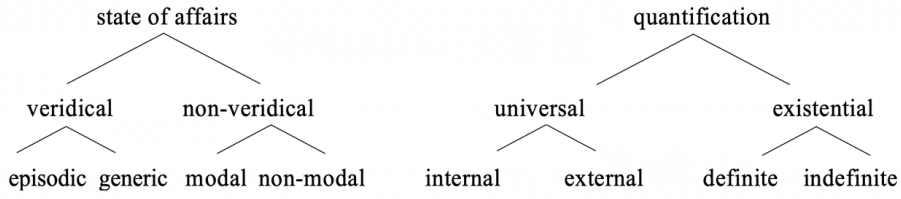


Figure 1. Classification trees for two parameters
(Gast and van der Auwera 2013: 24–26)

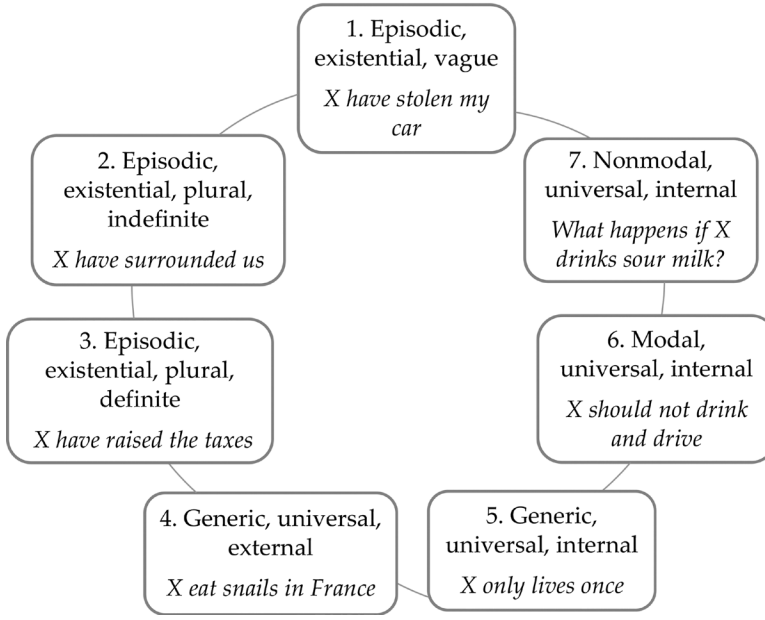


Figure 2. Semantic map of impersonal reference (adapted from Gast and van der Auwera 2013)[5]

2. Methodology

2.1. Research Questions

This study aims to describe the strategies used in Slavic languages to express impersonalization. Drawing on research by Gast and van der Auwera (2013) and Gast (2015), I introduced the notion of impersonalization in Section 1.1 and pointed out that the German pronoun *man* occupies a unique place in the semantic map of impersonal reference, since it can occur in all possible contexts. Slavic languages do not have a specialized impersonal pronoun like German *man*. The known Slavic counterparts such as Russian *čelovek*, Polish *człowiek*, or Czech *člověk* are not widely used, as will be seen in Section 3, and cannot be used in a wide range of contexts. Thus, the question arises as to which means of impersonalization are used in Slavic to render the meanings of the German pronoun *man*, as in example (4) above.

[5] Each context is illustrated with a representative sentence, where *X* stands for the impersonal argument. For an explanation of the map and the various contexts, the reader is referred to Gast and van der Auwera 2013.

The contrastive study of impersonalization in English and German (Gast 2015) has successfully implemented the pronoun *man* as a "filter" for identifying sentences expressing impersonalization (also Rudolf 2014). Following Gast, this study also uses German *man* as a "methodological anchor" to find the Slavic strategies corresponding to this pronoun (see Section 2.3 for more details). In particular, the current article seeks to address the following questions:

(i) How is impersonalization expressed in Slavic? (by asking what types of structures Slavic languages use to render propositional content expressed by the pronoun *man* in German)
(ii) How do Slavic languages differ with regard to the distribution and use of impersonalization strategies?
(iii) Are there translation effects?

The third question is motivated by previous findings showing that the direction of translation in the corpus influences the choice of the structure used. Such translation effects have been identified by Siewierska and Papastathi (2011) in their investigation of 3PL impersonals on the basis of ten European languages and by Gast (2015) in the above-mentioned contrastive study of German and English impersonalization strategies. We thus ask whether the *man*-equivalents in Slavic that we find in our data are strongly influenced by the structure in the original text.

2.2. Database of Slavic Impersonalization Strategies Using ParaSol

To reach the aim of the study, one would need a large parallel corpus including all Slavic languages, comprising different genres, and containing a good representation of spoken interaction as well as written language. Unfortunately, we do not yet have the required parallel language corpora in Slavic (Divjak et al. 2017). To the best of my knowledge, there are no parallel corpora of colloquial speech or spontaneous interactions in Slavic languages. Thus, the ParaSol corpus was chosen as the best alternative to study impersonalization strategies in Slavic.

ParaSol, formerly known as the Regensburg Parallel Corpus and originally developed by Ruprecht von Waldenfels and Roland Meyer, is a parallel aligned corpus of translated and original fictional texts in Slavic and other languages (von Waldenfels and Meyer 2006–; von Waldenfels 2006, 2011). ParaSol contains not only texts of Slavic languages, but also languages such as German, English, French, and Italian. It is free of charge to access but requires registration.

To get an initial idea of the types of strategies that we find in Slavic languages rendering the content of the German impersonal pronoun *man*, a da-

tabase of *man*-equivalents in six Slavic languages was developed on the basis of three parallel texts from ParaSol: (1) the German novel *Das Parfum: Die Geschichte eines Mörders* (*Perfume: The Story of a Murderer)* by Patrick Süsskind; (2) the Czech novel *Nesnesitelná lehkost bytí* (*The Unbearable Lightness of Being*) by Milan Kundera; and (3) the Italian novel *Il nome della rosa* (*The Name of the Rose*) by Umberto Eco. The choice of these three texts was dictated by the availability of parallel texts in German and the largest number of various Slavic languages in ParaSol. The chosen texts from the ParaSol corpus provide translations in German, Russian, Ukrainian, Bulgarian, Croatian, Czech, and Polish. Thus, the data used for the present study contain three types of texts: (1) German original sentences and their translations into six Slavic languages (n = 180 sentences), (2) Czech original sentences and their translations into German and five Slavic languages (n = 240 sentences), and (3) German and Slavic sentences that were translated from Italian (n = 524 sentences). For each language, approximately 942 sentences were analyzed, making up a total number of 6,594 sentences. For some languages in the ParaSol corpus, a translation of German *man* was missing or the necessary context was not available. In such cases, the sentence was deleted from the analysis completely. This accounts for the different number of evaluated constructions provided in Figure 3. A random sample from the database used in this paper is made available for the reader at https://uni.koeln/5JEML.

2.3. German *man* as a "Methodological Filter"

While there is no corpus available which is tagged for impersonal forms in Slavic languages, using the German impersonal pronoun *man* or the French *on* as a methodological anchor has already proved successful in a number of previous linguistic studies (Anochina 1981; Rudolf 2014; Gast 2015; Zaliznjak and Kružkov 2016; Mazzitelli 2019). In my database extracted from ParaSol (see Section 2.2), the German impersonal pronoun *man* was used as a "filter" to identify the sentences in each of the six Slavic languages corresponding to this pronoun, independent of the original language. As one of the reviewers has pointed out, we should be aware that by using *man* as a filter, we surely miss usage types where *man* is dispreferred as an impersonalization strategy. Passivization, for example, is another commonly used impersonalization strategy in German, for which we will not find the impersonal equivalents in Slavic using this method. As the German pronoun *man* is always unambiguously impersonal, almost all the correspondences in the six Slavic languages in the dataset are also impersonal, except for paraphrased sentences which include fully referential personal uses of pronouns. Sentences containing personal pronouns as *man*-equivalents in Slavic were counted under "paraphrase", provided that the context allowed for the personal interpretation. Instances of impersonal translations of *man* into Slavic were extracted from the ParaSol

corpus manually, with sufficient context to check their impersonal meaning, and inputted into Excel sheets. For each language, the impersonalization strategies to be discussed in Section 3 were identified, annotated, and counted.

One can think of a wide range of factors influencing the choice of impersonalization strategy in a given language (see also Prenner and Bunčić, this volume, for a number of grammatical factors). Gast (2015) shows that various parameters should be taken into account, such as the semantic and syntactic context of a sentence, register (conversation, scientific), and translation effects. In the case of translated language, we expect to find some systematic differences between the original and the translated version. Translations are secondary texts influenced by the language of the original text (von Waldenfels 2012). We might thus expect that an impersonalization strategy in translation may differ due to language pair and translation direction. As we are dealing with translated texts in this study, it is vital to explore the translation effects.

3. How is Impersonalization Expressed in Slavic?

This section analyzes the data and presents the results of the study. Section 3.1 provides a descriptive overview of all Slavic impersonalization means found in the database. Following that, I present the results separately according to the language of the original text. Section 3.2 shows the *man*-equivalents in translations of the German original text in six Slavic languages. Section 3.3 presents the impersonalization means in the translations from the Czech novel, and Section 3.4 demonstrates the impersonalization strategies in the translated examples from Italian. The differences among the Slavic languages will be dealt with in Section 4.

3.1. Impersonalization Strategies

In the six Slavic languages under study, impersonalization is expressed by a great variety of means, as illustrated in Figure 3. All six languages employ similar strategies (with only three exceptions[6]) but vary in their distribution. This study reveals that these languages do not differ significantly in the range of construction types that they employ for impersonalization.

The corpus data in this study show that sentences with the German impersonal pronoun *man* can be rendered in the Slavic languages by 18 different impersonalization strategies. The various strategies and cumulative frequen-

[6] In the present data, the *-no/-to* construction is restricted to Polish only, although this construction is also said to be used in Ukrainian (Billings 1993). The modal/reflexive construction occurs in South Slavic languages only (exemplified in this study by Bulgarian and Croatian), and Bulgarian does not have infinitives or modal/infinitive constructions.

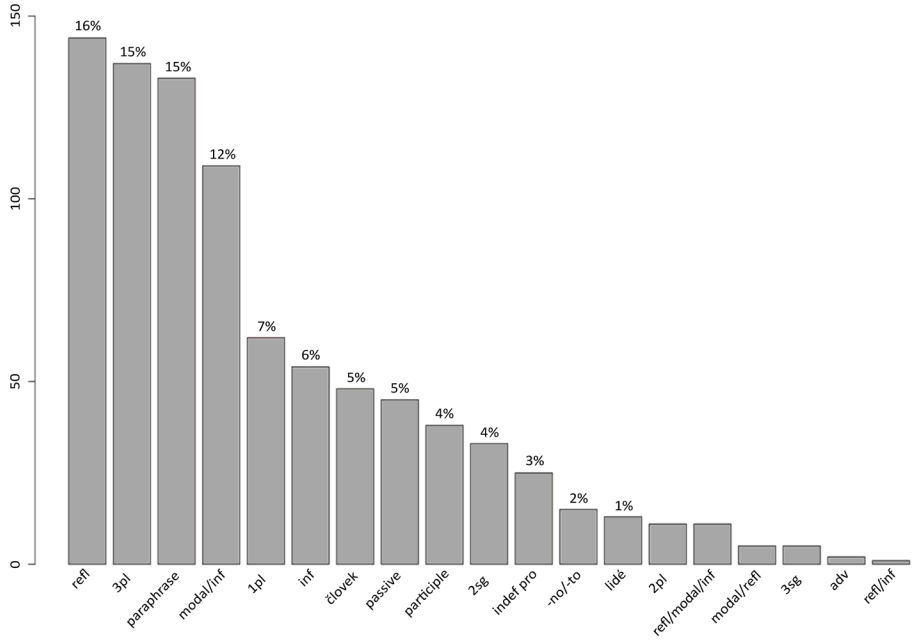


Figure 3. 18 Slavic impersonalization strategies and paraphrase in all three texts as mean numbers in absolute values and percentages (n = 5,345 sentences)

cies in all six languages are presented in Figure 3.[7] The classification used in this study is consistent across all six languages and is based on formal morphological description. All classification terms are to be understood as labels for certain surface configurations with no theoretical implications. Each label

[7] All statistics, as well as diagrams, in this paper were generated in R (R Core Team 2015) with RStudio (Version 1.2.5042). Figure 3 shows individual mean values of overall occurrences of impersonalization strategies in all six Slavic languages under study expressed in absolute values. The percentages of each strategy are represented above the bars. The diagram is based on the data presented in the Appendix (pp. 170–78). Here are the abbreviations used in Figure 3: refl—reflexive; 3pl—third-person plural constructions; modal/inf—modal infinitive constructions; 1pl—first-person plural constructions; inf—infinitive constructions; člověk—stands for all Slavic impersonally used nouns meaning 'human being' in singular: Cz *člověk*, Blg *čovek*, Rus *čelovek*, Pol *człowiek*, Cr *čovjek*, Ukr *ljudina*; 2sg—second-person singular constructions; indef pro—indefinite pronouns; lidé—stands for all Slavic impersonally used nouns meaning 'people' (Rus/Ukr/Cr *ljudi*, Cz *lidé*, Blg *xorata*, Pol *ludzie*); 2pl—second-person plural constructions; modal/refl—modal reflexive construction; 3sg—third-person singular constructions; and adv—adverbial constructions.

is demonstrated by an example. Some cases where multiple classifications are possible will be discussed in a short paragraph below. Note that the data were extracted from the ParaSol corpus by using German *man* as a filter in German sentences (see Section 2.3). Examples (5–22) are Slavic translations of the German *man*-sentences as they are found in the corpus.[8] The original text (German, Italian, or Czech) as well as the English translation as it appears in the corpus have been added for clarity. A larger random sample of the database is also represented in the Appendix.

(5) refl(exive)

PL Wtedy **zbierało się** je ostrożnie i rozkładało nowe kwiaty.

GE Dann zupfte **man** sie vorsichtig ab und streute frische Blüten aus. <original>

'Then they were carefully plucked off and new blossoms spread out.'

All reflexive forms with the successors of Proto-Slavic **sę*, appearing today as clitics or postfixes in Slavic languages, are deemed reflexive in this study as they are morphologically marked as such. Note that almost all of the examples in this study have an impersonal reading, and clear cases of personal reflexive forms, such as Rus *on breetsja* 'he shaves himself', are absent. Reflexive constructions with a "passive" meaning (e.g., Pol *buduje się willę* 'the villa is being built' or Rus *dveri otkryvajutsja* 'doors are opened') are also categorized as reflexive in this study.

(6) 3pl

CR Osam stotina godina **donosili** su tamo mrtvace […].

GE Achthundert Jahre lang hatte **man** hierher die Toten […] verbracht. <original>

'For eight hundred years the dead had been brought here….'

(7) modal/inf(initive)

UK V kožnomu zakutku **možna** ix **počuti**.

GE An jeder Straßenecke riecht **man** es. <original>

'You can smell it at every corner.'

[8] The language abbreviations used in the examples are the following: GE—German, BU—Bulgarian, CR—Croatian, RU—Russian, UK—Ukrainian, CZ—Czech, PL—Polish.

The Ukrainian example (7) demonstrates constructions defined in this study as modals with an infinitive. Typical polyfunctional modals (Rus *moč'*, Pol *móc*) as well as so-called modal content words or "semi-modals" (Rus *nužno, prixoditsja;* Pol *wolno*)—i.e., words with modal meaning that are not subject to an auxiliarization process and have only one modal meaning (Hansen 2005)—are deemed modals for the purpose of consistent classification across languages in this study. Therefore, such constructions as Rus *pridetsja priznat'* 'have to admit' are counted under the category modal infinitive.

(8) 1pl

CZ Ďábel je tupý, ve svých léčkách a ve svých svodech se drží svého, opakuje své obřady třeba po tisícilet, nemění se, a právě proto jej **můžeme** co nepřítele rozpoznat!

GE Der Dämon ist blöde und einfallslos, er hält sich in seinen Verlockungen und Verführungen an einen sturen Rhythmus, er wiederholt seine Riten über Jahrtausende, er bleibt sich immer gleich, und eben daran erkennt **man** ihn als den Feind!

IT Il demonio è ottuso, segue un ritmo nelle sue insidie e nelle sue seduzioni, ripete i propri riti a di stanza dimillenni, egli è sempre lo stesso, proprio per questo lo si riconosce come il nemico! <original>

'The Devil is stubborn, he follows a pattern in his snares and his seductions, he repeats his rituals at a distance of millennia, he is always the same, this is precisely why he is recognized as the enemy!'

(9) inf(initive)

RU Èto značit **poterjat'** vsjakuju silu.

GE Es bedeutet, daß **man** auf all seine Stärke verzichtet hat.

CZ To znamená pozbýt jakékoli síly. <original>

'It means losing all strength.'

Various infinitive verb forms (syntactically free, embedded, or combined with other constituents) are not further differentiated and are gathered together here under the category "infinitive".

(10) člověk

CZ **Člověk** to ucítí na každém rohu.

GE An jeder Straßenecke riecht **man** es. <original>

'You can smell it at every corner.'

"Člověk" stands in this classification for all Slavic impersonally used nouns meaning 'human being': Cz *člověk*, Blg *čovek*, Rus *čelovek*, Pol *człowiek*, Cr *čovjek*, and Ukr *ljudina*.

(11) passive

BU Xristijanskijat svjat ne može da **bădе izmenen**, ako otritnatite ne **bădat priobšteni** kăm nego.

GE Denn **man** kann das Gottesvolk nicht verändern, wenn **man** die Ausgeschlossenen nicht wieder integriert. <original>

'The people of God cannot be changed until the outcasts are restored to its body.'

The category "passive", as in (11), comprises periphrastic passive constructions with *be* and a past participle form of the verb, marked by the successors of the participial markings with **n/t*, which all Slavic languages possess. In some Slavic languages, such as Russian, only one auxiliary is used (e.g., Rus *bylo prikazano* 'was ordered'); in some other languages, such as Polish, two auxiliaries are used (e.g., Pol *willa była/została zbudowana* 'the villa was built'). Slavic languages are also known to have reflexive constructions formed with the verb in the 3rd person and the reflexive morpheme *-s'/sja*, *się*, or *se* (such as Pol *buduje się willę* 'the villa is built'). Such constructions are regarded by many linguists as passive. However, there is considerable disagreement in the literature with regard to their status as passives (Siewierska 1988). Due to their doubtful status, the decision was made to classify "reflexive passives" as reflexives, in this way ensuring consistent categorization across languages.

(12) participle

UK Zminiti Božij narod možna, liše **povernuvši** izgoïv nazad u jogo milo.

GE Denn man kann das Gottesvolk nicht verändern, wenn **man** die Ausgeschlossenen nicht wieder integriert. <original>

'The people of God cannot be changed until the outcasts are restored to its body.'

(13) 2sg

RU Kak ix tut **ponjuxaeš'**, tak i **poljubiš'**, vse odno—svoi oni ili čužie.

GE Wenn **man** sie da gerochen hat, dann liebt **man** sie, ganz gleich ob es die eignen oder fremde sind. <original>

'Once you've smelled them there, you love them whether they're your own or somebody else's.'

(14) indef(inite) pronoun

RU On snova obradovalsja, čto **kto-to** ešče ne zabyl, čto on xirurg!

GE Schon wieder war er erfreut, daß **man** noch nicht vergessen hatte, daß er Chirurg war!

CZ Už zase byl rád, že ještě někdo nezapomněl na to, že byl chirurgem! <original>

'Again he enjoyed the feeling that he had not been forgotten as a surgeon!'

(15) *-no/-to*

PL Nie **zamykano** go już na noc.

GE Zum Schlafen sperrte **man** ihn nicht mehr ein. <original>

'He was no longer locked in at bedtime.'

(16) lidé

UK **Ljudi** pritiskalis' bliže odne do odnogo.

GE Also rückte **man** näher zusammen. <original>

'So people huddled closer together.'

Similar to "člověk", the category "lidé" stands for all Slavic impersonally used nouns meaning 'people'. These are Rus/Ukr/Cr *ljudi*, Cz *lidé*, Blg *xorata*, and Pol *ludzie*.

(17) 2pl

CR Čim osjetite taj **miris**, **zavolite** ih, bila ona vaša ili tuda.

GE Wenn **man** sie da gerochen hat, dann liebt **man** sie, ganz gleich ob es die eignen oder fremde sind. <original>

'Once you've smelled them there, you love them whether they're your own or somebody else's.'

(18) refl/modal/infinitive

CR Eto, u jesen ima toliko toga što bi **se moglo donijeti**.

GE Es gibt doch im Herbst eine Menge Dinge, die **man** vorbeibringen könnte. <original>

'After all, in autumn there are lots of things someone could come by with.'

(19) modal/refl(exive)

BU **Može** da **se naduši** na păt i pod păt.

GE An jeder Straßenecke riecht **man** es. <original>

'You can smell it at every corner.'

(20) 3sg

RU No idti na risk tože ne **sledovalo**.

GE Andererseits durfte **man** aber auch nichts riskieren.

IT Ma non si poteva rischiare. <original>

'But he could run no risks.'

(21) adv(erbial construction)

PL „Nigdy nie **wiadomo**"—zakończył Wilhelm odprawiając go.

GE „Ja, **man** kann nie wissen", nickte William und entließ den jungen Studiosus.

IT "Non si sa mai," concluse Guglielmo accomiatandolo. <original>

'"You never can tell," William concluded, dismissing him.'

In quite a number of cases, none of the impersonalization strategies illustrated in (5–21) are used in Slavic. Instead, we find a complete rephrasing of the German *man*-sentence by using a personal reference or nominal, adjectival, prepositional, or other constructions, as in (22a–c):

(22) paraphrase

a. RU Prošlo čut' bol'še dvenadcati časov **posle naxodki** tela Venancija.

GE Es waren kaum mehr als zwölf Stunden vergangen, seit **man** Venantius' Leiche gefunden hatte.

IT Erano passate poco più di dodici ore da quando **si era scoperto** cadavere di Venanzio. <original>

'It had been just over twelve hours since the discovery of Venantius's corpse.'

b. PL **Po co** te wszystkie nowe ulice, […]?

GE Wozu brauchte **man** die vielen neuen Straßen, […]? <original>

'What was the need for all these new roads being dug up everywhere, and these new bridges?'

c. BU Tova ne be parfjum **kato parfjumite**.
GE Dieses Parfum war kein Parfum, wie **man** es bisher kannte.
<original>
'This perfume was not like any perfume known before.'

As shown in Figure 3 (on p. 133), a complete rephrasing is on average the third most frequent way of rendering a *man*-sentence in Slavic. Although paraphrase was included in Figure 3 as one of the means of rendering a *man*-construction, it does not belong to impersonalization strategies per se, and I will not go into further detail with respect to paraphrasing here.

Figure 3 highlights the most frequent means used for impersonalization in Slavic languages, shown by bars in decreasing order from left to right. The six most frequent strategies account for more than 75% of the cases[9] in Slavic based on the available corpus data. I will therefore refer to these strategies as the major Slavic impersonalization strategies. The major Slavic impersonalization strategies are the reflexive impersonal clitic or suffix; 3PL; the modal/infinitive construction; bare infinitive; and impersonally used nouns meaning 'human being' (Cz *člověk*, Rus *čelovek*, Pol *człowiek*, Blg *čovek*, Cr *čovjek*, Ukr *ljudina*). Although these impersonalization strategies, exemplified in (5–10), are commonly used in all Slavic languages, the languages do not behave uniformly and differ greatly in the extent to which they utilize a particular impersonalization strategy. While Figure 5 (on p. 142) represents the impersonalization strategies used in six Slavic languages on average in the whole dataset, it reveals no information about the distribution of these strategies in text types with different original languages and their translations or within each of the languages under study. Seeing as our data contain German, Italian, and Czech original texts, and the method of anchoring against *man* had to be done through the lens of translation, the following sections (3.2–3.4) will reveal how the impersonalization means (the actual number and the individual types) vary in three different types of texts. How the impersonalization strategies are used in each Slavic language will be discussed separately in Section 4.

3.2. Data Analysis of the German Original and Slavic Translations

This section presents the structures that Slavic languages use to convey the propositional content expressed by the German pronoun *man* in the translations of the German original text. The German polysemous impersonal pro-

[9] The six least frequent strategies—3SG constructions; *-no/-to* constructions; nouns meaning 'people' (Rus/Ukr/Cr *ljudi*, Cz *lidé*, Blg *xorata*, Pol *ludzie*); 2PL constructions; modal reflexive; and adverbial constructions—together account for less than 6% of the data. These constructions can be referred to as minor impersonalization devices.

noun *man* was taken as a filter to search for the corresponding *man*-equivalents in six Slavic languages.

The results are presented as cumulative frequencies in all six languages in Figure 4. As can be seen, sentences with the German impersonal pronoun *man* can be rendered in the Slavic languages by 18 different impersonalization strategies. Paraphrased sentences remain in the data but are not considered as impersonalization means since they lost their impersonal meaning. Modal reflexive constructions are not found in this part of the dataset, but there is a small number of constructions with modals (see Table 1 in the Appendix for a full breakdown of the data for each language, p. 166).

One of the striking differences between the results of the whole dataset (see Figure 3) and the German original part is the number of cases of impersonal uses of 1PL. While the first three major impersonalization strategies (reflexive, 3PL, and the modal/infinitive constructions) behave similarly, the 1PL is not very common in this part of the dataset. First-person plural is the third least frequent strategy and accounts for only 1% of all cases in the Slavic translations of the German original text, as can be seen in Figure 4. Another remarkable difference is the position of the 2SG strategy. This impersonalization strategy has slipped forward to the major impersonalization strategies, since it is used more frequently in the Slavic translations of the German text.

3.3. Data Analysis of the Czech Original and Translations

This section presents the Slavic impersonalization means which are used as *man*-equivalents in the Czech original text and its translations into five other Slavic languages. In this part of the dataset, visualized in Figure 5, the reflexive impersonal clitic, or postfix, is not the most frequent strategy, as it is in the overall data (see Figure 3 on p. 133). Rather, it presents the fourth most frequent strategy[10] and accounts for only 9% of all cases. Most striking and important for further analysis of the data is again the use of the 1PL, which is the third most frequent strategy, accounting for 11% of all cases. The nouns meaning 'human being' and 'people', e.g., Cz *člověk* and *lidé*, are also used more often in this part of the dataset and account for 9% of all cases (see Figure 5 on p. 142).

The use of 1PL presents a specific challenge in translated texts.[11] The fact that this part of the database includes German translations from the original Czech and the searches are carried out in translation potentially allows for a fully referential 1PL usage in the original Czech text, which could explain a higher percentage of the use of 1PL in this part of the data. To verify whether

[10] Paraphrase is not counted as a strategy here.

[11] I thank an anonymous reviewer for pointing this out.

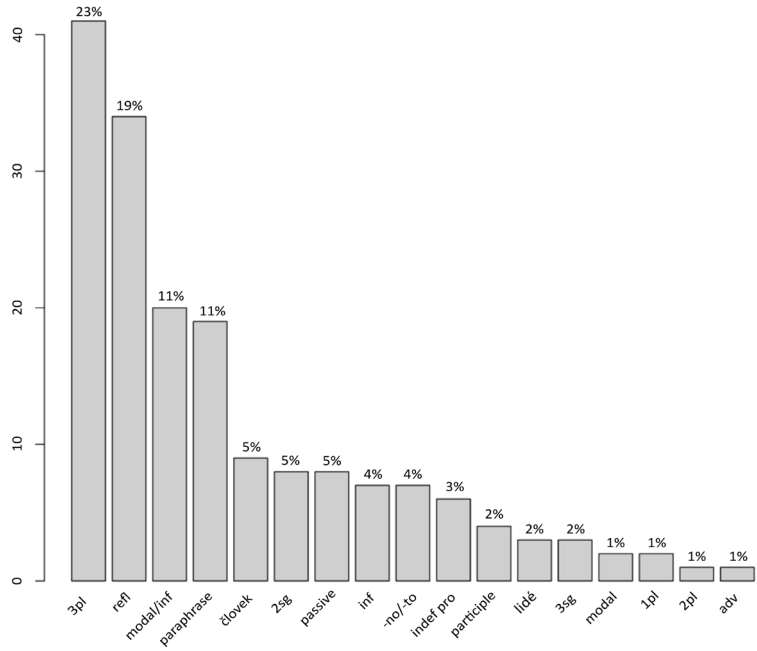


Figure 4. Slavic impersonalization means in the translations from German

the Czech 1PL constructions used in the original as corresponding to German *man* involve impersonal meanings, all 22 Czech sentences containing 1PL constructions were extracted (see Table 3 in the Appendix for a full breakdown of the data for each language, p. 168) and analyzed by a proficient speaker of Czech within a wider context of the original novel by Milan Kundera. Only one out of the 22 Czech examples was considered to be fully referential. This particular sentence with a personal 1PL construction is part of a dialogue and might refer to the speaker of the utterance and his companion (see example 7232 in the random dataset in the Appendix, p. 177). Due to its low frequency, this possible personal use of the 1PL construction remains in the data.

3.4. Data Analysis of the Italian Original and Slavic Translations

This section demonstrates the Slavic impersonalization means which are filtered through the German *man*-constructions in the German translations of the original Italian text. This set of data also reveals interesting contrasts with the results of the overall dataset. Compared to the whole dataset and subsets previously described, it has the least number of nouns meaning 'human

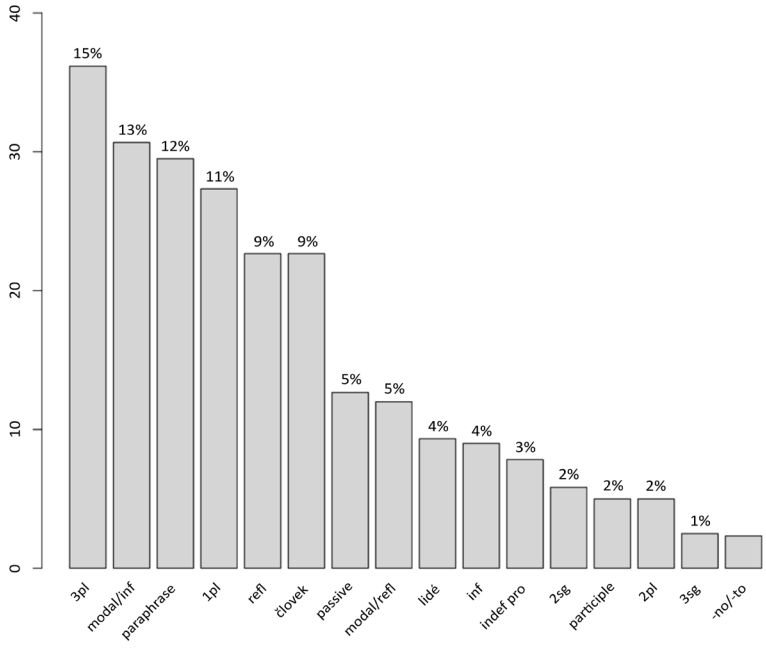


Figure 5. Slavic impersonalization means in the translations from Czech

being' and 'people', such as Cz *člověk* and *lidé*, accounting for only 3% of all numerous cases.

Additionally, 3PL is used insignificantly less often in this part of the dataset (see Figure 6 opposite), whereas the use of the reflexive is significantly higher (almost 20%) than in the whole dataset and other parts.

Similar to the verification of Czech impersonalization strategies, all Italian original sentences that were rendered in Slavic with the 1PL strategy were extracted (see Table 2 in the Appendix for a full breakdown of the data for each language) and analyzed by a native speaker of Italian within a wider context of the original novel by Umberto Eco to determine whether they contain impersonal meaning. Five out of 76 Italian examples (~6%) were considered to be possible in referential contexts.[12] However, some of these five Italian sentences without an impersonal meaning are still rendered with the impersonal 1PL in Slavic, as shown in the following Czech example (23):

[12] As mentioned in the previous section, a good example of a referential usage of 1PL can be found in the random dataset provided in the Appendix (see example 7232 on p. 177).

(23) CZ Zkrátka, každá kniha pro něho byla jako bájné zvíře, které **potkáme** v neznámé zemi.

GE Kurzum, jedes Buch war für ihn wie ein Fabelwesen, dem **man** in einem fremden Lande begegnet.

IT Insomma, ogni libro era per **lui** come un animale favoloso che **egli** incontrasse in una terra sconosciuta. <original>

'In short, for him every book was like a fabulous animal that he was meeting in a strange land.'

Due to this fact, a decision was made to leave all five Italian sentences without an impersonal meaning in the dataset.

4. Cross-Slavic Comparison of Impersonalization Strategies

This section reveals how the impersonalization strategies outlined in 3.1 are distributed in each Slavic language under study. This breakdown of data for

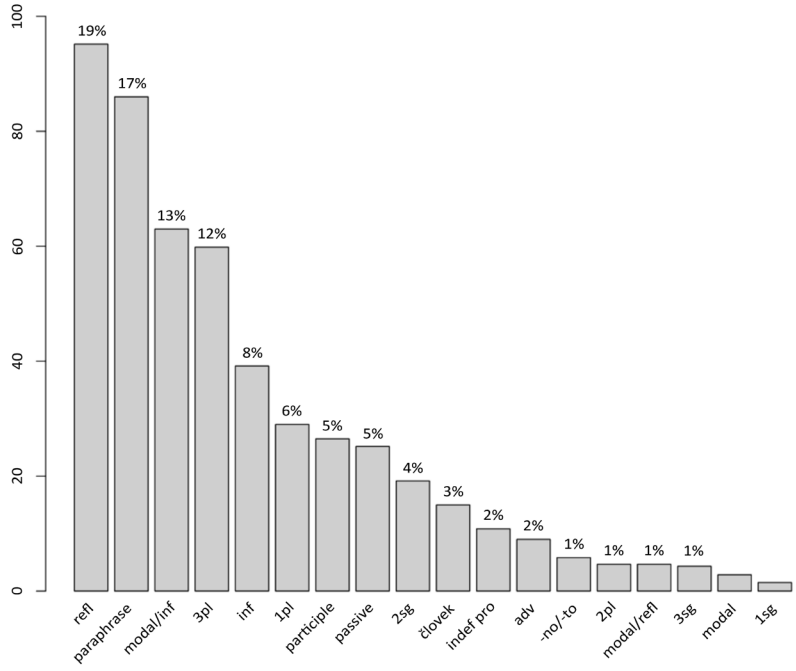


Figure 6. Slavic impersonalization means in the translations from Italian

each language allows us to see how diverse or similar Slavic languages behave in the domain of impersonalization.

Figure 7 (on pp. 145–48) shows how Slavic languages vary with regard to the distribution of impersonalization strategies. The frequencies of impersonalization devices in the six Slavic languages under study are represented in raw numbers. The diagrams in Figure 7 do not order the strategies in each of the Slavic languages by frequency, but rather keep the same sequence of strategies for a better cross-linguistic comparison of the data.

Comparing the frequencies of reflexives[13] in Figure 7, one notices that these constructions are by far the most frequent strategy in Croatian and Polish. In Bulgarian and Czech, reflexives are the second most frequent category after the 3rd-person plural and modal with infinitive constructions, respectively. The situation is, however, quite different in the East Slavic languages, where the reflexive is realized as the verbal affix *-sja*. Russian and Ukrainian make significantly less use of this impersonalization strategy in our data. Reflexives are the fifth most frequent strategy in East Slavic.[14]

We get almost the opposite picture in the case of 3rd-person plural constructions. Third-person plural constructions also reveal substantial differences in distribution across the six languages (see Siewierska and Papastathi 2011). While the 3PL impersonals are exceptionally frequent in Russian, Ukrainian, and Bulgarian, this strategy is less common in Croatian and scarce in Czech and Polish. According to Siewierska and Papastathi (2011), Russian employs 3PL impersonals five times as often as French and considerably more often than Spanish or Italian. Our data show that Russian employs 3PL impersonals twice as often as Czech, which might be attributed to the different properties of the 3PL impersonal in Russian and Czech (cf. Berger 1991). In Polish, 3PL impersonals appear to be less common and are used the least frequently of all the Slavic languages. This might be explained by the strong competition between three quasi-synonymous impersonalization constructions: the *-no/-to* construction, the reflexive impersonal *się*, and the 3PL impersonals (see the contribution by Prenner and Bunčić in this volume). The most frequent impersonalization strategies in Polish are reflexives (22%), closely

[13] For the purposes of this paper, this heterogeneous group of reflexive constructions has not been further differentiated. See Meyer 2010 for differences in the diachronic development of these constructions and their synchronic properties in Slavic.

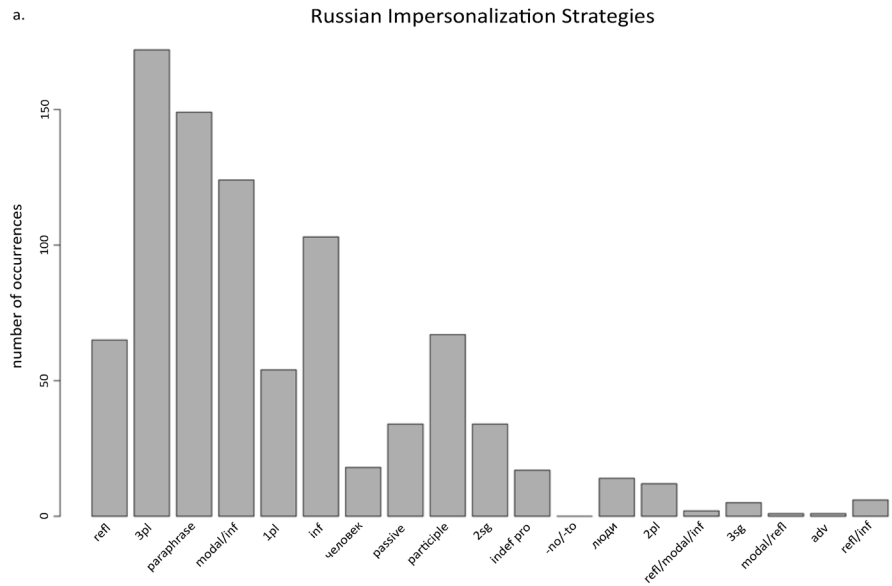
[14] One reviewer pointed out a conceivable explanation for the fact that reflexives are more frequent in Polish and Croatian than they are in Russian and Ukrainian. It is that Polish and some dialectal variants of Croatian distinguish two syntactic patterns of reflexives that in many contexts have the same meaning; one is the non-agreeing accusative construction, and the other is the agreeing nominative construction. Russian, however, does not have the non-agreeing reflexive accusative construction, which might explain why it less frequently employs reflexives in the translations—i.e., there is one less syntactic means to render impersonality than in Polish/Croatian.

followed by modal/infinitive constructions (18%), *-no/-to* (9%), and 3pl (6%). The same cannot be said for Ukrainian. Although the *-no/-to* construction is also said to be used in Ukrainian (Billings 1993), it is absent in the present data and is restricted to Polish.

A combination of modal predicative words with (bare) infinitives turns out to be a very widely used impersonalization strategy in four out of six Slavic languages. It is the second most frequent category in Russian, Polish, Ukrainian, and Czech. However, it is not widely used in Croatian and is completely absent in Bulgarian as it is known to lack the infinitive. There are some quasi-infinitive constructions—i.e., modal constructions with finite verbs—but these are not very numerous and account for less than 10% of Bulgarian strategies.

Some enlightening differences are found in the distribution of infinitives with an impersonal meaning. This type of construction seems to be quite heterogeneous within Slavic. Syntactically this type may involve bare infinitives, as in the Russian example (9); and embedded infinitives, as in Pol *słuchać* or *videć* 'to hear, to see'; or it may also contain infinitives combining with other constituents to form complex predicates. Our data show that infinitives are extensively used as an impersonalization device in Russian and Ukrainian. Infinitives are thus the fourth most frequent strategy in East Slavic. Czech and

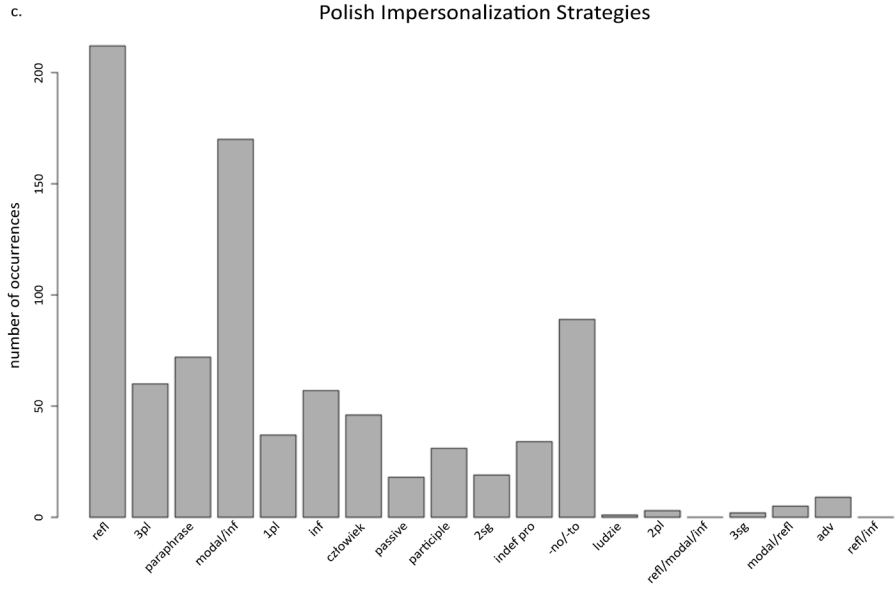
Figure 7. Impersonalization strategies in each Slavic language under study



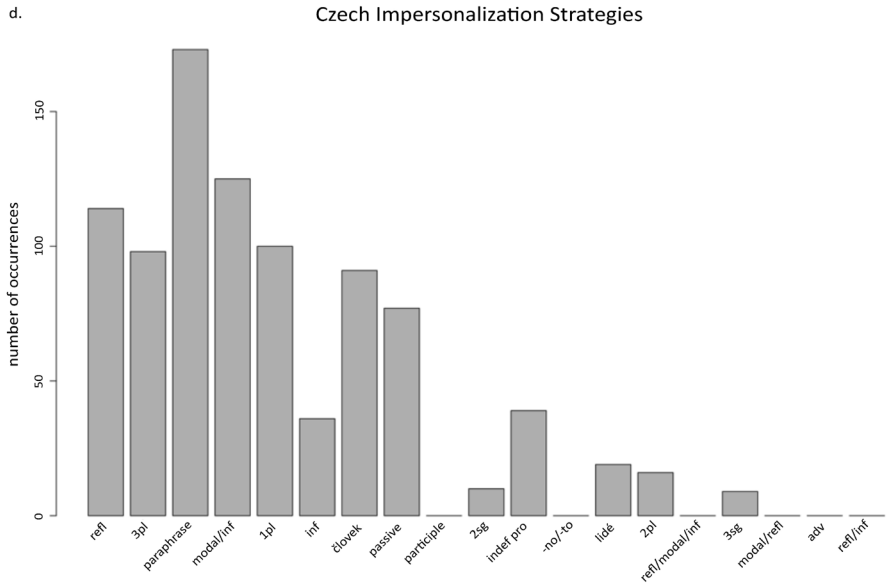
b.


Ukrainian Impersonalization Strategies
number of occurrences
0
50
100
150
refl
3pl
paraphrase
modal/inf
1pl
inf
людина
passive
participle
2sg
indef pro
-но/-то
люди
2pl
refl/modal/inf
3sg
modal/refl
adv
refl/inf

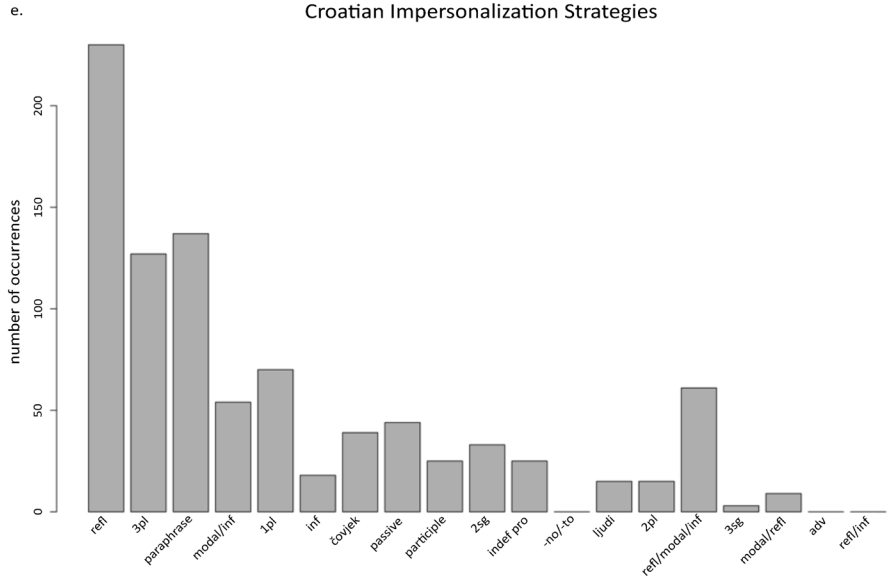
c.

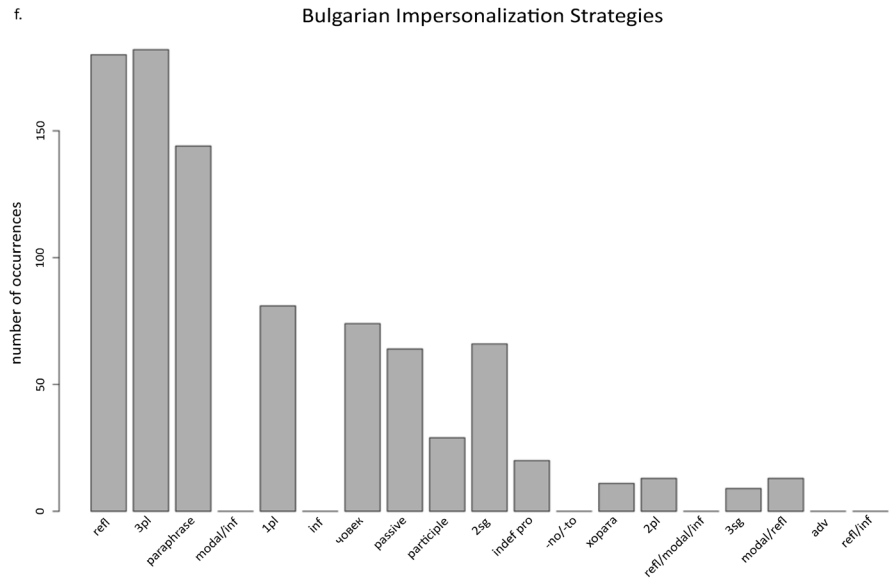

Polish Impersonalization Strategies
number of occurrences
0
50
100
150
200
refl
3pl
paraphrase
modal/inf
1pl
inf
człowiek
passive
partticiple
2sg
indef pro
-no/-to
ludzie
2pl
refl/modal/inf
3sg
modal/refl
adv
refl/inf

d.


Czech Impersonalization Strategies
number of occurrences
0
50
100
150
refl
3pl
paraphrase
modal/inf
1pl
inf
člověk
passive
participle
2sg
indef pro
-no/-to
lidé
2pl
refl/modal/inf
3sg
modal/refl
adv
refl/inf

e.


Croatian Impersonalization Strategies
number of occurrences
0
50
100
150
200
refl
3pl
paraphrase
modal/inf
1pl
inf
čovjek
passive
participle
2sg
indef pro
-no/-to
ljudi
2pl
refl/modal/inf
3sg
modal/refl
adv
refl/inf



Polish use infinitives with an impersonal meaning less frequently. In Croatian, infinitives as an impersonalization device are used very rarely. There are only 18 occurrences in the data (which approximates to 2% of all cases), and Bulgarian is well known to have no infinitive form.

Interestingly, 1PL has not yet been regarded as an impersonalization strategy in the literature. Our data show strong variation in the distribution of 1PL with impersonal meaning across Slavic. In Czech 1PL is used quite extensively with an impersonal meaning. Czech employs 1PL impersonals twice as often as Russian and almost three times as often as Ukrainian or Polish. Czech shows the most frequent use of 1PL with an impersonal meaning, followed by Bulgarian and Croatian.

The Slavic nouns meaning 'human being' (Cz *člověk*, Rus *čelovek*, Pol *człowiek*, Blg *čovek*, Cr *čovjek*, Ukr *ljudina*) have been largely ignored in their impersonal function in Slavic grammars and linguistic research. According to Giacalone Ramat and Sansò (2007), the distribution of impersonal nouns meaning 'human being' in the languages of Europe is consistent with the so-called Charlemagne area (in the sense of van der Auwera 1998: 823ff.); these nouns are widespread in German, Dutch, and French. However, their usage tends to diffuse eastwards to West and South Slavic languages, whereas East Slavic languages do not exhibit clear instances of these nouns (Giacalone Ramat and Sansò 2007: 66). This is not quite supported by the data in this study. Instead, the two East Slavic languages in our data clearly exhibit numerous

instances of Rus *čelovek* or Ukr *ljudina*[15] as equivalents of *man*-constructions in the dataset, as illustrated by (24), and behave like other Slavic languages in the distribution and usage of this strategy.

(24) GE […] das Ziel, das **man** verfolgt, bleibt immer verschleiert.
RU […] cel', kotoruju **čelovek** presleduet, vsegda skryta.
UK […] meta, do jakoï pragne **ljudina**, zavždi prixovana.
BU […] vsjaka cel, kojato presledva č**ovek**, e zabulena v măgla.
CR […] cilj kome **čovjek** teži uvijek je obavijen maglom.
PL […] cel, do którego **człowiek** dąży, jest zawsze niejasny.
CZ […] cil, za kterým se **člověk** žene, je vždycky zahalen. <original>
'… the goals we pursue are always veiled.'[16]

However, this impersonalization strategy is indeed used more frequently in other Slavic languages than it is in Russian and Ukrainian, according to our data. In contrast to the study by Rudolf (2014),[17] Rus *čelovek* (as well as Ukr *ljudina*) is not the least frequent impersonalization strategy in the present Russian (Ukrainian) data. It is true that this strategy is used much more frequently in Czech (see Figure 7d, p. 147), as can be seen in the following corpus example (25), where only the Czech text chooses the noun meaning 'human being' (Cz *člověk*) as a *man*-equivalent:

(25) GE […] auch wenn **man** diesen Teil nicht zahlt. <original>
RU […] daže esli **ne platit'** emu ètoj časti […]
UK […] navit' jakščo cju častinu **ne splačuješ**
BU […] Dori da **ne** go **plaštaš** […]
CR […] a da se taj dio i **ne plaća** […]
PL […] nawet jeśli **się nie płaci** […]
CZ […] i když ji **člověk** celou nezaplatí […]
'… even if you didn't pay…'

[15] It should be kept in mind that the singular and plural nouns are treated here as two separate strategies (see Figure 3, p. 133).

[16] This example nicely reveals how the impersonalization means of using the noun *člověk* in the original Czech is mirrored in all Slavic translations. This is considered a translation effect here (see Section 5). The sentences originally available in Cyrillic in the corpus (Russian, Ukrainian, and Bulgarian) were transliterated here by the author.

[17] In her study, Rudolf used the same method of anchoring against German *man* in the same corpus, but with a smaller sample of examples.

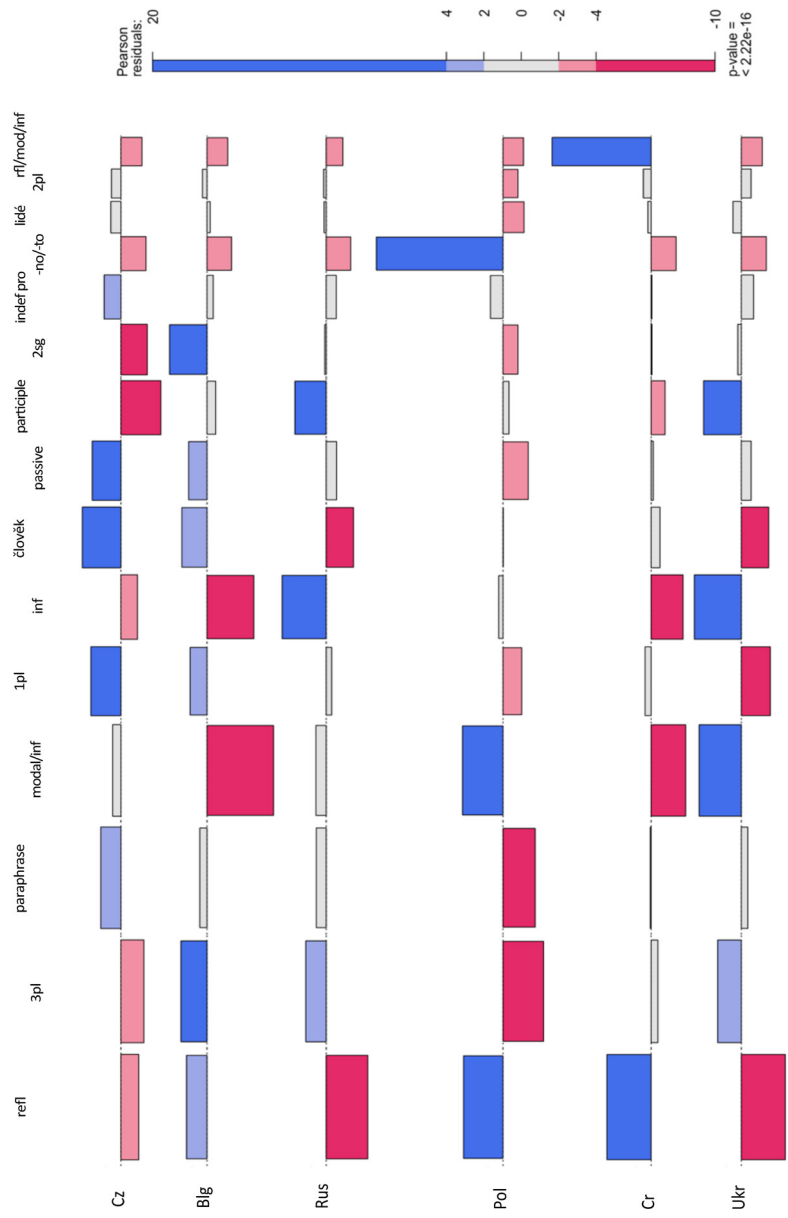


Figure 8. A comparison of the impersonalization strategies in six Slavic languages (Cohen-Friendly association plot)

In Bulgarian, Polish, and Croatian, impersonal nouns meaning 'human being' are used less frequently than in Czech but more often than in Russian or Ukrainian.

The aforementioned differences in the distribution of impersonalization strategies are illustrated by the association plot presented in Figure 8 on the opposite page. Figure 8 shows a comparison of the frequencies in the six languages in the form of a Cohen-Friendly association plot.[18] It shows the observed frequencies in relation to the expected frequencies, on the assumption of the statistical independence of the variables. For the bars that rise above the baseline, e.g., the impersonal reflexive in Polish or Croatian, the observed frequency is greater than expected. For the bars that fall below the baseline, e.g., the reflexives in Russian and Ukrainian, the observed frequency is smaller than expected. The height of each bar signifies the value of the corresponding Pearson residual, and the width stands for the squared root of the expected value. The shading color corresponds to the residual—large positive residuals are indicated by blue, large negative residuals by red—and intensity represents its relative importance: the more intense the color, the greater the deviation.[19]

The differences are thus visualized in the plot in Figure 8. The reflexives are strongly overrepresented in Polish and Croatian and massively underrepresented in Russian and Ukrainian.[20] The impersonal use of the 3PL strategy is greatly underrepresented in Polish and prominently overrepresented in Bulgarian. Modal infinitive constructions as well as infinitives are particularly overrepresented in Ukrainian and underrepresented in Bulgarian and Croatian, while the člověk strategy is enormously underrepresented in Rus-

[18] To create an association plot, I have used the *assoc()* function from the package *vcd*.

[19] For a more detailed explanation of an association plot in R, see Levshina 2015: 220.

[20] This speaks to a fundamental difference in the behavior of reflexives across Slavic. Marelj (2004) points out, for example, a syntactic divergence in reflexives in Slavic languages. See examples (i) and (ii) below and the observation about syntactic patterns provided in footnote 14 (p. 144).

(i) Owe przesąda dzisiaj inaczej się (Polish)
these prejudices$_{ACC}$ today differently REFL
interpretuje/interpretują.
interpret$_{3SG}$/interpret$_{3PL}$
'One interprets these prejudices differently today.'

(ii) Roditelje/Roditelji se poštuje/poštuju. (Croatian)
parents$_{ACC}$ REFL respect$_{3SG}$
'One should respect parents.' (Marelj 2004: 267–69)

sian. All these differences in the distribution of the major impersonalization strategies appear to be highly significant ($p < 0.001$).[21]

The association plot in Figure 8 also shows some significant differences in the distribution of other non-major impersonalization strategies such as passive, *-no/-to*, and 2nd-person singular constructions. While passives are overrepresented in Czech, participial and 2nd-person singular constructions are greatly underrepresented. Constructions with *-no/-to* occur exclusively in the Polish data, and the 2nd-person singular constructions occur most frequently in Bulgarian, followed by Ukrainian and Russian.

The *-no/-to* construction is also known to be used in Ukrainian (Billings 1993), but we do not find it in our data. This may be due to the fact that the Ukrainian construction, but not the Polish one, can be used in both agentive (26) and non-agentive (27) contexts, and thus has a broader reference in terms of the external argument than *man*-constructions.

(26) Tabir bulo **zajnjato** amerykans'kym vijs'kom. (Ukrainian)
camp AUX$_{3SG.N}$ occupied American troops
'The camp was occupied by American troops.'

(27) Bereh **rozmyto** tečijeju.
shore washed.away current
'The shore was washed away due to the current.' (Lavine 2017: 190)[22]

Modern Slavic languages fall into three major groups, according to linguistic and historical factors (Sussex and Cubberley 2006). Considering the languages under study, there is South Slavic, which includes Bulgarian and Croatian; East Slavic, which includes Russian and Ukrainian; and West Slavic, which includes Polish and Czech. The question arises whether the languages fall into these three groups and behave similarly within the Slavic subfamilies in the domain of impersonalization. Observing the data on the distribution of various impersonalization strategies in six Slavic languages, I have identified some group tendencies in the use of impersonal strategies. While the East Slavic languages use the impersonal reflexive or the člověk strategy less frequently than other languages, the West Slavic languages do not use the 3PL strategy as often as the other languages do. In order to compare the distribution of the impersonalization strategies among the three Slavic groups, one needs to show that the six languages actually divide up into these three

[21] If a standardized residual value is greater than 1.96 or smaller than –1.96, the cell makes a statistically significant contribution to the obtained χ^2-statistic value at the significance level of 0.05 (Levshina 2015: 221).

[22] I thank an anonymous reviewer for this point.

groups by verifying that the differences inside the groups are smaller than the differences between the groups.[23] To test this, I first consider the impersonalization strategies within each group—that is, between Russian and Ukrainian (for East Slavic), Bulgarian and Croatian (for South Slavic), and between Czech and Polish (for West Slavic)—and formulate the null hypothesis that the proportions of the impersonalization strategies are equal between the languages. Three Fisher's exact tests[24] show that the null hypothesis can be rejected for the East (p-value = 0.005497), West (p-value = 0.0004998), and South Slavic group (p-value = 0.0004998), since the p-value is smaller than the conventional level of 0.01. This means that the differences in the distribution of impersonalization strategies within the groups are highly significant and the six Slavic languages do not divide up into three groups in the domain of impersonalization, since the differences within the three groups are not smaller than the differences between groups (p-value < 0.01). Despite the similarities observed, the statistical tests show that there are significant cross-Slavic differences between all six languages and the languages that belong to the same Slavic group do not utilize impersonalization strategies in a similar way.

There are, of course, some similarities which may carefully be interpreted as tendencies within particular Slavic subfamilies in the domain of impersonalization. Specifically, we note a comparable distribution of several strategies within East Slavic. These similarities can be illustrated by a neighbor net, shown in Figure 9 on the following page. The neighbor net in Figure 9 shows that there are indeed East and South Slavic groups, the West Slavic languages do not form any cluster, and it is completely unclear whether the differences between the groups are greater and statistically more significant than in-group differences.

5. Translation Effects

It can be assumed that the distribution of impersonalization strategies is largely impacted by the source language of the translation, e.g., the original language from which the sentences have been translated into German and Slavic in our database, namely German, Italian, or Czech (see Section 2.2 for the description of the data). To test this assumption, I consider the impersonalization strategies in just two different Slavic languages due to space constraints (namely, Czech and Russian) and formulate the following hypotheses (H1 and H0) concerning the translation effects:

[23] I thank another anonymous reviewer for pointing this out.

[24] Since several values in the data are smaller than 5, Pearson's χ^2-test is reported to be not robust enough in such cases. I therefore use an additional recommended test, Fisher's exact test (Levshina 2015: 29), in all three cases.

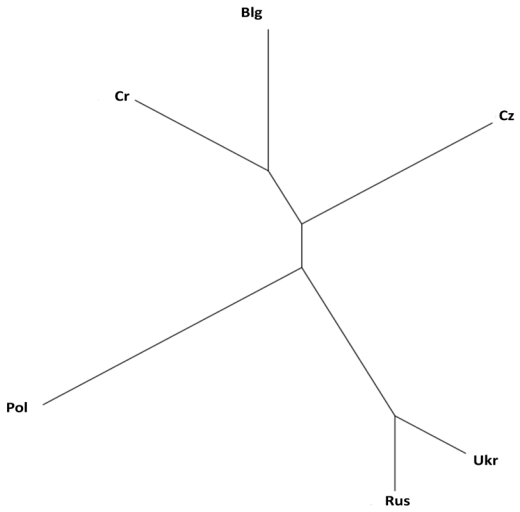


Figure 9. Cross-Slavic comparison of impersonalization strategies. Phylogenetic tree via UPGMA method, unrooted.

H1: The source language has an impact on the choice of impersonalization strategy found in Russian or Czech.

H0: The source language has no impact on the choice of impersonalization strategy found in Russian or Czech.

Figure 10 on the opposite page visualizes the data from Russian in the form of a Cohen-Friendly association plot. A χ^2-test shows that the null hypothesis can be rejected ($\chi^2 = 127.23$, $df = 28$; $p < 0.001$).[25]

Figure 10 shows that the main differences in the Russian corpus data consist in the distribution of five impersonalization strategies as well as the option of paraphrasing the whole sentence. The use of 3PL impersonals in Russian is strongly overrepresented in the sentences translated from German, whereas the use of 1PL as well as Rus *čelovek* and *ljudi* is massively overrepresented in

[25] As mentioned in footnote 24, the χ^2-test is said not to be robust when at least one expected value in the table is smaller than 5. In the Russian data there were no instances of adverbial constructions in the sentences translated from Czech. I therefore used Fisher's exact test (Levshina 2015: 29). Fisher's exact test for count data with simulated *p*-value (based on 2,000 replicates) also shows that the null hypothesis can be rejected: *p*-value = 0.0004998.

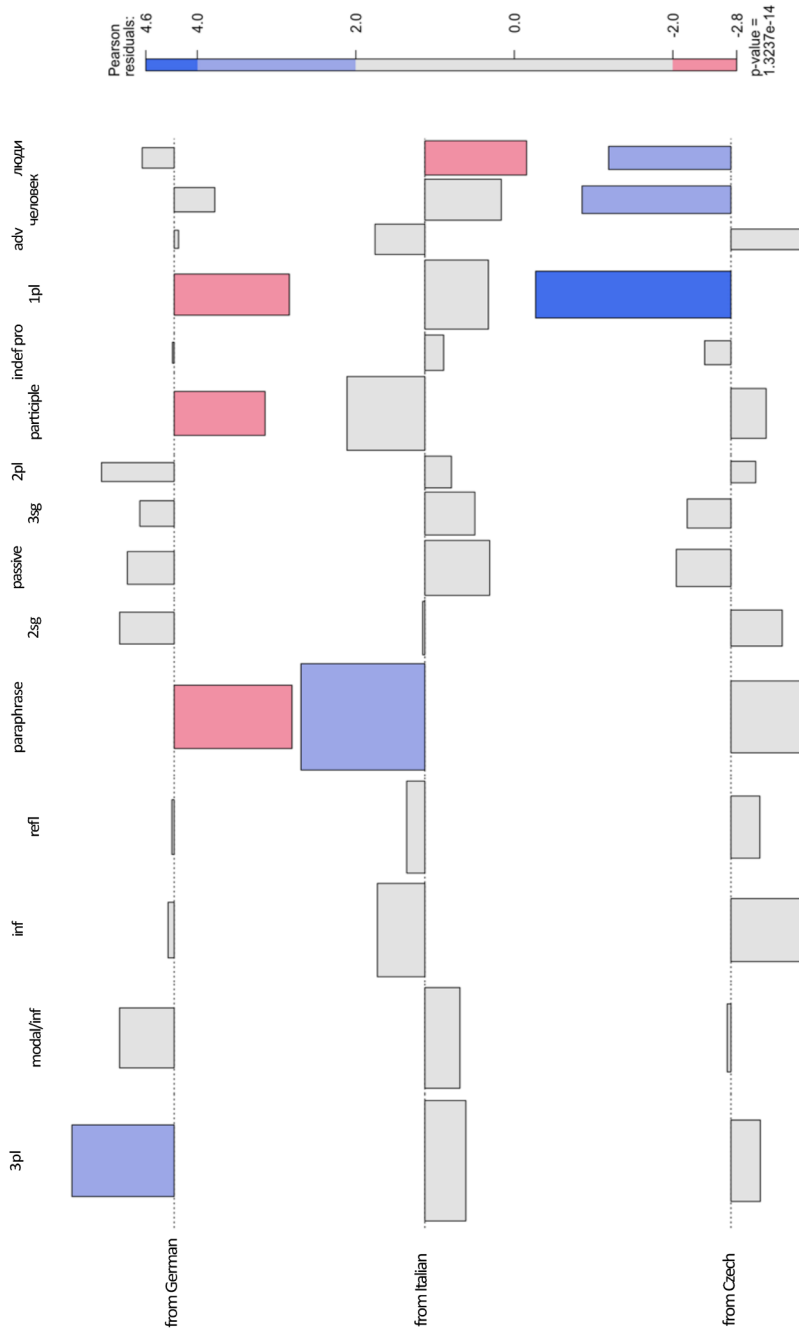


Figure 10. Translation effects in Russian

the sentences translated from Czech.[26] This influence from the German or Czech source language is not very surprising and can be attributed to the prevalence of the *man*-sentences in German, which are usually rendered by 3PL in Russian, and the frequent usage of the 1PL and *člověk* (as well as *lidé*) in original Czech sentences, which was attested in the Czech data as well as Czech translations (see Table 4 in the Appendix).

Figure 11 on the opposite page presents the data from Czech in the form of a Cohen-Friendly association plot. A χ^2-test shows that the null hypothesis can also be rejected ($\chi^2 = 165.94$, $df = 26$; $p < 0.001$).[27] As the bars rise or fall with respect to the baseline in Figure 11, they reveal that the main differences in the Czech data concern seven impersonalization strategies (reflexive impersonals, 1PL, 3PL, noun *člověk*, infinitive constructions, noun *lidé*, and 2PL) as well as paraphrase.

Figure 11 reveals that the choice of impersonalization strategy in Czech translation is determined by the source language. The main Czech strategies vary according to whether a sentence is translated from German or Italian. We see that Czech translations from German and from Italian choose reflexive as an impersonalization strategy more often than the original. In the German translations, the člověk strategy is strongly overrepresented. In Italian translations, the 3PL strategy is strongly underrepresented.

In both cases (Figures 10 and 11), statistical analysis reveals significant translation effects, which leads to the conclusion that the distribution of impersonalization strategies is heavily influenced by the source language.

6. Discussion of the Data

One of the major aims of this study is to exemplify the types of structures Slavic languages use to render propositional content expressed by the German pronoun *man*. By using a method based on German *man*, I found that the Slavic languages use 18 linguistic means to express impersonal meaning. I do not claim that all the strategies revealed by the data in this study are impersonal constructions. I rather suggest that the structures presented here (as shown in Figures 3–6) are used as impersonalization strategies in particular contexts in the six selected Slavic languages.

[26] I also notice this influence from Czech in other Slavic languages in the corpus data. The Slavic sentences translated from Czech tend to have a more frequent use of 1PL and nouns meaning 'human being' (see Table 3 in the Appendix, p. 168).

[27] Here it must again be mentioned, as in the Russian case above, that the χ^2-test is not robust here. In the Czech data there were no instances of *lidé* in the sentences translated from Italian, and therefore Fisher's exact test was used. Fisher's exact test also shows that the null hypothesis can be rejected: p-value = 0.0004998.

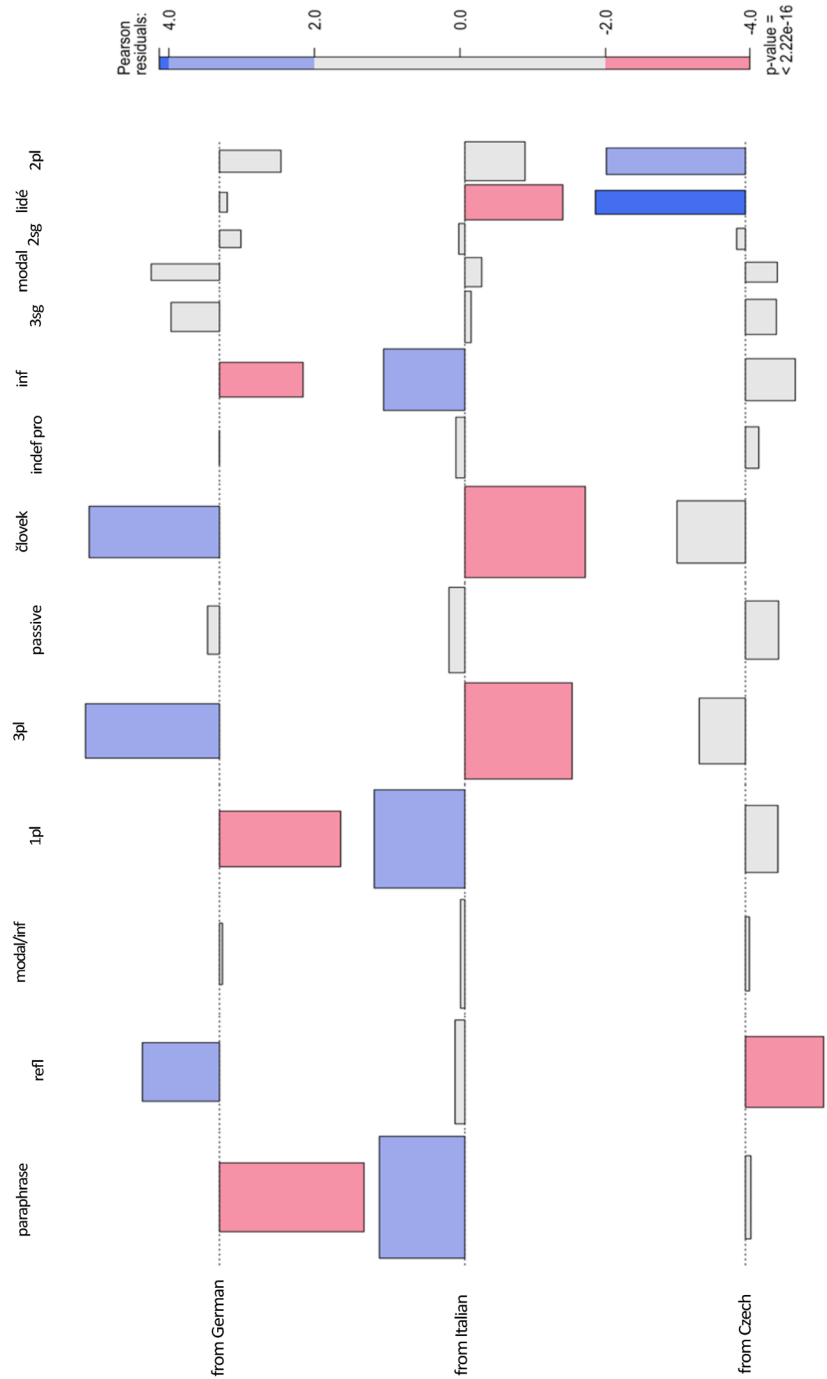


Figure 11. Translation effects in Czech

Since the analysis presented in this paper is based on parallel translation data, the corpus data is interpreted with diligence.

Two important methodological issues should be discussed here. First, using *man* as a filter has specific weaknesses. As already mentioned in Section 3.3, this approach neglects the other strategies used in German to express impersonalization. Although *man* is the most widely used impersonalization strategy in German, there are also other means to render impersonal meaning, e.g., the impersonal passive (Primus 2011). Therefore, to supplement the parallel corpus, my future study will take random samples of some of the Slavic strategies discovered here in a comparable corpus of original texts in the same language and investigate how much has been missed by using *man* as a filter.

One of the advantages of filtering through German *man* is that it arguably maximizes precision in large amounts of data, since all the occurrences are clearly impersonal. Moreover, anchoring against *man* allows us to identify means of expressing impersonalization that have not been identified heretofore or have been paid little attention in the literature, as for example, the Slavic strategies of using 1PL and nouns meaning 'human being' or 'people' in impersonal contexts.

The fact that two texts included in the database are German translations from a third language, namely, Czech and Italian, raises another important methodological issue regarding whether the German *man* presents an adequate anchor in translated texts. That is, this method has a possible limitation in that examples extracted from German *man* in Italian or Czech may not necessarily involve impersonalization strategies. To ensure that the correspondences of German *man* in the original languages do involve impersonal meaning, Italian and Czech examples were explored in detail within a wider context (see Sections 3.2–3.4 for details). It was found that the overwhelming majority of these constructions is unambiguously impersonal.

It turns out that working with translated texts revealed strong translation effects, which comparative linguists should be aware of (von Waldenfels 2012). Data show that the Slavic translations reflect the structures of the original in the domain of impersonalization (see example (23)). Thus, the frequent use of impersonal reflexives in Slavic translations from the Italian text can be explained by the numerous usages of the impersonal reflexive in Italian. The available data also show that Czech prefers the use of nouns meaning 'human being' and 'people' (Cz *člověk*, *lidé*) in impersonal contexts and uses the strategy more often than other Slavic languages do in the translations of German or Italian text (see Tables 1 and 2 in the Appendix, pp. 166–67). Translation effects are clearly seen in the Slavic translations from Czech, where the Slavic languages are influenced by the Czech structure of the impersonalization strategy and choose to use a similar one (e.g., Rus *čelovek*, *ljudi*; Pol *człowiek*, *ludzie*; Blg *čovek*, *xorata*; Cr *čovjek*, *ljudi*; and Ukr *ljudina*, *ljudi*).

Besides translation effects, there is another point to be raised here with regard to translated texts. As a comparative linguist, I was not interested in the translation process as such or in the translation-specific characteristics of the data contained in this parallel corpus. Nevertheless, a straightforward equivalence between original and translated texts is not assumed here (von Waldenfels 2012). Translation is known to be closely connected with stylistics (Boase-Beier 2019), and some impersonalization strategies exemplified in this study might have been chosen by the authors and translators for various stylistic purposes, which are not dealt with in this study. Moreover, the native Italian speakers who assisted with the annotation of the data reported an extensive use of the impersonal reflexive in the Italian novel by Umberto Eco, and translators have previously encountered challenges in translating Eco's texts (Dixon 2016). Important for this study was the case of 1pl corresponding to German *man*-constructions, which could have been used as fully referential in Slavic translations. Our analysis of the data in Section 3 reveals, however, that the 1pl is indeed used as an impersonalization strategy in Czech, as well as other Slavic languages, as an equivalent of the German *man*-construction.

The general point made in this paper is that using *man* as a filter proves to be a valid approach even in translated texts. The potential false positive results, such as a probable usage of a fully referential 1pl as an artifact of the translation-driven approach, have been validated by looking at the individual data points and kept to a minimum. The data analyzed here, however, reveal strong translations effects.

7. Summary and Conclusion

This paper has employed a comparative corpus-driven approach to identify the types of structures Slavic languages use to express propositional content conveyed by the pronoun *man* in German and to compare these impersonalization means across Slavic. The overview of specific types of strategies used for impersonalization in Slavic is given in Section 3. The corpus data from Russian, Ukrainian, Bulgarian, Croatian, Czech, and Polish reveal that Slavic languages possess a large variety of impersonalization strategies, represented by 18 constructions: reflexives, 3pl constructions, modal infinitive constructions, 1pl constructions, infinitive constructions, impersonally used nouns meaning 'human being' (Cz *člověk*, Rus *čelovek*, Pol *człowiek*, Blg *čovek*, Cr *čovjek*, Ukr *ljudina*) and 'people' (*ljudi*, *lidé*, *xorata*, *ludzie*), participial constructions, passives, 2sg constructions, indefinite pronouns, *-no/-to* constructions, 2pl constructions, reflexive modal infinitive constructions, modal reflexive constructions, 3sg constructions, adverbial constructions, and reflexive infinitive constructions. The distribution of these impersonalization strategies across Slavic languages is illustrated in Figure 3. An alternative option of rendering a German *man*-sentence is a complete rephrasing by resolving the impersonal

meaning and using either a personal pronoun or a nominal, adjectival, prepositional, or other construction. This option has not been dealt with in much detail in the current study. Reflexives turn out to be the most frequent cross-Slavic impersonalization strategy, followed by the 3PL constructions. Infinitives and impersonally used nouns meaning 'human being' reveal a great deal of variation in distribution across the six Slavic languages under study. One of the most surprising results of the study is the impersonal use of 1PL constructions in Slavic languages. The use of 1PL has not yet been described as an impersonal construction in the literature. Our data and analysis convincingly show that 1PL constructions are used as an impersonalization means, mostly in Czech but also in other Slavic languages, to convey propositional content expressed by the pronoun *man* in German. The contexts in which impersonal 1PL in Slavic are used represent a clear need for future study.

On the one hand, this study shows a range of expressions which are used in Slavic to express impersonalization; on the other hand, it reveals cross-Slavic variation in the distribution and use of these impersonalization strategies (Section 4). The domain of impersonalization is clearly not homogeneous across the Slavic languages. Interestingly, the six Slavic languages under study do not divide up into the typical West, East, and South Slavic subfamilies in their distribution of the impersonalization strategies. Statistical analysis reveals significant differences in the domain of impersonalization between Slavic languages. While some similarities in the distribution of particular strategies are detected, for example, within the East Slavic group (e.g., the frequent use of 3PL or infinitive constructions), the data also reveal some noteworthy differences between Ukrainian and Russian, such as a more frequent use of modal infinitive constructions in Ukrainian.

The study has also shown that the distribution of impersonalization strategies is greatly influenced by the original language of translation, i.e., the question of whether a sentence in the corpus data is a German original or translation from Italian or Czech. It turns out that the source language is highly relevant for the choice of the impersonalization strategy.

Other factors are also known to influence the choice of impersonalization strategies. Therefore, more data from different registers and the analysis of the impact of grammatical, semantic, and pragmatic parameters on the choice of impersonalization devices in Slavic are desirable in future studies. As suggested by von Waldenfelds (2012: 265), "assessments based on a corpus such as ParaSol need to be examined critically in the light of independent monolingual corpora". To control for the obvious translation effects that have been revealed by this study, the present study should be combined in the future with the analysis of comparable corpora, i.e., the collection of original texts in the languages compared.

Sources

ParaSol, a corpus of Slavic and other languages. (2006–) Ruprecht von Waldenfels and Roland Meyer. Available at: http://parasolcorpus.org/Ursynow/. Last accessed 11 May 2020.

References

Anochina, Svetlana P. (1981) "Neopredelënno-ličnye predloženija s podležaščim-mestoimeniem *man* v sovremennom nemeckom jazyke i sootvetstvujuščie im neopredelënno-ličnye predloženija sovremennogo russkogo jazyka". I. L. Klimovič and E. A. Fedotova, eds. *Problemy sopostovitel'noj tipologii rodnogo (russkogo) i inostrannyx jazykov.* Leningrad: RIS LGPI, 3–12.

Berger, Tilman. (1991) "Zum denotativen Status ‚unbestimmt-persönlicher' Konstruktionen im Tschechischen". Maciej Grochowski and Daniel Weiss, eds. *Words are physicians for an ailing mind: Ὀργῆς νοσούσης εἰσὶν ἰατροὶ λόγοι.* Munich: Sagner, 71–81.

Billings, Loren A. (1993) "A note on expletives in Ukrainian *-no/-to* passives". *Slavic syntax newsletter* 3(1)*:* 1–8.

Boase-Beier, Jean. (2019) *Translation and style.* 2nd edition. London: Routledge. [Translation Theories Explored.] DOI 10.4324/9780429327322.

Bunčić, Daniel. (2018) "Impersonal constructions in Slavic languages and the agentivity of the verb". Sebastian Kempgen, Monika Wingender, and Norbert Franz, eds. *Deutsche Beiträge zum 16. Internationalen Slavistenkongress, Belgrad 2018*. Wiesbaden: Harrassowitz, 101–10.

————. (2019) "Agent prominence in the Polish *-no/-to* construction". Anastasia Bauer and Daniel Bunčić, eds. *Linguistische Beiträge zur Slavistik.* Vol. 24. *JungslavistInnen-Treffen in Köln, 17–19 September 2015.* Frankfurt am Main: Peter Lang, 63–76.

————. (2020) "Agent prominence and movement in Russian third person plural impersonals ('indefinite-personal sentences')". *Die Welt der Slaven* 65(2): 297–310.

Cabredo Hofherr, Patricia. (2017) "Impersonal passives". Martin Everaert and Henk C. Riemsdijk, eds. *The Wiley Blackwell companion to syntax*. 2nd edition. DOI 10.1002/9781118358733.wbsyncom031.

Divjak, Dagmar, Serge Sharoff, and Tomaž Erjavec. (2017) "Slavic corpus and computational linguistics". *Journal of Slavic linguistics* 25(2): 171–99. DOI 10.1353/jsl.2017.0008.

Dixon, Richard. (2016) "Playing on words: Challenges in translating Umberto Eco's *Numero zero*". *Signata* 7: 377–90.

Dušková, Libuše. (1973) "*Man*-Sätze in Czech and English". *Philologica Pragensia* 16: 5–37.

Gast, Volker. (2015) "On the use of translation corpora in contrastive linguistics: A case study of impersonalization in English and German". Bart Defrancq, ed. "Contrasting contrastive approaches". Special issue, *Languages in contrast* 15(1): 4–33.

Gast, Volker and Johan van der Auwera. (2013) "Towards a distributional typology of human impersonal pronouns, based on data from European languages". Dik Bakker and Martin Haspelmath, eds. *Languages across boundaries: Studies in memory of Anna Siewierska.* Berlin: de Gruyter Mouton, 119–58.

Gast, Volker, Lisa Deringer, Florian Haas, and Olga Rudolf. (2015) "Impersonal uses of the second person singular: A pragmatic analysis of generalization and empathy effects". Hogeweg Lotte and Helen de Hoop, eds. "The flexibility of pronoun reference in context". Special issue, *Journal of pragmatics* 88: 148–62.

Giacalone Ramat, A. and A. Sansò. (2007) "The indefinite usage of *uomo* ('man') in early Italo-Romance: Grammaticalization and areality". *Archivio glottologico italiano* 92(1)*:* 65–111.

Guiraud-Weber, Marguerite and Irina Kor Chahine. (2013) "Impersonals and beyond in Slavic". Irina Kor Chahine, ed. *Current studies in Slavic linguistics.* Amsterdam: John Benjamins, 1–22. [Studies in Language Companion Series, 146.]

Hansen, Björn. (2005) "How to measure areal convergence: A case study of contact-induced grammaticalization in the German-Hungarian-Slavonic contact area". Björn Hansen and Petr Karlík, eds. *Modality in Slavonic languages: New perspectives.* Munich: Sagner, 219–37.

Ivanova, Elena Ja. and Alla A. Gradinarova. (2015) *Sintaksičeskaja sistema bolgarskogo jazyka na fone russkogo.* Moscow: Jazyki slavjanskoj kul'tury.

Kimmelman, Vadim. (2018) "Impersonal reference in Russian Sign Language (RSL)". *Sign language and linguistics* 21(2): 204–31.

Kitagawa, Chisato and Adrienne Lehrer. (1990) "Impersonal uses of personal pronouns". *Journal of pragmatics* 14: 739–59.

Lavine, James E. (2017) "Syntactic change and the rise of transitivity: The case of the Polish and Ukrainian *-no/-to* construction". *Studies in Polish linguistics* 12(3): 173–98. DOI 10.4467/23005920SPL17.009.7201.

Levshina, Natalia. (2015) *How to do linguistics with R: Data exploration and statistical analysis.* Amsterdam and Philadelphia: John Benjamins.

Malchukov, Andrej and Akio Ogawa. (2011) "Towards a typology of impersonal constructions: A semantic map approach". Andrej Malchukov and Anna Siewierska, eds. *Impersonal constructions: A cross-linguistic perspective.* Amsterdam: John Benjamins, 19–56.

Malchukov, Andrej and Anna Siewierska. (2011) *Impersonal constructions: A cross-linguistic perspective*. Amsterdam: John Benjamins. [Studies in Language Companion Series, 124].

Marelj, Marijana. (2004) *Middles and argument structure across languages*. Utrecht: LOT Publishing.

Mazzitelli, Lidia Federica. (2019) "Referential and pragmatic-discourse properties of Lithuanian reference impersonals: 2sg-imp, 3-imp and ma/ta-imp". *Kalbotyra* 72: 32–57.

Mel'čuk, Igor. (1974) *Opyt teorii lingvističeskix modelej "Smysl—Tekst": Semantika, sintaksis*. Moscow: Nauka.

Meyer, Roland. (2010) "Reflexive passives and impersonals in North Slavonic languages: A diachronic view". *Russian linguistics* 34: 285–306. DOI 10.1007/s11185-010-9063-6.

Nikitina, Elena N. (2011) "Neopredelenno-ličnye predloženija". D. V. Sičinava, *Časti reči: Materialy dlja proekta korpusnogo opisanija russkoj grammatiki*. Available at: http://rusgram.ru/Неопределенно-личные_предложения. Last accessed 9 December 2021.

Padučeva, Elena V. (2012) "Neopredelenno-ličnoe predloženie i ego podrazumevaemyj sub"ekt". *Voprosy jazykoznanija* 1: 27–41.

Plungjan, Vladimir A. (2016) *Vvedenie v grammatičeskuju semantiku: Grammatičeskie značenija i grammatičeskie sistemy jazykov mira.* Moscow: RGGU.

Prenner, Maria K. (forthcoming) *Agentivity in impersonal constructions in Polish and Russian.* Ph.D. thesis, University of Cologne.

Prenner, Maria Katarzyna and Daniel Bunčić. (this volume) "The competition of arb constructions in Polish". *Journal of Slavic linguistics* 29(2): 201–20.

Primus, Beatrice. (2011) "Animacy and telicity: Semantic constraints on impersonal passives". *Lingua* 121(1): 80–99.

R Core Team. (2015) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rivero, María Luisa and Milena Milojevic Sheppard. (2003) "Indefinite reflexive clitics in Slavic: Polish and Slovenian". *Natural language & linguistic theory* 21(1): 89–155. DOI 10.1023/A:1021841517604.

Rudolf, Olga. (2014) "A corpus-based study of human impersonal constructions in Russian". Silwester Jaworski and Jacek Witkoś, eds. *New insights into Slavic linguistics*. Frankfurt: Lang, 327–40.

Schlund, Katrin. (2018a) "A unifying approach to Russian impersonal constructions". *Zeitschrift für Slawistik* 63(1): 120–68.

————. (2018b) *The Russian 'Elemental Construction' from a synchronic and diachronic perspective: An empirical contribution to the understanding of an impersonal construction in Slavic*. Post-doctoral thesis, University of Heidelberg.

————. (2020) "Active transitive impersonals in Slavic and beyond: A parallel corpus analysis". *Russian linguistics* 44(1): 39–58.

Siewierska, Anna. (1988) "The passive in Slavic". Masayoshi Shibatani, ed. *Passive and voice*. Amsterdam: John Benjamins, 243–90.

————. (2008) "Ways of impersonalizing: Pronominal vs. verbal strategies". María de los Ángeles Gómez González, J. Lachlan Mackenzie, and Elsa González Álvarez, eds. *Current trends in contrastive linguistics: Functional and cognitive perspectives.* Amsterdam: John Benjamins, 3–26.

Siewierska, Anna and Maria Papastathi. (2011) "Towards a typology of third person plural impersonals". *Linguistics* 49(3): 575–610.

Sussex, Roland and Paul Cubberley. (2006) *The Slavic languages*. Cambridge: Cambridge University Press.

Szucsich, Luka. (2007) "Nothing wrong with finite T: Non-agreeing accusative impersonal sentences". Magdalena Goledzinowska, ed. *Proceedings of Formal Approaches to Slavic Linguistics 15*. Ann Arbor: Michigan Slavic Publications, 401–19.

Šaxmatov, Aleksej A. (2001) *Sintaksis russkogo jazyka.* Moscow: URSS.

Švedova, Natalja Ju., ed. (1980) *Russkaja grammatika.* Moscow: Nauka.

Uhlik, Mladen and Andreja Žele. (2018) "Bresosebne zdradbe v Slovenšči: Kontrastiva z drugimi južnoslovanskimi jeziki in Ruščino". *Jezikoslovni zapiski* 24: 99–112.

Valgina, Natalija. (2003) *Teorija teksta*: *Učebnoe posobie*. Moscow: Logos.

van der Auwera, Johan. (1998) "Conclusion". Johan van der Auwera, ed. *Adverbial constructions in the languages of Europe*. Berlin and New York: De Gruyter, 813–36.

van der Auwera, Johan, Volker Gast, and Jeroen Vanderbiesen. (2012) "Human impersonal pronouns in English, Dutch and German". Esther Ruigendijk, ed. "A Germanic sandwich". Special issue, *Leuvense Bijdragen* 98(1): 27–64.

van Olmen, Daniel and Adri Breed. (2018) "Human impersonal pronoun in Afrikaans: A double questionnaire-based study". *Language sciences* 69: 1–29.

Vinogradov, Vladimir V., ed. (1954) *Grammatika russkogo jazyka*. Moscow: Izdatel'stvo Akademii Nauk SSSR.

von Waldenfels, Ruprecht. (2006) "Compiling a parallel corpus of Slavic languages: Text strategies, tools and the question of lemmatization and alignment". Bernhard Brehmer, Vladislava Zhdanova, and Rafał Zimny, eds. *Beiträge der europäischen slavistischen Linguistik (POLYSLAV)* 9. Munich: Sagner, 123–38.

————. (2011) "Recent developments in ParaSol: Breadth for depth and XSLT based web concordancing with CWB". Daniela Majchráková and Radovan Garabík, eds. *Natural language processing, multilinguality: Proceedings of Slovko 2011, Modra, Slovakia, 20–21 October 2011*. Bratislava: Tribun, 156–62.

————. (2012) "Polish *tea* is Czech *coffee*: Advantages and pitfalls in using a parallel corpus in linguistic research". Andrea Ender, Adrian Leemann,

and Bernhard Wälchli, eds. *Methods in contemporary linguistics.* Berlin, New York: De Gruyter Mouton, 263–81. [Trends in Linguistics, 247.]
Zaliznjak, Anna A. and Michail Kružkov. (2016) "Baza dannyx bezličnyx glagol'nyx konstrukcij russkogo jazyka". *Informatika i ee primenenija* 10(4): 136–45.
Zobel, Sarah. (2012) *Impersonally interpreted personal pronouns*. Ph.D. thesis, University of Göttingen.

Anastasia Bauer
Slavic Department
University of Cologne
Cologne, Germany
anastasia.bauer@uni-koeln.de

Appendix

A full breakdown of the data for each language and each text type in the dataset and a random sample of the database

Table 1. Impersonalization means in Slavic translated from German text (*n* = 1,033)

Strategies	UK	RU	CZ	PL	BU	CR
3pl	53	48	33	6	57	49
modal/inf	31	30	22	23	0	12
inf	13	21	1	4	0	2
refl	22	20	30	43	41	42
paraphrase	16	16	8	16	32	25
2sg	10	10	1	1	23	4
passive	7	10	15	4	5	7
participle	11	6	0	3	3	3
2pl	0	5	0	0	1	1
lidé	3	4	8	0	0	2
indef pro	9	3	7	5	2	7
3sg	0	2	5	7	0	1
1pl	1	2	3	1	1	2
člověk	0	2	32	9	7	2
adv	0	1	0	2	0	0
modal	1	1	4	3	2	0
-no/-to	0	0	0	40	0	0
refl/modal/inf	0	2	0	0	0	10
refl/inf	0	2	0	0	0	0

Table 2. Impersonalization means in Slavic translated from Italian text (n = 3,008)

Strategies	UK	RU	CZ	PL	BU	CR
paraphrase	74	105	105	58	91	83
refl	35	59	66	147	110	149
modal/inf	101	62	71	92	0	38
1pl	6	23	66	9	47	36
3pl	93	86	33	25	79	44
passive	12	13	46	14	48	19
člověk	4	5	27	7	45	2
indef pro	7	9	23	10	7	9
inf	85	66	33	36	0	15
3sg	1	3	9	4	8	1
modal	0	0	4	13	0	0
2sg	18	19	6	12	35	25
1sg	0	0	0	7	2	0
lidé	2	0	0	0	0	0
2pl	3	5	4	1	6	9
modal/refl	0	0	0	1	2	25
participle	57	51	0	14	16	21
adv	5	0	0	19	0	0
-no/-to	0	0	0	35	0	0
refl/modal/inf	0	2	0	0	0	44
refl/inf	0	4	0	0	0	0

Table 3. Impersonalization means in Slavic translated from Czech text (*n* = 1,384)

Strategies	UK	RU	CZ	PL	BU	CR
3pl	37	39	32	29	46	34
1pl	20	32	22	25	33	32
modal/inf	48	31	32	51	11	11
paraphrase	33	29	40	15	31	29
refl	8	22	18	19	29	40
inf	12	17	6	18	0	1
participle	3	13	0	4	9	1
passive	18	13	16	0	11	18
člověk	17	13	32	28	22	24
lidé	13	9	11	1	9	13
2sg	9	5	3	6	8	4
indef pro	0	5	9	13	11	9
2pl	3	2	12	2	6	5
3sg	1	1	5	2	5	1
-no/-to	0	0	0	14	0	0
modal/refl	0	2	0	0	0	12
ref/modal/inf	0	0	0	0	0	6

Table 4. Impersonalization means in Slavic (*n* = 5,425)

Strategies	CZ	BU	RU	PL	CR	UK
refl	114	180	65	212	230	64
3pl	98	182	172	60	127	183
paraphrase	173	144	149	72	137	124
modal/inf	125	0	124	170	54	180
1pl	100	81	54	37	70	27
inf	36	0	103	57	18	109
člověk	91	74	18	46	39	19
passive	77	64	34	18	44	36
participle	0	29	67	31	25	75
2sg	10	66	34	19	33	37
indefinite pro	39	20	17	34	25	16
-no/-to	0	0	0	89	0	0
lidé	19	11	14	1	15	18
2pl	16	13	12	3	15	6
refl/modal/inf	0	0	2	0	61	0
3sg	9	9	5	2	3	2
modal/refl	0	13	1	5	9	1
adv	0	0	1	9	0	0
refl/inf	0	0	6	0	0	0

GE (original)	BU	CZ	CR	PL	RU	UK
455 Achthundert Jahre lang hatte **man** hierher die Toten des Krankenhauses Hotel-Dieu und der umliegenden Pfarrgemeinden verbracht, achthundert Jahre lang Tag für Tag die Kadaver zu Dutzenden herbeigekarrt und in lange Gräben geschüttet, achthundert Jahre lang in den Grüften und Beinhäusern Knöchelchen auf Knöchelchen geschichtet.	Цели осемстотин години тук бяха **заравяли** умрелите от болницата "Божи дом" и околните енории, цели осемстотин години, ден подир ден, бяха извозвали с тарги трупове, за да ги изтърсват в дългите ровове, цели осемстотин години в гробници и костници се бяха пластили кокал върху кокал. **(3pl)**	Po osm set let sem **nosili** nebožtíky z městské nemocnice a z okolních farností, po osm set let sem dennodenně sváželi na kárách desítky mrtvol a házeli je do dlouhých hrobů, po osm set let vrstvili v hrobkách a kostnicích kostičku na kostičku. **(3pl)**	Osam stotina godina **donosili** su tamo mrtvace iz bolnice Hôtel-Dieu i okolnih žup- [*sic*], osam stotina godina, dan za danom, dovozili su u kolicima na desetke leševa i istresali ih u duge rake, osam stotina godina slagali su u grobnice i kosturnice koščice na koščice, sloj po sloj. **(3pl)**	Przez osiemset lat **chowano** tu zmarłych ze szpitala Hótel-Dieu i okolicznych parafii, przez osiemset lat dzień w dzień tuzinami zwożono tu trupy i zrzucano do podłużnych dołów, przez osiemset lat w kryptach i kostnicach składano warstwami kości. **(*-no*/*-to*)**	Восемьсот лет подряд сюда **доставляли** покойников из Отель-Дье и близлежащих приходов, восемьсот лет подряд сюда на тачках дюжинами свозили трупы и вываливали в длинные ямы, восемьсот лет подряд их укладывали слоями, скелетик к скелетику, в семейные склепы и братские могилы. **(3pl)**	Протягом восьмисот років **спроваджували** сюди мерців із лікарні Отель Дью та навколишніх парафій, протягом восьмисот років день у день привозили сюди десятки трупів, звалювали їх у довжелезні могили, нашаровуючи у склепах та на звалищах кісток. **(3pl)**

GE (original)	BU	CZ	CR	PL	RU	UK
558 Und erst später, am Vorabend der Französischen Revolution, nachdem einige der Leichengräben gefährlich eingestürzt waren und der Gestank des überquellenden Friedhofs die Anwohner nicht mehr zu bloßen Protesten, sondern zu wahren Aufständen trieb, wurde er endlich geschlossen und aufgelassen, wurden die Millionen Knochen und Schädel in die Katakomben von Montmartre geschaufelt, und man errichtete an seiner Stelle einen Marktplatz für Viktualien.	Едва по-късно, в навечерието на Френската революция, когато някои гробове застрашително пропаднаха и смрадта от преливащото гробище подтикна жителите не към безобиден ропот, а към истински бунт, то най-сетне бе затворено и **изоставено**. **(passive)**	A teprve později, v předvečer Francouzské revoluce, když se některé hroby nebezpečně propadly a zápach překypujícího hřbitova dohánět sousední obyvatele nejen k pouhým protestům, nýbrž k opravdovým vzpourám, teprve tehdy byl hřbitov konečně zavřen a zrušen a milióny hnátů a lebek pak zahrabali do katakomb na Montmartre a na původním místě **zřídili** trh s potravinami. **(3pl)**	Tek poslije, uoči Francuske revolucije, pošto su se neki grobovi opasno urušili, pa je smrad što je provalio s groblja natjerao susjedno stanovništvo ne samo na obične prosvjede nego i na prave ustanke, napokon su ga zatvorili i napustili, milijune kosti i lubanja nabacali lopatama u katakombe Montmartrea, a na njegovu mjestu **podignuli** tržnicu. **(3pl)**	I dopiero później, w przededniu Rewolucji Francuskiej, kiedy to niektóre groby zapadły się niebezpiecznie i smród występującego z brzegów cmentarzyska skłonił mieszkańców już nie do gołych protestów, ale do prawdziwych buntów, cmentarz został nareszcie zamknięty i zlikwidowany, miliony piszczeli i czaszek zsypano do katakumb Montmartre, a na tym miejscu **urządzono** targowisko. **(-*no*/-*to*)**	И лишь позже, накануне Французской революции, после того как некоторые из могил угрожающе обвалились и вонь переполненного кладбища [*sic*] побудила жителей предместья не только к протестам, но и к настоящим бунтам, кладбище было наконец закрыто и разорено, миллионы костей и черепов сброшены в катакомбы Монмартра, а на этом месте **сооружен** рынок. **(passive)**	І лише згодом, напередодні Французької революції, після того як кілька могил небезпечно обвалились і нестерпний сморід змусив жителів не просто запротестувати, а повстати, цвинтар нарешті закрили і, закопавши мільйони кісток та черепів у катакомбах Монмартру, **влаштували** на тому місці ринковий майдан для торгівлі харчем. **(3pl)**

GE (original)	BU	CZ	CR	PL	RU	UK
1025 Geschrei, Gerenne, im Kreis steht die glotzende Menge, **man** holt die Polizei.	Врява, топуркане, тълпа зяпачи се струпва около й. Притичва стражата. **(paraphrase)**	Křik, shon, dokola čumící dav a **někdo** zavolá policii. **(indef pro)**	Krika, strka, u krugu mnoštvo izbuljenih očiju, **zovu** policiju. **(3pl)**	Krzyk, rwetes, gapie gromadzący się dookoła, **ktoś** sprowadza policję. **(indef pro)**	Крик суматоха, толпа зевак окружает тело, **приводят** полицию. **(3pl)**	Викрики метушня, довкола вирячені очі, натовп **кличе** поліцію. **(paraphrase)**
1186 Und weil sie geständig ist und ohne weiteres zugibt, dass sie das Ding bestimmt würde haben verrecken lassen, wie sie es im übrigen schon mit vier anderen getan habe, macht **man** ihr den Prozess, verurteilt sie wegen mehrfachen Kindermords und schlägt ihr ein paar Wochen später auf der Place de Greve den Kopf ab.	Тъй като тя си признава и без да усуква, заявява, че пак е щяла да остави това изчадие да пукне, както впрочем е сторила вече с предишните четири, я **изправят** пред съд, осъждат я за многократно детеубийство и няколко седмици по-късно отсичат главата й на площад "Грев". **(3pl)**	A protože se beze všeho přizná, že by byla nechala nemluvně zahynout, jako to ostatně udělala už se čtyřmi předchozími, **je postavena** před soud, odsouzena pro několikanásobnou vraždu novorozenců a o několik týdnů později jí pak na náměstí de Grěve setnou hlavu. **(passive)**	I budući da priznaje i bez ustručavanja dodaje da bi ga svakako pustila da crkne, kako je već, uostalom, učinila i sa četiri prijašnja, **izvode** je na sud, osude je zbog višestrukog čedomorstva i za nekoliko tjedana odrube joj glavu na Place de Grèveu. **(3pl)**	A że przyznaje się do winy i nie zaprzecza, iż chciała robaka zostawić na zatracenie, jak zresztą zrobiła to już z czterema poprzednimi, **wytaczają** jej proces, skazują za wielokrotne dzieciobójstwo i w kilka tygodni potem na place de Gréve ucinają głowę. **(3pl)**	И так как она ничего не отрицает и без лишних слов признает, что собиралась бросить ублюдка подыхать с голоду, как она, впрочем, проделывала уже четыре раза, ее **отдают** под суд, признают виновной в многократном детоубийстве и через несколько недель на Гревской площади ей отрубают голову. **(3pl)**	І, позаяк вона визнає свою провину й відверто признається, що мала намір залишити дитя вмирати, як робила це доти з чотирма іншими дітьми, на неї **заводять** справу, засуджують за багаторазове дітовбивство, а через кілька тижнів на майдані де Грев відтинають їй голову. **(3pl)**

GE (original)	BU	CZ	CR	PL	RU	UK
2101 Es gibt doch im Herbst eine Menge Dinge, die **man** vorbeibringen könnte.	Ето на, есен е, и **човек** може да донесе какво ли не. **(човек)**	Na podzim je přece spousta věcí, které by **se daly přinést. (refl/inf)**	Eto, u jesen ima toliko toga što bi **se moglo donijeti. (refl modal/inf)**	Jesienią jest przecież mnóstwo rzeczy, które **można by podrzucić. (modal/inf)**	Мало ли по осени вещей, которые **можно было бы занести** идучи мимо. **(modal/inf)**	Восени багато чого **можна було б сюди принести**. **(modal/inf)**

GE	IT (original)	BU	CZ	CR	PL	RU	UK
9406 Könnte **man** daher nicht sagen, daß uns das Buch der Natur lediglich abstrakte Wesenheiten verkündet, wie zahlreiche ehrwürdige Theologen lehren? “	Non dobbiamo dunque dire che il libro della natura ci parla solo per essenze, come insegnano molti insigni teologi?”	Значи ли това, че **не можем да твърдим**, че книгата на природата ни говори само по същество, както учат мнозина видни теолози? **(1pl)**	Neměli **bychom tedy říci**, že kniha přírody k nám promlouvá pouze svými abstraktními pojmy, jak tomu učí mnozí slavní teologové? **(1pl)**	Zar **ne moramo dakle reći** da nam knjiga prirode govori samo o biti, kako naučavaju mnogi znameniti teolozi? **(1pl)**	Czyż **nie powinni śmy więc powiedzieć**, że księga przyrody przemawia do nas jedynie przez esencje, jak nauczają liczni znamienici teologowie? **(modal/inf)**	**Значит** книга природы изъясняется только общими понятиями, как и учат многие именитые богословы? **(paraphrase)**	Отож **чи не мусим ми сказати**, що книга природи мовить нам лише про сутності, як це навчають багато славетних богословів? **(1pl)**

GE	IT (original)	BU	CZ	CR	PL	RU	UK
26206 Bedenke, wenn selbst den aufbegehrenden Engeln so wenig genügte, um ihre Inbrunst der Anbetung und der Demut umschlagen zu lassen in eine Inbrunst der Hoffart und der Rebellion gegen Gott, was soll **man** dann von den schwachen Menschen sagen?	Se bastò così poco agli angeli ribelli per mutare il loro ardore d'adorazione e umiltà in ardore di superbia e di rivolta, cosa dire di unessere umano?	Щом на разбунтувалите се ангели им е било нужно толкова малко, за да превърнат пламенното си обожание и смиреност в надменна и бунтовна ярост, **какво да кажем** за една човешка твар? **(1pl)**	Tak málo stačilo vzpurným andělům, aby přeměnili horoucí zbožňování a pokoru v horoucí pýchu a vzpouru, a **co teprve** slabé bytosti lidské! **(paraphrase)**	Ako je pobunjenim anđelima tako malo trebalo da svoj žar obožavanja i poniznosti prometnu u žar oholosti i bune, što **reći** o ljudskom biću? **(inf)**	Skoro tak niewiele wystarczyło zbuntowanym aniołom, by zmienić żar uwielbienia i pokory w żar pychy i buntu, cóż **powiedzieć** o istocie ludzkiej? **(inf)**	Если ангелам-бунтовщикам столь немного хватило, чтоб огнь обожания и смирения стал в них огнем гордыни и бунта, что **говорить** о слабом роде человеческом? **(inf)**	Якщо бунтівним ангелам так мало треба було, щоб перемінити свою жагу віри і покори в жагу гордині й бунту, то що ж тоді **казати** про людину? **(inf)**
223158 Das Übel treibt **man** nicht aus.	"Il male non si esorcizza.	Злото не **се прогонва**. **(refl)**	Neduh **se** z těla **nevyhání**. **(refl)**	Zlo **se** ne **izaginje**. **(refl)**	Choroby **się** nie **egzorcyzuje**. **(refl)**	Болезнь не **изгоняют**. **(3pl)**	Від недуги не зцілюють її **знищують**. **(3pl)**

GE	IT (original)	BU	CZ	CR	PL	RU	UK
1318 Bedenkt man, daß er zur Zeit des Geschehens Novize war und zur Zeit der Niederschrift seiner Erinnerungen an der Schwelle des Todes stand, so ist anzunehmen, daß sein geheimnisumwittertes Manuskript in den letzten zehn oder zwanzig Jahren des 14. Jahrhunderts entstand.	Calcolando che si dice novizio nel '27 e ormai vicino alla morte quando stende le sue memorie, possiamo congetturare che il manoscritto sia stato stilato negli ultimi dieci o vent'anni del Quattordicesimo secolo.	Самият той отбелязва, че през 1327 г. е бил послушник и че когато се е заловил да пише своите спомени, е бил пред прага на смъртта; това ни навежда на мисълта, че ръкописът е бил създаден през последните десет или двайсет години на XIV в. **(paraphrase)**	není však jasné, kdy je napsal. **Vezmeme**-li v úvahu, že v roce 1327 byl novic a v době, kdy paměti psal, jednou nohou v hrobě, můžeme říci, že rukopis vznikl v posledních desíti nebo dvacíti letech čtrnáctého století. **(1pl)**	Ako r**ačunamo** da za sebe kaže kako je '27. bio iskušenik te da se, u trenutku dok zapisuje uspomene, bliži smrti, možemo nagađati da je rukopis sastavio u posljednih deset ili dvadeset godina 14. stoljeća. **(1pl)**	**Wziąwszy** pod uwagę fakt, że jak sam powiada, w roku 27 był nowicjuszem, a kiedy spisuje wspomnienia, jest już bliski śmierci, możemy przypuścić, iż manuskrypt powstał w ostatnim dziesięcioleciu lub dwudziestoleciu czternastego wieku. **(part)**	**Исходя** из того, что автор в 1327 году был послушником, а во время, когда пишется книга, он уже близок к окончанию жизни, можно предположить, что работа над рукописью велась в последнее десяти—или двадцатилетие XIV века. **(part)**	**Зважаючи**, що, за його словами, у 1327 році він був новіцієм, а спогади свої писав уже близьким до смерті старцем, рукопис можна гіпотетично датувати останніми десятиліттями XIV сторіччя. **(part)**

GE	IT (original)	BU	CZ	CR	PL	RU	UK
50283 Eins weiß ich allerdings: Wenn man den Katalog der Bücher durchblättert, findet **man** unter den Bemerkungen, die nur der Bibliothekar versteht, oft eine, die heißt „Africa“, und ich habe sogar schon „finis Africae“ gefunden.	Ma io so una cosa: chi sfogli il catalogo dei libri, troverà, tra le indicazioni che solo il bibliotecario conosce, una che dice sovente ‘Africa’ e ne ho trovata persino una che diceva ‘finis Africae’.	всеки, **който** прелисти каталога на ръкописите, **ще намери** сред сиглите, които са известни само на библиотекаря, една, която се среща често — „Africa“; дори съм се натъквал и на обозначението „finis Africae“. **(indef pro)**	Jenže já vím ještě něco: **prolistujeme**-li katalog knih, **najdeme** mezi odkazy, ve kterých se vyzná jen knihovník, jeden, který se často opakuje a zní ‘Afrika’, a našel jsem dokonce finis Africae. **(1pl)**	Ali ja znam jedno: **tko** prelista katalog knjiga, među naznakama koje poznaje samo knjižničar, često će naći jednu koja glasi ‘Africa’, a našao sam čak i jednu koja glasi ‘finis Africae’. **(indef pro)**	Wiem jedno: **kto** przejrzy katalog ksiąg, wśród wskazań, które zna tylko bibliotekarz, często spotka słowo „AFRICA“, a znalazł em nawet jedno mówiące „FINIS AFRICAE“. **(indef pro)**	**Полистайте** каталог. Среди тайных, понятных лишь библиотекарю обозначений найдете визу „Африка“. А я отыскал даже визу „предел Африки“. **(2pl)**	Але одне я знаю: **гортаючи** каталог книг, серед позначок, відомих лише бібліотекареві, часто **можна побачити** позначку “Africa”,—і я бачив навіть позначку, що гласить “finis Africae”. **(part/modal/inf)**

GE	BU	CZ (original)	CR	PL	RU	UK
7232 „Wird das ihre sexuelle Entwicklung nicht stören, wenn **man** sie Karenin nennt?“	Но ако и викаме Каренин, това няма *ли* *да* разстрои сексуа*л*ния и живот? **(1pl)**	“Ale nenaruší to její sexualitu, když jí **budeme říkat** Karenin? **(1pl)**	Ali neće li to narušiti njenu seksualnost, ako je **budemo nazivali** Karenjin? **(1pl)**	—Ale czy nie skomplikuje jej tożsamości seksualnej, jeśli **będziemy na nią wołać** Karenin? **(1pl)**	Но если мы **станем звать** ее Каренин, не повлияет *ли* это на ее сексуальность? **(1pl)**	—
1619 Man kann nie wissen, was **man** wollen soll, weil **man** nur ein Leben hat, das **man** weder mit früheren Leben vergleichen noch in späteren korrigieren kann.	**Човек** няма как да знае какво би трябвало да иска, защото живее единединствен живот и не може да го сравнява с предишните си животи, нито пък да го поправи в следващите. **(човек)**	**Člověk** nikdy nemůže vědět, co má chtít, protože žije jen jeden život a nemůže ho nijak porovnávat se svými předchozími životy, ani ho opravit v následujících životech. **(člověk)**	**Čovjek** nikad ne može znati šta treba da želi, jer živi samo jedan život i nikako ga ne može usporediti sa svojim prethodnim životima, niti ga u sljedećim životima popraviti. **(čovjek)**	**Człowiek** nigdy nie może wiedzieć, czego ma chcieć, ponieważ dane mu jest tylko jedno życie i nie może go w żaden sposób porównać ze swymi poprzednimi życiami ani skorygować w następnych. **(człowiek)**	**Мы** никогда **не можем** знать, чего мы должны хотеть, ибо проживаем одну-единственную жизнь и не можем ни сравнить ее со своими предыдущими жизнями, ни исправить ее в жизнях последующих. **(1pl)**	**Людина** нічого не знає, до чого мусить прагнути, бо живе одне-єдине життя й не може ні порівняти його зі своїми попередніми життями, ні виправити його в наступних життях. **(людина)**

GE	BU	CZ (original)	CR	PL	RU	UK
14561 Solche Koinzidenzen sind so häufig, daß **man** sie oft nicht wahrnimmt.	—	Obrovské většiny takových koincidencí si **člověk** vůbec nevšimne. **(člověk)**	Ogromnu većinu takvih koincidencija **čovjek** uopće ne primjećuje. **(čovjek)**	—	Огромное множество таких совпадений **человек** не замечает вовсе. **(человек)**	—
16888 Wäre der Traum nicht schön, könnte **man** ihn schnell wieder vergessen.	Ако не беше красив, **човек** бързо щеше да го забрави. **(човек)**	Kdyby sen nebyl krásný, **bylo by možno na něj rychle zapomenout**. **(modal/inf)**	Kad san ne bi bio lijep, ne **bi bilo teško brzo zaboraviti** na njega. **(inf)**	Gdyby sen nie był piękny, **można by** o nim szybko **zapomnieć**. **(modal/inf)**	Не будь сон красивым, о нем **можно было бы мигом забыть**. **(modal/inf)**	—
2636 Mit Metaphern spielt **man** nicht.	Не е желателно да си **играеш** с тях. **(2sg)**	S metaforami **není radno si hrát**. **(inf)**	Nije preporučljivo **igrati se** s njima. **(refl)**	Z metaforami nie **należy igrać**. **(modal/inf)**	С метафорами шутки плохи. **(paraphrase)**	З ними небезпечно **жартувати**. **(inf)**
693 Was also soll **man** wählen?	Кое тогава да **изберем**? **(1pl)**	Co si tedy **máme** zvolit? **(1pl)**	Pa za šta onda da **se odlučimo**? **(refl)**	Cóż więc **mamy wybrać**? **(1pl)**	Так что же предпочтительнее? **(paraphrase)**	То чому ж **віддамо** перевагу? **(1pl)**

Slavic "Quirky Subject" Constructions with *ē*-Statives: Origin and Development

Jasmina Grković-Major

Abstract: This paper discusses the origin and development of constructions with *ē*-statives denoting sensation, emotion, perception, and cognition, which exhibit traces of non-nominative alignment in the history of Slavic languages. Patterns where the experiencer is encoded by the accusative or dative case were inherited from an earlier semantically aligned system, whose relics are found in other Indo-European languages as well. These structures have been subjected to various syntactic, morphological, and semantic changes in the history of Slavic, leading to the establishment of transitive constructions and thus the strengthening of syntactic alignment. The analysis shows that the pace of this process and the types of changes that *ē*-stative constructions underwent were determined by the level of the participant's volitivity and control.

1. Introduction[1]

Constructions in Indo-European (IE) languages whose alignment differs from the canonical nominative alignment of late PIE have presented a great challenge for the past 150 years. As already pointed out by Delbrück (1900: 23–37), they fall into three categories: denoting (a) weather conditions, (b) experiences, and (c) modality. Although such structures vary, their common features are the following: (a) the verb is in the 3SG, and (b) the affected or experiencing person, if specified, is expressed by an oblique morphological case, most often accusative or dative. Being aberrant from the canonical structures with nominative subjects, these constructions instigated a fruitful discussion on subject types ("logical", "psychological", "grammatical") already in the epoch of *Junggrammatiker* (Graffi 2001: 73–109). The authors, who considered the subject to be strictly a grammatical category, called these structures "subjectless" (Miklosich 1883).

[1] The glosses adhere to the Leipzig Glossing Rules, with the additional abbreviations AOR 'aorist'; CONJ 'conjunction'; IMPF 'imperfect'; PART 'particle'. Another abbreviation used in this paper is OCS 'Old Church Slavonic'.

Journal of Slavic Linguistics 29(2): 179–200, 2021.

Further investigations, which included genetically unrelated languages, led to the question of whether the subject is a universal category present in the grammar of every natural language, as, for example, proposed by Keenan in 1976. However, a year later Foley and van Valin (1977) stated that "subject is not a valid theoretical construct (universal) in linguistic theory", and that the idea of the subject-predicate dichotomy as elemental was brought to us through traditional logic and grammatical traditions since the Greeks. They based their claims on the analysis of three languages typologically different from IE systems. Especially interesting in this respect was Lakhota, an active-stative language, which they claim does not appear to have any clause-level referential structure, being a semantic-role dominated language. This is in accordance with a proposition given earlier by Klimov (1983: 106–07) in the framework of his contentive typology. He states that the nominative language type is characterized by a specific sentence type, determined by a transitivity feature, where, on the morphological level, the nominative and the accusative case have a clear "subject and object orientation" (see also Klimov 1972). Following Klimov, Gamkrelidze and Ivanov (1995: 271–76) argued that transitivity is the semantic basis both for nominative ("accusative") and ergative languages, while the active-stative ones do not have this feature at all. Similarly, Desnickaja (1951: 143) stated that "transitivity and intransitivity in their mutual opposition are historical categories, and their role and significance in a lexical-semantic system as well as in the grammatical system of a given language may not be viewed as primarily given or stable". Within a different theoretical framework, Hale (1983: 25) stated that there are languages in which lexical structure is configured differently from phrase structure. Recently we also encounter the term "semantic alignment", describing "the phenomenon whereby basic alignment property of a language can best be described by appealing to semantic factors, rather than syntactic ones" (Donohue 2008: 24). What is common in all these explanations is that there are languages that have only semantic valency.[2]

The first one to observe such a typological profile of early Proto-Indo-European (PIE) was Meillet (1908: 321–30). He argued that PIE was a system with autonomous sentence elements, where a word was self-sufficient to indicate its role in the discourse, with no "governing" of one word by another. Morphological cases were used depending on the intended meaning, expressing semantic roles, and there was no verb valency.[3] The basic principles of syntactic structuring were apposition and agreement, connecting semantically

[2] While *syntactic valency* refers to a number of arguments in a clause, *semantic valency* of a verb refers to "the number of the semantic roles associated with it" (van Valin 2003: 92). Cf. Payne 2007: 169–70.

[3] For the function of the nominative case, Meillet (1908: 308) used the term "subject", but for him it was a *topic*: "Le nominatif indique de quoi il est question dans la phrase,

related elements. Subsequently, a number of linguists elaborated the idea that PIE was a non-nominative, semantically aligned language, and that syntactic changes of PIE and its daughter languages are the result of typological transformation leading to the creation of syntactic alignment, caused by the rise of transitivity (see Burridge 1993; Gamkrelidze and Ivanov 1995; Krys'ko 1997; Bauer 2000; Lehmann 2002; Hewson and Bubenik 2006; Grković-Major 2007, 2010a; Barðdal and Eythórsson 2009; Luraghi 2010a; Pooth et al. 2019).

In the older stages of IE languages, we see a number of "syntactic archaisms" revealing this earlier typological profile: the so-called "absolute" verbs with no transitivity feature (Desnickaja 1984: 148) and only semantic valency, impersonal and absolute constructions (Bauer 2000), free word order, discontinuous constituents, null anaphora (Ponti and Luraghi 2018), etc. These syntactic archaisms are abundantly represented in the old Slavic languages as well (Grković-Major 2007, 2010a, 2011, 2012; Pavlović 2011). In the course of time, some of them disappeared, being replaced by new structures; some were reanalyzed; and some took up a marginal place in the system. However, some "syntactic residues" survived, and being non-canonical, formally marked structures, they became both semantically and functionally marked (cf. Havránek 1958: 79–80).

Among such syntactic archaisms in the Slavic languages are structures wherein the first participant is encoded by an oblique case. Their typology in contemporary systems is well described (Mrazek 1990). In this paper, we will focus on the origin and development of Proto-Slavic constructions with experiential *ē*-statives (infinitives in *-ěti*).

2. IE Constructions with Accusative and Dative Experiencers

Constructions with impersonal verbs and accusative or dative experiencers are well known from various IE languages (e.g., Bauer 2000: 93–145; Barðdal and Eythórsson 2009; Grković-Major 2012; Matasović 2013). This is a restricted set of verbs, mostly denoting unpleasant, negative experiences, both physical and emotional, such as Hittite *istarak-*, *irmaliya-*, *armaniya-* 'be(come) ill', *arsana*, *arsaniya-* 'envy', *kistanziya-* 'be hungry', etc. (Luraghi 2010b); Latin *paenitet*$_{3SG}$ 'regret', *miseret*$_{3SG}$ 'be sorry', *piget*$_{3SG}$ 'bother', etc. (Matasović 2013); Lithuanian *gelti* 'ache', *skaudėti*, *sopėti* 'hurt', *niežėti* 'itch', etc. (Piccini 2008); and Old English *grisan* 'fear', *hreowsian* 'trauern' ['mourn'], *yfelian* 'suffer', *tweogan* 'doubt' (Pishwa 1999), etc. We will illustrate them with well-known examples from Latin:

le 'sujet' ". Much later, Lehmann (1976) argued that early PIE was a topic-prominent language.

(1) a. me pudet
I_{ACC} be.ashamed$_{3SG.PRS}$
'I am ashamed'

b. mihi dolet
I_{DAT} be.in.pain$_{3SG.PRS}$
'I ache'

While the encoding of the experiencer may vary, the common feature is a verb in the 3SG. As noted by Benveniste (1966: 227–36), 3rd person, as opposed to 1st and 2nd, is not a "person". This is reflected in the fact that the form of the PIE 3SG *-m* conjugation differs from the 1SG and 2SG, which have endings of pronominal provenance. The original status of the 3SG can be clearly seen in meteorological verbs with null valency[4] (Delbrück 1900: 23–24), probably belonging to the oldest chronological layer of impersonals (cf. Savčenko 1974: 333). They refer to the existence of natural phenomena, which, as Wackernagel (2009: 154) stated, "could be represented in this simple form without any thought of a subject". He also points to the "strange" archaic use of the "indefinite" 3SG in early Latin legal phraseology, e.g., *si in ius uocat* 'if (one man) calls (another) to court', in the "Laws of the Twelve Tables" (5th c. BC) and the analogous Greek examples, claiming that "the agreement must be based on common inheritance, and this linguistic feature must be something very ancient" (Wackernagel 2009: 149–51). This brings us to the conclusion that 3SG initially just denoted a process.

The accusative experiencer seems to be older than the dative one (Delbrück 1900: 33). The morphological reconstruction of PIE cases, first given by Popov in 1879–81 (Popov 2012; see Krys'ko 1990, Danylenko 2016),[5] speaks in favor of this assumption; a form called "proto-accusative", which subsequently gave the accusative case, was the first general oblique "case" in early PIE. A semantically diffuse form, expressing all kinds of circumstances under which an action or state took place, it generally meant "in reference to *x*". Its residues in the form of *accusativus relationis* are present in many old IE languages (Ernout and Thomas 1953; Whitney 2004; Fraenkel 1928; Krys'ko 1997; Grković-Major 2007, 2010b). Ernout and Thomas (1953: 19) considered it to be an appositive "autonomous determination" not governed by the verb. See, for example, (2):

[4] Even Paul (1970: 131), who insisted that every IE sentence must have a subject and a predicate, admits that such sentences are truly subjectless. For more details about the origin of IE constructions with "meteorological verbs", see Grković-Major 2013.

[5] Popov's reconstruction of IE morphological cases was later accepted and developed by Gamkrelidze and Ivanov (1995: 233–52) within the active-stative typology of early PIE.

(2) timeo Danaos (Latin)
fear$_{1SG.PRS}$ Greeks$_{ACC}$
'I fear *in reference to* Greeks' > 'I fear Greeks'

Accordingly, impersonal constructions with the accusative experiencer may be interpreted as follows:

(3) me pudet (Latin)
I$_{ACC}$ be.ashamed$_{3SG.PRS}$
'*in reference to* me there is shaming' > 'I am ashamed'

With the rise of the dative case in PIE, originally restricted to the category [+animate] or personified notions (Kuryłowicz 1964: 191, 196), the dative, as the "recipient case", started entering constructions with experiencer verbs. Its competition with the older accusative is seen in the daughter languages (1).

These patterns were subjected to various changes in Indo-European languages. They could be replaced by nominative alignment constructions in different ways and at a different pace, even within the same subgroup of languages. The history of the Romance languages shows that the majority of impersonal emotion verbs shifted to a personal conjugation (Bauer 2000: 129). Old English had approximately 40 impersonal verbs, some of them having both dative and accusative experiencers, e.g., *maetan* 'dream' (Bauer 2000: 132), but in the Middle English period they were being replaced by agent-like experiencers (Pishwa 1999: 132). On the other hand, the process has been slower in German, which offers, according to von Seefranz-Montag (1981: 536), "a slow motion picture of syntactic change in progress", with a tendency to replace them with dummy subject constructions.

3. Slavic Constructions with *ē*-Statives

Proto-Slavic (PS) had a number of *ē*-statives denoting physical, emotional, and mental states.[6] Their infinitives (*-ěti*) are built with the PIE suffix **ē* (<**eH1*) (see Yakubovich 2014), which was used in Balto-Slavic to form intransitive aorists (Meillet 1934: 244).[7] In the development of the IE verbal systems, the same suffix was used to form different categories expressing a state or a "situ-

[6] On verbs in *-ěti*, see Vaillant 1966: 377–405.

[7] The present tense of the primary statives was in $e^i/i > i$ (**mьněti$_{INF}$*, *mьniši$_{2SG.PRS}$* 'think'), while the present tense of denominals was in *(ě)je* (**cělěti$_{INF}$*, *cělěješi$_{2SG.PRS}$* 'heal'). Some of them have "anomalous" presents, such as **xotěti$_{INF}$*, *xošteši$_{2SG.PRS}$* 'want, wish', where the old form is preserved only in the 3PL *xotętъ*, while the rest of the present paradigm represents the old optative (Vaillant 1966: 403).

ation" (Beekes 1995: 230). According to Ivanov (1981: 221), PS *ē*-statives present a transformation of the inactive series of PIE verbs with "centripetal" semantics".[8] In comparison to the *-mĭ* paradigm, this series was originally "defective", having only the 3SG. In other words, it was "structurally impersonal, without paradigmatic oppositions for person" (Gamkrelidze and Ivanov 1995: 257).

PS monovalent verbs denoting physical states (*stojati* 'stand', *běžati* 'run away', *sědĕti* 'sit', *ležati* 'lie', etc.) became intransitives already in OCS (4a). Semantically bivalent statives (**dьržati* 'hold', *vьrtĕti* 'turn around', etc.) developed syntactically transitive constructions, with nominative subjects and accusative objects (4b):

(4) a. ideže stoěste nodzě ego
where stand$_{3DU.AOR}$ feet$_{NOM.DU}$ he$_{GEN.SG}$
'where his feet stood' (OCS; *PsSin* 131.7)

b. drъžaaxǫ i
hold.back$_{3PL.IMPF}$ he$_{ACC}$
'they held him back' (OCS; *Mar* Lk 4:42)

On the other hand, statives denoting sensations, emotions, perception, and cognition do not always exhibit nominative alignment and have kept the accusative or dative experiencers throughout the history of Slavic.

3.1. Sensation and Emotion Verbs

Sensation and emotion verbs are analyzed together since the division between physical and emotional states is historically fuzzy. Emotion designations are the result of metaphorical and metonymical changes of words denoting concrete states, actions, and activities causing emotions or caused by emotions.[9] Some *ē*-statives had undergone semantic shifts by the time of the first written records (e.g., OCS *skьrbĕti* 'be sad, worry, hurt' < 'be sharp, cut'; Petleva 1988–90: 52), but some of them were still polysemous, meaning both sensation and emotion (e.g., OCS *bolĕti* 'be in pain physically, be in pain emotionally').

Verbs denoting negative sensations or emotions marked [−volitive] and [−control], such as PS **bolĕti* 'be in pain, be sick, hurt' or **svьrbĕti* 'itch' (see

[8] The PIE "semantically centripetal subject-version forms naturally became the means for marking intransitive semantics" (Gamkrelidze and Ivanov 1995: 292).

[9] The first study dealing with the semantic sources of the words for emotions in (some) Indo-European languages was Kurath 1921, followed by Buck's 1949 comprehensive "dictionary of ideas".

Miklosich 1868–74: 353), have the construction corresponding to Latin *pudet me*:[10]

(5) a. srbi me (Slovenian)
itch$_{3SG.PRS}$ I$_{ACC}$
'it itches me'

b. boli me (Serbian)
hurt$_{3SG.PRS}$ I$_{ACC}$
'it hurts me' / 'I am sick'[11]

A body part affected by a negative sensation is expressed by the nominative case in all three branches of Slavic:

(6) a. zělo mja glava bolitъ
very.much I$_{ACC}$ head$_{NOM}$ hurt$_{3SG.PRS}$
'I have a strong headache' (Old Russian; *SRJa* 1: 281)

b. bolĭ ju zzyrdcze
hurt$_{3SG.PRS}$ she$_{ACC}$ heart$_{NOM}$
'her heart hurts/aches' (Old Czech; *VW*)

c. kada čoika boli glava
when man$_{ACC}$ hurt$_{3SG.PRS}$ head$_{NOM}$
'when a man has a headache' (Old Serbian; *L*)

We also find the dative experiencer:

[10] Such verbs are rarely attested in old Slavic texts (not once in OCS), but this is understandable in view of the type and genre of the documents. For example, the majority of Old Serbian medieval documents are juridical texts, which do not permit expressions of sensations and emotions. The accusative experiencer is thus very rare, attested, for instance, in letters, in another type of experiential construction expressing the same semantics:

(i) stidъ nasъ jestъ
shame$_{NOM.SG}$ we$_{ACC}$ exist$_{3SG.PRS}$
'we are ashamed' (*PP*: 581, 1422)

The same kind of impersonal structure is found also in Old Czech, for instance, *mě jest túha* 'I am longing/I am sad'; *mě bude hněv* 'I will be angry'; *hrozno mě jest* 'I am terrified', etc. (Gebauer 2007: 316).

[11] Found in the 19th century and in some contemporary dialects. In the *RJA* (1: 537), such sentences are explained as subjectless, with the following specification: "it is not said what causes pain, but that there is pain".

(7) a. ašte čeloveku bolit serce
if man$_{DAT}$ hurt$_{3SG.PRS}$ heart$_{NOM}$
'if a man's heart hurts' (Old Russian; *SK*)

b. hlava mu boli
head$_{NOM}$ he$_{DAT}$ hurt$_{3SG.PRS}$
'he has a headache' (Old Czech; Gebauer 2007: 386)

c. ili ti e zaboléla glava
or you$_{DAT}$ AUX begin.to.hurt$_{3SG.PTCP.PRF}$ head$_{NOM}$
'or you got a headache' (Old Serbian; *PTP* 71a)

Taking into account comparative IE data, we might presume that the affected body part was originally expressed by *accusativus relationis* (see Desnickaja 1984: 89–199).[12] The replacement of the accusative (8a) by the nominative (8b) is seen in the history of Lithuanian:

(8) a. mane visą skauda
I$_{ACC}$ all$_{ACC}$ hurts
'I am aching all over' (older Lithuanian; Piccini 2008: 445)

b. man viskas skauda
I$_{DAT}$ all$_{NOM}$ hurts
'I am aching all over' (Lithuanian; Piccini 2008: 445)

In some contemporary Slavic languages, such constructions are still impersonal, with no agreement between the nominative and the verb:

(9) boljalo go zăb (Bulgarian)
hurt$_{PTCP.PRF.N}$ he$_{ACC}$ tooth$_{NOM.M}$
'he apparently had a toothache'

In others it has further developed into a structure where the nominative controls the agreement:

(10) bolela me je glava (Serbian)
hurt$_{PTCP.PRF.F}$ I$_{ACC}$ AUX head$_{NOM.F}$
'I had a headache'

[12] Its trace in Slavic impersonal constructions is Russian *mne golovu bol'no*, while in personal constructions it was replaced by the instrumental case (see Gadolina 1958: 209–12).

Although it has a formal nominative subject (*glava*) and an accusative object (*me*), this is a pseudo-transitive, "quirky" construction, since the nominative denotes a stimulus, the accusative the experiencer, and the predicate (*boleti*) is intransitive.[13]

Eventually within the scheme of nominative alignment, the meaning of the predicate was reinterpreted as 'cause pain/restlessness' (see *RSANU* 2: 49; *SSKJ*). This change was supported by the fact that statives in *-ěti* shared the present *-i-* stem with the productive class of factitives in *-iti*, such as PS **gubiti* 'kill', **staviti* 'put', and **umoriti* 'kill'.[14] The reinterpretation of sensation predicates as causatives led to the introduction of verbs that originally signified actions causing unpleasant sensations. The experiencer became an object, with no specification of a subject:

(11) a. menja znobit (Russian)
I$_{ACC}$ shiver$_{3SG.PRS}$
'I shiver'

b. bode me (Slovenian)
pierce$_{3SG.PRS}$ I$_{ACC}$
'it is piercing me'

c. guši me[15] (Serbian)
choke$_{3SG.PRS}$ I$_{ACC}$
'it is choking me'

Since emotions, unlike sensations, imply evaluation as a conscious mental activity, their experiencer has a certain degree of control, depending on the type of emotion and level of the volitivity feature. This semantic class of verbs gradually developed nominative subjects.

Negative emotion verbs with low or no volitivity and control features, such as **bojati sę* 'be afraid' and **styděti sę* 'be ashamed', were transformed

[13] Sentences such as Serbian **Petar$_{NOM}$ boli$_{3SG}$ Jovana$_{ACC}$*, with the intended meaning that Petar causes Jovan pain, are not acceptable at all. The construction was gradually generalized by including other nouns denoting stimuli: *Petra$_{ACC}$ boli$_{3SG}$ istina$_{ACC}$* 'Petar is hurt by the truth'. See Grković-Major 2012.

[14] Although their present tense originally differed in accentuation, the two paradigms eventually were unified (Vaillant 1966: 437–38).

[15] For more examples, see Mrazek 1990: 95–96. The affected body part can be specified with different prepositional phrases, which is a language-specific feature (Běličová and Uhlířova 1996: 57).

into reflexives expressing "middle", "centripetal" semantics already in PS.[16] In this way, they developed "internal transitivity", with the accusative (object) **sę* being coreferential with the subject:

(12) azъ esmь ne boite sę
I am NEG be.afraid$_{2SG.IMP}$ REFL
'it is me, do not be afraid' (OCS; *Mar* Mt 14:27)

Negative emotion verbs exhibiting volitivity and control became intransitives. In the following example, OCS *trъpěti* 'suffer' expresses not only that a subject will suffer but that he is willing to do so. This further caused its semantic shift into 'endure':

(13) trъpljǫ do kon'ca
suffer/endure$_{1SG.PRS}$ until end
'I will (suffer >) endure until the end' (OCS; *SS*: 705)

A difference between a negative sensation and a negative emotion can be seen in the polysemous verb *boleti*. If it meant 'be in pain, hurt physically', the experiencer was, as shown earlier, patient-like, but if it meant 'be in pain, hurt emotionally', it had an agent-like experiencer:[17]

(14) dětišťь plačetъ i mati bolitъ
child cries and mother$_{NOM}$ hurt$_{3SG.PRS}$
'the child is crying and the mother is hurting (in emotional pain)'
(OCS; *Supr* 312.8–9)

Changes in the constructions with negative emotion verbs were gradual and depended on their semantics. For example, PS **mьrzěti* 'be loathsome, repellent' is found in OCS only in the 3rd person with the dative experiencer:

(15) vesъ denъ slovesa moě mrъzěaxǫ imъ
all day words my be.loathsome$_{3PL.IMPF}$ they$_{DAT}$
'my words were loathsome to them all day long' (OCS; *PsSin* 55.6)

[16] Reflexive verbs developed in IE languages lacking middle voice as a grammatical category (Večerka 1993: 130). Both categories express the same "centripetal" semantics. Cf. Shenker 1988.

[17] This is in accordance with Seržant's (2013: 305) conclusions that "there is a change in meaning concomitantly with the change from the original oblique case-marking into the nominative one".

Old Slavic languages had not only the dative but also the accusative experiencer (see Dal' 1881: 326; *VW*; *SłStp*: 233; *RJA* 7: 100–01), which points to the common PS origin of such patterns. They have been subjected to various language-specific changes, both syntactic and semantic. In the history of Serbian and Croatian, the original meaning 'be loathsome, repellent' is preserved with the non-nominative experiencers:

(16) a. taj te dar sad mrzi
this$_{NOM}$ you$_{ACC}$ gift$_{NOM}$ now be.loathsome$_{3SG.PRS}$
'that gift is loathsome to you' > 'you do not like that gift'
(older Serbian and Croatian; *RJA* 7: 100–01)[18]

b. Bogu to mnogo mrzi
God$_{DAT}$ this$_{NOM}$ very.much be.loathsome$_{3SG.PRS}$
'that is loathsome to God' > 'God does not like it'
(older Serbian and Croatian; *RJA* 7: 100–01)[19]

The dative pattern was lost, while the accusative one is kept with the infinitive or the *da*-clause complement in contemporary Serbian and Croatian,[20] denoting a feeling of not wanting to do something. Its preservation was probably supported by the generalization of the type (11):

(17) a. mrzi me pisati loše kritike (Croatian)
be.loathsome$_{3SG.PRS}$ I$_{ACC}$ write$_{INF}$ bad reviews$_{ACC}$
'I do not feel like writing bad reviews'

b. mrzi me da učim (Serbian)
be.loathsome$_{3SG.PRS}$ I$_{ACC}$ COMP study$_{1SG.PRS}$
'I do not feel like studying'

On the other hand, the verb also developed a transitive construction, accompanied by its semantic change into 'hate':

(18) mrzim da učim (Serbian)
hate$_{1SG.PRS}$ COMP study$_{1SG.PRS}$
'I hate to study.'

[18] The example is from the works of M. Nelješković, a 16th-century writer from Dubrovnik (see *RJA* 6: 947).

[19] The example is from a Croatian Glagolitic book, *Korizmenjak* (1508) (see *RJA* 6: 943).

[20] They are present today in both standards, although in different ratio (see the Serbian web corpus, http://nlp.ffzg.hr/resources/corpora/srwac/, and Croatian web corpus, http://nlp.ffzg.hr/resources/corpora/hrwac/).

Positive emotion statives are found in two kinds of constructions from the earliest records. When meaning 'want', which presumes volitivity and control,[21] PS **hotěti/hъtěti* 'want, wish' gradually developed transitive syntax with agent-like experiencers. However, the traces of semantic alignment are still found in the history of the Slavic languages; a patient is (rarely) attested in the genitive, mostly with abstract nouns (see *SDrJa* 3: 1381; *RJA* 3: 663–64), but often with the dative, which was dominant in OCS:

(19) a. emuže ašte xošteši damь ti
what$_{DAT.SG}$ PART want$_{2SG.PRS}$ give$_{1SG.PRS}$ you$_{DAT}$
'whatever you want I will give to you' (OCS; *Mar* Mr 6:22)

b. zlěmь dinarem ne htě grьci
bad dinars$_{DAT}$ NEG want$_{3PL.PRS}$ Greeks
'the Greeks do not want bad dinars' (Old Serbian; *PP*: 43)

c. ne tolma xotja pobědě
NEG only want$_{PTCP.NOM.SG}$ victory$_{DAT}$
'not only wanting victory' (Old Russian; Pravdin 1956: 72)

d. jakému chceš, panno, muži
which$_{DAT.SG}$ want$_{2SG.PRS}$ maiden$_{VOC}$ husband$_{DAT.SG}$
'which husband do you want, maiden?'
(Old Czech; Gebauer 2007: 378)

On the other hand, the semantics of 'wish' (volitivity and no control) was expressed by the reflexive 3SG and a dative experiencer. The reflexive marks the "centripetal" predicate force, compatible with the recipient status of the experiencer. We find this type of construction already in OCS, but also in the old Slavic languages:

(20) a. poslušati sę jemu xoštetь zapovedii
obey$_{INF}$ REFL he$_{DAT}$ wish$_{3SG.PRS}$ commandments$_{GEN.PL}$
n<e>b<e>sьnyxь
divine
'he wishes to obey the heavenly commandments'
(OCS; *SSJa* 4: 785)[22]

[21] Wanting, as a simple intentional state (*MIT*: 132), implies a degree of control of the first participant, insofar as intentions presume cognitive processes.

[22] This example is from a 13th-century Russian Church Slavonic text whose archetype was translated from Latin in Bohemia. The angle brackets in the example indicate letters that are omitted in the original manuscript.

b. němaju oprava kako b-i-mь se htělo
not.have$_{3PL.PRS}$ things conj aux-they$_{DAT}$ refl wish$_{PTCP.PRF}$
'they do not have as many things as they wish'
(Old Serbian; *PP* 286)

c. mne s nim rostatisja ne xočetsja
I$_{DAT}$ with him part$_{INF.REFL}$ neg wish$_{3SG.PRS.REFL}$
'I do not wish to part with him'
(Old Russian; Borkovskij 1968: 139)

d. zachtělo se mi masa
wish$_{PTCP.PRF}$ refl I$_{DAT}$ meat$_{GEN.SG}$
'I wished for meat' (Old Czech; Gebauer 2007: 13)

As the transitive agent-like type was grammaticalized, the "quirky" reflexive pattern was reinterpreted into '*x* feels like' and extended to incorporate other verb classes in all three branches of Slavic (Borkovskij 1968: 137–43; Georgieva 1969: 74–75; Grković-Major 2004: 198). Today this modal construction is productive in South and East Slavic (Běličová and Uhlířová 1996: 60), denoting a recipient-like experiencer situation. Ivić (1973: 86) distinguishes two basic types—the first one marked [+volitive] (21a), the second [−volitive] (21b)—while Mitkovska (2019: 283) thoroughy analyzes a continuum of the "various modal nuances from necessity and urge through need, craving, desire, inclination to determination" in South Slavic:

(21) a. ide mi se u bioskop
go$_{3SG.PRS}$ I$_{DAT}$ refl in movies
'I feel like going to the movies' (Serbo-Croatian; Ivić 1973: 86)

b. kija mi se
sneeze$_{3SG.PRS}$ I$_{DAT}$ refl
'I have an urge to sneeze' (Serbo-Croatian; Ivić 1973: 86)

3.2. Perception and Cognition Verbs

Perception and cognition verbs are analyzed together because they historically constitute a continuum: physical perception evolves into "mental perception" ('see' > 'know', 'listen' > 'obey').[23]

[23] Due to the general closeness of these two domains, Talmy (2003: 139) postulates a cognitive domain of "ception, which encompasses the traditional notions of 'perception' and 'conception'".

These predicates are found in two types of constructions, which mark different degrees of their experiencer's volitivity and control. On the one hand, they gradually developed transitive syntax with agent-like experiencers. But still in the oldest records we see traces of semantic alignment. In OCS, as well as in the early Slavic languages, there is a competition in formalizing the second participant, which could be expressed by different cases. For example, with *slyšati* 'listen' it could be denoted by genitive (source), dative (goal), oreven by *accusativus relationis* (22); and with *mьněti* 'think', by the double accusative[24] (23):

(22) ioanъ že slyšavъ vъ ǫzilišti děla °xva
John PART hear$_{PTCP.PST}$ in prison deeds$_{ACC}$ Christ's
'when John heard in prison about the deeds of the Christ'
(OCS; *Mar* Mt 11:2)

(23) vy bo běsъni sǫšte ny
you$_{NOM}$ PART insane$_{NOM.PL}$ be$_{PTCP.PRS.NOM.PL}$ we$_{ACC}$
cělomǫdrъnyę běšeny mьnite
wise$_{ACC.PL}$ insane$_{ACC.PL}$ think$_{2PL.PRS}$
'you, being insane, think that we, who are wise, are insane'
(OCS; *Supr* 116.6)

At the same time, the process of establishing transitive syntax with the accusative object (24) or complement clause (25) was underway:

(24) da bǫ slyšali slovo °božie
CONJ AUX listen$_{PTCP.PRF}$ word$_{ACC.SG}$ God's
'in order to listen to God's word' (OCS; *Mar* Lk 5:1)

(25) ne mnite ěko pridъ razoriti zakona
NEG think$_{2PL.PRS}$ COMP come$_{1SG.AOR}$ abolish$_{INF}$ law$_{GEN.SG}$
'do not think that I have come to abolish the Law'
(OCS; *Mar* Mt 5:17)

The second type of construction found in the old Slavic languages denotes lower control of the first participant. It consists of a reflexive 3SG and a dative

[24] The double accusative with perception, cognition, and communicative verbs is a syntactic archaism, replaced by complement clauses in the history of Indo-European languages (Ambrazas 1990: 148–49).

experiencer. Among perception verbs it is found only with the non-volitive *viděti* 'see',[25] when a person is not sure what s/he perceives:

(26) a. mně sja vidit
I$_{DAT}$ REFL see$_{3SG.PRS}$
'it seems to me' (Old Ukrainian; Borkovskij 1968: 141)

b. vidí mi se
see$_{3SG.PRS}$ I$_{DAT}$ REFL
'it seems to me' (Old Czech; Kosek 2012: 10)

c. ako vi se vidi
if you$_{DAT}$ REFL see$_{3SG.PRS}$
'if it seems to you' (Old Serbian; *PP* 800)

The identical pattern, with the same semantics, is found with *mьněti* in OCS and Old West and South Slavic:[26]

(27) a. čьto ti sę mьnitъ
what you$_{DAT}$ REFL think$_{3SG.PRS}$
'how does it seem to you?' (OCS; *Mar* Mt 17:25)

b. mnyeſſe mi sě, bych stál na břežě
think$_{3SG.IMPF}$ I$_{DAT}$ REFL AUX stand$_{PTCP.PRF}$ on shore
'it seemed to me that I stood on the shore' (Old Czech; *VW*)

c. mněše mu se da je Ancilešь ubijenь
think$_{3SG.IMPF}$ he$_{DAT}$ REFL COMP AUX Achilles killed
'it seemed to him that Achilles was killed'
(Old Serbian; *T*: 58–59)

This type of construction with perception and cognition *ē*-statives is today almost completely lost and is found only in some dialects (*BER* 4: 191).

[25] The group of visual (as well as auditory) perception verbs exhibits the opposition [–volitive] : [+volitive], e.g., OCS *viděti* 'see' : *zьrěti*, *ględati*, *sъmotriti*, all 'watch'—i.e., active vs. inactive perception (cf. Verhoeven 2007: 50).

[26] This pattern also existed in Old East Slavic but belonged to the higher registers (Borkovskij 1968: 138), which is indicative of its Church Slavonic origin.

4. Conclusions

The syntactic development of PIE and its daughter languages testifies to the gradual establishment of syntactic alignment caused by the rise of transitivity. This is reflected in the changes of PS constructions with *ē*-statives denoting negative sensations, emotions, perception, and cognition. The pace of this process and the types of changes the *ē*-stative constructions were subjected to were determined by the level of the participant's volitivity and control.

Statives denoting negative bodily sensations, characterized by the features [–volitive] and [–control], exhibit traces of semantic alignment in the history of Slavic: their participant was encoded by the accusative or dative. The accusative pattern was eventually reinterpreted and generalized by including causatives, and being aberrant from the dominant alignment, it became a marked structure, denoting the aberrant status of the participant: a patient-like experiencer. The fact that semantic markedness corresponds to syntactic markedness points to a kind of isomorphism between the two linguistic levels.

Since emotions always include evaluations, experiencers of emotion statives have a certain degree of volitivity and control. However, they evolved differently depending on the level of that degree. If denoting "centripetal" non-volitive negative states, they gave reflexives, becoming "internally transitive" already in PS. In this way, they formalized the double "middle" nature of their agent- and patient-like experiencer. Verbs marked [+volitive] evolved into intransitives with agent-like experiencers from the earliest records. Positive emotion statives marked as [+volitive] eventually gave transitives with agent-like experiencers. Although they have nominative subjects from the earliest written sources on, the process of creating transitive constructions was gradual, since it took time for the accusative objects to be grammaticalized. The same applies to perception and cognition verbs, which also eventually developed transitive syntax with agent-like experiencers.

Non-volitive positive emotion, perception, and cognition statives also had impersonal reflexives with dative-like experiencers (recipients) in the history of Slavic. While the pattern with emotion statives was preserved and then grammaticalized as a modal construction denoting a recipient-like experiencer with other verb classes as well, this possibility no longer exists with perception and cognition verbs. This is because, compared to other semantic classes of experiential statives, their experiencer has the highest control over a situation. This is in accordance with Haspelmath's (2001: 63–64) conclusion that "cognition predicates show the strongest affinity with the agent-like experiencer construction".

Finally, we want to point out that the gradual changes of PS experiencer *ē*-stative constructions caused by the rise of transitivity leading to the creation of the syntactically aligned systems encompassed different linguistic levels.

This was a multifaceted process which included morphological and syntactic innovations, followed by semantic reinterpretations and shifts, while semantic shifts could also lead to syntactic changes.

Sources

[*L*] (18th c.) *Lekaruša*. Manuscript. National Library of Serbia, sign. RS 79.
[*Mar*] (1883/1960) *Codex Marianus glagoliticus*. V. Jagić, ed. Graz: Akademische Druck–u. Verlagsanstalt.
[*PP*] (1929–34) *Stare srpske povelje i pisma.* Vol. 1. Parts 1–2. Ljubomir Stojanović, ed. Belgrade: SKA. [Zbornik za istoriju, jezik i književnost srpskog naroda, vol. 19, no. 24.]
[*PsSin*] (1922) *Sinajskaja psaltyr´: Glagoličeskij pamjatnik XI věka*. Sergěj Sever'janov, ed. Petrograd: Izdanie Rossijskoj akademii nauk".
[*PTP*] (1763) *Pesmarica Teodora Popovića Dobraševića*. Manuscript. Novi Sad: Library of Matica srpska, sign. BMS: Mk 168.
[*RJA*] (1880–1976) *Rječnik hrvatskoga ili srpskoga jezika*. Vols. 1–24. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.
[*RSANU*] (1959–) *Rečnik srpskohrvatskog književnog i narodnog jezika SANU*. Vols. 1– . Belgrade: SANU.
[*SDrJa*] (1893–1912) *Materialy dlja slovarja drevne-russkogo jazyka po pis´mennym pamjatnikam*. Vols. 1–3. I. I. Sreznevskij. St. Petersburg: Tipografija Imperatorskoj akademii nauk.
[*SK*] *Starorusskij korpus*. Available at: http://www.ruscorpora.ru/new/search-mid_rus.html. Last accessed 30 March 2020.
[*SłStp*] (1963–65) *Słownik staropolski*. Vol. 4. Wrocław: Polska Akademia Nauk.
[*SRJa*] (1975–) *Slovar' russkogo jazyka XI–XVII vv.* Vols. 1– . Moscow: Nauka.
[*SS*] (1994) *Staroslavjanskij slovar', po rukopisjam X–XI vekov*. R. M. Cejtlin, R. Večerka, and È. Blagova, eds. Moscow: Russkij jazyk.
[*SSJa*] (1966–97/2006) *Slovar' staroslavjanskogo jazyka*. Vols. 1–4. St. Petersburg: Izdatel'stvo Sankt-Peterburgskogo universiteta.
[*SSKJ*] *Slovar slovenskega knjižnega jezika*. Available at: http://bos.zrc-sazu.si/sskj.html. Last accessed 30 March 2020.
[*Supr*] (1904) *Suprasl'skaja rukopis'*. Vols. 1–2. Sergěj Sever'janov, ed. St. Petersburg: Izdanie Imperatorskoj akademii nauk.
[*T*] (1951) *Eine altserbische Trojasage*. Allan Ringheim. Prague, Upsal: Imprimerie de l'état à Prague.
[*VW*] *Vokabulář webový* [online]. Version 0.4.2. Oddělení vývoje jazyka Ústavu pro jazyk český AV ČR. Available at: http://vokabular.ujc.cas.cz. Last accessed 30 March 2020.

References

Ambrazas, Vitautas. (1990) *Sravnitel'nyj sintaksis pričastij baltijskix jazykov*. Vilnius: Mokslas.

Barðdal, Jóhanna and Thórhallur Eythórsson. (2009) "The origin of the oblique-subject construction: An Indo-European comparison". Vit Bubenik, John Hewson, and Sarah Rose, eds. *Grammatical change in Indo-European languages*. Amsterdam: John Benjamins, 179–93.

Bauer, Brigitte. (2000) *Archaic syntax in Indo-European: The spread of transitivity in Latin and French*. Berlin, New York: Mouton de Gruyter.

Beekes, Robert S. P. (1995) *Comparative Indo-European linguistics: An introduction*. Amsterdam, Philadelphia: John Benjamins.

Běličová, Helena and Ludmila Uhlířová. (1996) *Slovanská věta*. Prague: Euroslavica.

Benveniste, Émile. (1966) *Problèmes de linguistique générale.* Vol. 1. Paris: Gallimard. [Bibliothèque des sciences humaines.]

[*BER*] (1962–) *Bŭlgarski etimologičen rečnik*. Vols. 1– . V. Georgiev, I. Gŭlbov, J. Zaimov, and S. Ilčev. Sofia: Izdatelstvo na Bŭlgarskata akademija na naukite.

Borkovskij, V. I. (1968) *Sravnitel'no-istoričeskij sintaksis vostočnoslavjanskix jazykov: Tipy prostogo predloženija*. Moscow: Nauka.

Buck, Carl Darling. (1949) *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago, London: University of Chicago Press.

Burridge, Kate. (1993) *Syntactic change in Germanic: Aspects of language change in Germanic with particular reference to Middle Dutch*. Amsterdam: John Benjamins.

Dal', Vladimir. (1881) *Tolkovyj slovar' živago velikoruskago jazyka*. Vol. 2. St. Petersburg, Moscow.

Danylenko, Andrii. (2016) "Oleksandr Popov (1855–80) and the reconstruction of Indo-European noun inflection". *Language & history* 59(2): 112–30.

Delbrück, Berthold. (1900) *Vergleichende Syntax der indogermanischen Sprachen*. Vol. 3. Strassburg: Karl J. Trübner.

Desnickaja, A. V. (1951) "Iz istorii razvitija kategorii glagol'noj perexodnosti". *Pamjati akademika L. V. Ščerby*. Leningrad: Izdatel'stvo LGU.

————. (1984) *Sravnitel'noe jazykoznanie i istorija jazykov*. Leningrad: Nauka.

Donohue, Mark. (2008) "Semantic alignment systems: What's what, and what's not". Mark Donohue and Søren Wichmann, eds. *The typology of semantic alignment*. Oxford: Oxford University Press, 24–75.

Ernout, Alfred and François Thomas. (1953) *Syntaxe latine*. Paris: Libraire C. Klincksieck.

Foley, William and Robert Van Valin, Jr. (1977) "On the viability of the notion of 'subject' in universal grammar". *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, 293–320.

Fraenkel, Ernest. (1928) *Syntax der litauischen Kasus*. Kaunas: L.U. humanitarinių mokslų fakulteto leidinys.

Gadolina, M. A. (1958) "Tvoritel'nyj priimennoj". S. B. Bernštejn, ed. *Tvoritel'nyj padež v slavjanskix jazykax*. Moscow: Izdatel'stvo AN SSSR, 289–312.

Gamkrelidze, T. V. and V. V. Ivanov. (1995) *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture*. Berlin, New York: De Gruyter.

Gebauer, Jan. (1929/2007) *Historická mluvnice jazyka českého.* Vol. 4. *Skladba*. Prague: Academia.

Georgieva, V. L. (1969) "Bezličnye predloženija po materialam drevnejšix slavjanskix pamjatnikov (osobenno staroslavjanskix)". *Slavia* 38: 63–90.

Graffi, Giorgio. (2001) *200 Years of syntax: A critical survey*. Amsterdam: John Benjamins.

Grković-Major [Mejdžor], Jasmina. (2004) "Impersonalne konstrukcije u starosrpskom jeziku". *Naučni sastanak slavista u Vukove dane* 33(1): 249–61.

————. (2007) "Razvoj sintaksičke tranzitivnosti". *Zbornik Matice srpske za slavistiku* 71–72: 417–33.

————. (2010a) "Kognitivni aspekti razvoja tranzitivnosti". Jasmina Grković-Mejdžor and Milorad Radovanović, eds. *Teorija dijahronijske lingvistike i proučavanje slovenskih jezika*. Belgrade: SANU, 43–62.

————. (2010b) "The role of syntactic transitivity in the development of Slavic syntactic structures". Björn Hansen and Jasmina Grković-Major, eds. *Diachronic Slavonic syntax: Gradual changes in focus*. Munich, Berlin, Vienna: Verlag Otto Sagner, 63–74. [Wiener slawistischer Almanach, 74.]

————. (2011) "The development of predicative possession in Slavic languages". Nomachi Motoki, ed. *The grammar of possessivity in South Slavic languages: Synchronic and diachronic perspectives*. Sapporo: Slavic Research Center, Hokkaido University, 35–54. [Slavic Eurasian Studies, 24.]

————. (2012) "Obrazac srpskog 'boli me glava' kao sintaksički arhaizam". *Glas SANU—Odeljenje jezika i književnosti* 419(28): 43–60.

————. (2013) "Povodom Miklošičeve studije o rečenicama bez subjekta". Jasmina Grković-Mejdžor and Aleksandar Loma, eds. *Miklosichiana bicentennalia: Zbornik u čast dvestote godišnjice rođenja Franca Miklošiča*. Belgrade: SANU, 293–310.

Hale, Ken. (1983) "Warlpiri and the grammar of non-configurational languages". *Natural language & linguistic theory* 1(1): 5–47.

Haspelmath, Martin. (2001) "Non-canonical marking of core arguments in European languages". Alexandra Aikhenvald, R. M. W. Dixon, and Masayuki Onishi, eds. *Non-canonical marking of subjects and objects*. Amsterdam, Philadelphia: John Benjamins, 53–84.

Havránek, Bohuslav. (1958) "Metodická problematika historickosrovnávacího studia syntaxe slovanských jazyků". *K historickosrovnávacímu studiu slovanských jazyků*. Prague: Státní pedagogické nakladatelství, 77–88.

Hewson, John and Vit Bubenik. (2006) *From case to adposition: The development of configurational syntax in Indo-European languages*. Amsterdam: John Benjamins.

Ivanov, Vjač. Vs. (1981) *Slavjanskij, baltijskij i rannebalkanskij glagol: Indoevropejskie istoki*. Moscow: Nauka.

Ivić, Milka. (1973) "Problematika modalnih rečenica". Jaroslav Burian, ed. *Otázky slovanské syntaxe.* Vol. 3. *Sborník symposia "Modální výstavba výpovědi v slovanských jazycích", Brno 27.–30. září 1971.* 1st edition. Brno: Universita J. E. Purkyně, 85–91.

Keenan, Edward L. (1976) "Towards a universal definition of subject". Charles N. Li, ed. *Subject and topic*. London, New York: Academic Press, 303–34.

Klimov, G. A. (1972) "K xarakteristike jazykov aktivnogo stroja". *Voprosy jazykoznanija* 4: 3–13.

————. (1983) "On contentive typology". *Lingua e stile* 18(3): 327–41.

Kosek, Pavel. (2012) "Vývoj slovosledných vlastností auxiliáru préterita ve starší češtině". Kateřina Najbrtová, ed. *Gramatika a korpus 2012: 4. mezinárodní konference*. Hradec Králové: Gaudeamus, Univerzita Hradec Králové, 1–17.

Krys'ko, V. B. (1990) "Istorija indoevropejskogo akkuzativa v 'Sintaksičeskix issledovanijax' A. Popova". *Voprosy jazykoznanija* 4: 119–30.

————. (1997) *Istoričeskij sintaksis russkogo jazyka: Ob˝jekt i perexodnost'*. Moscow: Indrik.

Kurath, Hans. (1921) *The semantic sources of the words for the emotions in Sanskrit, Greek, Latin, and the Germanic languages*. Menasha, WI: George Banta Publishing Company.

Kuryłowicz, Jerzy. (1964) *The inflectional categories of Indo-European*. Heidelberg: Carl Winter Universitätsverlag.

Lehmann, Winfred P. (1976) "From topic to subject in Indo-European". Charles N. Li, ed. *Subject and topic*. London, New York: Academic Press, 445–56.

————. (2002) *Pre-Indo-European*. Washington: Institute for the Study of Man. [Journal of Indo-European Studies, 41.]

Luraghi, Silvia. (2010a) "The rise (and possible downfall) of configurationality". Silvia Luraghi and Vit Bubenik, eds. *Continuum companion to historical linguistics*. London, New York: Continuum International Publishing Group, 212–29.

————. (2010b) "Experiencer predicates in Hittite". Ronald Kim, Norbert Oettinger, Elisabeth Rieken, and Michael Weiss, eds. *Ex Anatolia Lux: Anatolian and Indo-European studies in honor of H. Craig Melchert on the occasion of his sixty-fifth birthday.* Ann Arbor, MI; New York: Beech Stave Press, 249–64.

Matasović, Ranko. (2013) "Latin *paenitet me*, *miseret me*, *pudet me* and active clause alignment in Proto-Indo-European". *Indogermanische Forschungen* 118: 93–110.

Meillet, Antoine. (1908) *Introduction a l'étude comparative des langues indoeuropéennes*. Paris: Librarie Hachette.

————. (1934) *Le slave commun*. Paris: Librairie ancienne Honoré Champion.

Miklosich, Franz (1868–1874) *Vergleichende Syntax der slavischen Sprachen*. Vienna: Wilhelm Braumüller.

————. (1883) *Subjectlose Sätze*. Vienna: Wilhelm Braumüller.

[*MIT*] (1999) *The MIT encyclopedia of the cognitive sciences*. Robert A. Wilson and Frank C. Keil, eds. Cambridge, London: The MIT Press.

Mitkovska, Liljana. (2019) "Types of modality in South Slavic stative reflexive-dative constructions". *Slověne* 8(2): 260–87.

Mrazek, Roman. (1990) *Sravnitel'nyj sintaksis slavjanskix literaturnyx jazykov: Isxodnye struktury prostogo predloženija*. Brno: Univerzita J. E. Purkyně.

Paul, Hermann. (1880/1970) *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer Verlag.

Pavlović, Slobodan. (2011) "Elementi nekonfigurativne sintakse u starim slovenskim pismenostima". *Zbornik Matice srpske za slavistiku* 80: 47–62.

Payne, Thomas E. (2007) *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.

Petleva, I. P. (1993) "Ètimologičeskie zametki po slavjanskoj leksike. XVII". Ž. Ž. Varbot, L. A. Gindin, G. A. Klimov, V. A. Merkulova, V. N. Toporov, and O. N. Trubačev, eds. *Ètimologija 1988–1990: Sbornik naučnyx trudov*. Moscow: Nauka, 52–57.

Piccini, Silvia. (2008) "Traces of non-nominative alignment in Lithuanian: The impersonal constructions in Indo-European perspective". *Baltistica* 43(3): 437–61.

Pishwa, Hanna. (1999) "The case of the 'impersonal' construction in Old English". *Folia linguistica historica* 20(1–2): 129–51.

Ponti, Edoardo Maria and Silvia Luraghi. (2018) "Non-configurationality in diachrony: Correlations in local and global networks of Ancient Greek and Latin". *Diachronica* 35(3): 367–92.

Pooth, Roland, Peter Alexander Kerkhof, Leonid Kulikov, and Jóhanna Barðdal. (2019) "The origin of non-canonical case marking of subjects in Proto-Indo-European: Accusative, ergative, or semantic alignment". *Indogermanische Forschungen* 124(1): 245–64.

Popov, A. V. (2012) *Sravnitel'nyj sintaksis imenitel'nogo, zvatel'nogo i vinitel'nogo padežej v sanskrite, drevnegrečeskom, latinskom i drugix jazykax*. Moscow: Librokom (reprint).

Pravdin, A. B. (1956) "Datel'nyj priglagol'nyj v staroslavjanskom i drevnerusskom jazykax". *Učenye zapiski Instituta slavjanovedenija* 13: 3–120.

Savčenko, A. N. (1974) *Sravnitel´naja grammatika indoevropejskix jazykov*. Moscow: Vysšaja škola.

Schenker, Alexander M. (1988) "Slavic reflexive and Indo-European middle". Alexander M. Schenker, ed. *American contributions to the Tenth International Congress of Slavists. Linguistics*. Columbus, OH: Slavica, 363–83.

Seržant, Ilja A. (2013) "Rise of canonical subjecthood". Ilja A. Seržant and Leonid Kulikov, eds. *The diachronic typology of non-canonical subjects*. Amsterdam: John Benjamins, 283–310.

Talmy, Leonard. (2003) *Toward a cognitive semantics*. Vol. 1. *Concept structuring systems*. Cambridge, MA: MIT Press.

Vaillant, André. (1966) *Grammaire comparée des langues slaves.* Vol. 3. *Le verbe*. Paris: Éditions Klincksieck.

van Valin, Robert D., Jr. (2003) *An introduction to syntax*. Cambridge: Cambridge University Press.

Večerka, Radoslav. (1993) *Altkirchenslavische (altbulgarische) Syntax*. Vol. 2. *Die innere Satzstruktur*. Freiburg: U. W. Weiher.

Verhoeven, Elisabeth. (2007) *Experiential constructions in Yucatec Maya: A typologically based analysis of a functional domain in a Mayan language*. Amsterdam, Philadelphia: John Benjamins.

von Seefranz-Montag, Ariane. (1981) "'Subjectless' constructions and syntactic change". Jacek Fisiak, ed. *Historical syntax*. Berlin, New York, Amsterdam: Mouton Publishers, 521–53.

Wackernagel, Jacob. (2009) David Langslow, ed. *Lectures on syntax, with special reference to Greek, Latin, and Germanic*. Oxford: Oxford University Press.

Whitney, William Dwight. (2004) *Sanskrit grammar (Including both the classical language, and the older dialects, of Veda and Brahmana)*. New Delhi: Munshiram Manoharlal Publishers Pvt. Ltd. (reprinted from the second edition of 1889).

Yakubovich, Ilya. (2014) "Reflexes of Indo-European '*ē*-Statives' in Old Indic". *Transactions of the Philological Society* 112(3): 386–408.

Jasmina Grković-Major
Department of Serbian Language and Linguistics
Faculty of Philosophy
University of Novi Sad
Novi Sad, Serbia
jgrkovicns@gmail.com

The Competition of ARB Constructions in Polish*

Maria Katarzyna Prenner and Daniel Bunčić

Abstract: Polish has three quasi-synonymous impersonal constructions: the *-no/-to* construction, the reflexive impersonal, and the 3PL impersonal. This raises the question of what the differences between them are and how one of them is selected. This paper presents the results of an acceptability judgment test, which is informed by an explorative corpus study and examines the following factors: colloquial vs. neutral register; perfective vs. imperfective aspect; present tense vs. preterite; and generic vs. specific reading. The main findings are that the 3PL impersonal turned out to be better in colloquial discourse and the reflexive impersonal is much more acceptable with the imperfective than with the perfective aspect. Furthermore, the corpus data and the acceptability judgments show numerically that both reflexive and 3PL impersonal have a certain tendency towards present tense and generic reading, and that the *-no/-to* construction is more typical with the perfective than the imperfective aspect.

1. Introduction

The Polish language has several impersonal constructions. This paper is concerned with three of them: the *-no/-to* construction (Polish *bezosobnik*, cf. e.g., Fellerer 2008) as in (1), the reflexive impersonal (*się* construction) as in (2), and the impersonal third-person plural as in (3).

(1) Wróco**no** do swoich domów.
return$_{PST.IMPRS}$[1] to one's houses
'People returned to their homes.' (Wiemer 1995: 314)

* Research for this article was funded by the German Research Foundation (DFG), project ID 281511265, SFB 1252 "Prominence in Language".

[1] The following abbreviations, which follow the Leipzig Glossing Rules where possible, are used in the glosses and diagrams: 1 'first person'; 2 'second person'; 3 'third person'; ACC 'accusative'; AUGM 'augmentative'; COND 'conditional'; F 'feminine'; FUT 'future'; GEN 'genitive'; IMPRS 'impersonal'; INSTR 'instrumental case'; M 'masculine'; N 'neuter'; NEG 'negative particle'; NVIR 'non-virile (= not masculine-human)'; PL 'plural'; PRS 'present'; PST 'past, preterite'; REFL 'reflexive (pronoun)'; SG 'singular'; VIR 'virile (= masculine-human)'.

(2) Czyta **się** książkę.
read$_{PRS.3SG}$ REFL book$_{ACC}$
'One reads a book.' (Wiese 1973: 625)

(3) Znowu **podnieśli** cenę paliwa.
again raise$_{PST.3PL.VIR}$ price$_{ACC}$ fuel$_{GEN}$
'They have raised the price of fuel again.' (Kibort 2008: 263)

Gast and van der Auwera (2013: 123) subsume these expressions under the category of "human impersonal pronouns" in the wider sense. More specifically, all three constructions belong to a category of impersonals called "*arb* constructions" (Cinque 1988: 544) or "arbs" and defined as "constructions with arbitrary interpretations" (Malamud 2013: 1).[2] (This definition excludes some of the "human impersonal pronouns", see (8–11) on p. 204) Although these constructions and their arbitrary nature have been extensively investigated (cf. e.g., Cabredo Hofherr 2003; Mel'čuk 1974; Padučeva 2012), the proposed analyses are diverse, and the semantics of ARBs is still far from being thoroughly understood (Malamud 2013: 2).

The quasi-synonymity of these three ARB constructions raises the obvious question of the choice of expression: When does a Polish native speaker choose which of these constructions? The extant literature on these expressions does not give sufficient answers to this question. Therefore, this paper attempts to clarify the differences between the three Polish ARBs on the basis of a corpus analysis and an acceptability judgment test. Specifically, we will investigate the effects of register, tense, aspect, and generic vs. specific reading of the implicit subject.

The problem addressed here arose in the context of a project within the Collaborative Research Center 1252 at Cologne University, "Prominence in Language". This project is based on the finding that ARBs are not equally good with all verbs, which can be explained by assuming a prominence relation in the sense of Himmelmann and Primus 2015, according to which the constructions are deemed more acceptable the more agentive the verbs are (cf. Bunčić 2018, 2019, 2020). However, in order to be able to determine this effect (and in very practical terms, to construct valid test items), we have to know the influence of other factors on the grammaticality and/or acceptability (cf. Haider 2019) of these constructions.

[2] Note that Cinque (1988: 529) used "arb" for a certain class of meanings and "*arb* constructions" (ibid. 544) as a cover term for constructions that can have these meanings (in this case, the Italian *si* and 3PL impersonal), whereas we follow Malamud (2013: 1) in calling the constructions themselves ARBs and using the terminology introduced in (4–7) for the meanings these constructions can have (cf. also Egerland 2003: 76, fn. 4 on differences in terminology).

In the following, a short definition of the category of ARBs (§2) will be given. On the basis of what is known so far about Polish ARBs, we can then make predictions about the possible factors influencing the choice of construction (§3). The predictions will be tested in an explorative corpus study (§4) and an acceptability judgment test (§5), the results of which will be discussed in detail (§6).

2. Definition of ARBs

All three constructions have a demoted subject/agent (and are thus impersonal constructions in the sense of Siewierska 2008a: 116) with reduced referentiality, which is why Malchukov and Ogawa (2011: 44) classify them as "R-impersonals"—a category that, however, also includes sentences like *It dawns* (ibid. 25). Gast and van der Auwera (2013: 124) describe this reduction of referentiality as "impersonalization", by which they mean "the process of filling an argument position of a predicate with a variable ranging over sets of human participants without establishing a referential link to any entity from the universe of discourse". A number of different meanings can be associated with such expressions, which have been classified with varying degrees of granularity; see (4–8):

(4) a. *neopredelënno-ličnye predloženija* 'indefinite-personal sentences'
b. *obobščënno-ličnye predloženija* 'generalized-personal sentences'
(Vinogradov 1954: 5–12)[3]

(5) a. quasi-existential
b. quasi-universal (Cinque 1988: 545)

(6) a. specific ("1st person singular, 'I'")
b. arbitrary ("a non-specific group of individuals")
c. generic ("a quasi-universal set of individuals")
(Egerland 2003: 76)

(7) a. specific existential reading (temporally anchored)
b. vague existential reading (not temporally anchored)
c. inferred existential reading (inferred from a result)
d. corporate reading (predicates with a designated subject)
e. universal reading (licensed by a locative)
(Cabredo Hofherr 2003: 83)

[3] Note, however, that apart from this semantic definition, these terms are also used to distinguish different forms. Thus, Padučeva (2012: 27) uses the term *neopredelënno-ličnye predloženija* to refer to the Russian 3PL impersonal regardless of its meaning.

(8) a. episodic; existential, vague
b. episodic; existential, plural, indefinite
c. episodic; existential, plural, definite
d. generic; universal, external
e. generic; universal, internal
f. modal; universal, internal
g. non-assertive; universal, internal

(Gast and van der Auwera 2013: 140f)

However, the "universal"/"generic" readings can easily be achieved with many other expressions as well (cf. Bauer this volume.), e.g., with *człowiek* 'man, human being' as in (9), with *ludzie* 'people' as in (10), with the second person as in (11), or with the first-person plural as in (12).[4]

(9) W takim domiszczu **człowiek** spodziewa się
in such house$_{AUGM}$ man expect REFL
większych rzeczy.
greater things
'In such a huge house one expects greater things.'
(ParaSol; Polish original)

(10) **Ludzie** nigdy nie dowiedzieli się, dokąd.
people never NEG found.out REFL whither
'Nobody ever found out where to.'
(ParaSol; translation from German)

(11) Prawdy nie **dojdziesz**, **choćbyś** i tam był.
truth NEG reach$_{FUT.2SG}$ even.if$_{2SG}$ also there be$_{PST}$
'One will not learn the truth even if one has been there.'
(ParaSol; Polish original)

(12) wszystko, **cośmy** raz przeżyli, miałoby się
all that$_{1PL}$ once experienced must.COND REFL
kiedyś powtórzyć
someday repeat
'everything one once experienced would have to recur someday'
(ParaSol; translation from Czech)

[4] Examples (9–12) were retrieved from the ParaSol corpus, a parallel corpus of Slavic and other languages (von Waldenfels and Meyer 2006–).

None of the expressions in (9–12) can be interpreted in any of the non-universal readings associated with ARBs (i.e., Vinogradov's reading from (4a), Cinque's reading from (5a), Egerland's reading from (6b), Cabredo Hofherr's readings from (7a–d), Gast and van der Auwera's readings from (8a–c)). Therefore, it seems sensible not to treat these expressions as ARBs and to restrict the definition of ARBs to those constructions that can have non-universal (i.e., existential), or arbitrary, readings.

3. Predictions

The three Polish ARB constructions are characterized by important differences and constraints. For example, the *-no/-to* construction and the 3PL impersonal trigger virile (plural) marking in agreement, while the reflexive can occur with any agreement, depending on the implicit referent. See, for example, (13) and (14):

(13) Pracowano jako nauczyciele / *nauczycielki / *nauczyciel / work$_{PST.IMPRS}$ as teacher$_{PL.VIR}$ teacher$_{PL.NVIR}$ teacher$_{SG.M}$
*nauczycielka.
teacher$_{SG.F}$
'People worked as teachers.' (Kibort 2008: 267)

(14) Pracowało się jako nauczyciele / nauczycielki / work$_{PST.3SG.N}$ REFL as teacher$_{PL.VIR}$ teacher$_{PL.NVIR}$
nauczyciel / nauczycielka.
teacher$_{SG.M}$ teacher$_{SG.F}$
'People worked as (female) teachers. / One worked as a (female) teacher.' (Kibort 2008: 273)

Furthermore, the *-no/-to* impersonal is restricted to the preterite and conditional, whereas the reflexive and the 3PL impersonal can be used in all tenses (preterite, present, future; cf. (2), (14), (19), (22)). This might mean that in certain contexts in which the *-no/-to* construction would be preferred to the other two constructions in the preterite, these constructions are perfectly acceptable in the present tense. We therefore predict an effect of tense on the 3PL and reflexive impersonal.

Although in the Slavic languages verbal aspect is such a pervasive category that it plays a role in almost all areas of grammar, so far it has not been considered as a factor influencing the acceptability of ARBs, and no studies have been carried out to investigate this. We will therefore investigate whether aspect has an effect on ARBs.

In the literature, the 3PL impersonal is often described as having "a strongly colloquial flavour" (Siewierska 2008b: 22). This leads to the simple assumption that the 3PL impersonal should be rated better in informal registers. In more formal registers, however, there seems to be a certain preference for the *-no/-to* construction. In the literature, this construction is sometimes described as "stylistically neutral" (Siewierska 2008b: 22) and sometimes as formal or bookish (Bartnicka and Lehmann 2004: 522; Puzynina 1993: 40; Skibicki 2016: 420). A possible explanation for this is that the use of *-no/-to* is perceived as more detached, perhaps because it mainly occurs in newspaper articles where the main function is the "objectivization" of information and where the irrelevance of the agent is emphasized (Laskowski 1984: 147). For this reason, the *-no/-to* construction might be considered less suitable for everyday communication. On this basis, we predict that the *-no/-to* impersonal should be rated better in a formal register. The reflexive impersonal, however, is described as stylistically neutral, which is why we do not expect any differences here.

As seen in Section 2, ARBs can express both universal (generic) and non-universal (specific, arbitrary) readings. The morphosyntactic differences between the three constructions might cause them to express one reading more easily than another, which in the situation of competition among the constructions might have an effect on their choice and acceptability.

We therefore predict that preferences for one of the three constructions might depend on the following factors: register, aspect, tense, and universal vs. non-universal reading.

4. Explorative Corpus Study

Before designing an experiment with native speakers, we conducted an explorative corpus study to see if there are any significant effects of these four factors on the frequency of the ARBs.

However, the problems associated with finding the three constructions in a monolingual Polish corpus like the NKJP (National Corpus of Polish; cf. Przepiórkowski et al. 2012) are very unevenly distributed. Thus, the *-no/-to* construction is completely unambiguous (since the form of the passive participle from which it originally derived now ends exclusively in *-ne/-te*), so that a search for *pracowano* 'one worked', for example, does not yield any false positives or false negatives.[5] The reflexive impersonal is homonymous with a range of other constructions with the reflexive pronoun: real reflexives, recip-

[5] The only exceptions are purely coincidental homonyms. For example, a search for the rare *padano* 'one fell' in the NKJP yielded a great number of false positives, all of which were preceded by *grana* and often capitalized to *Grana Padano* (a kind of Italian cheese).

rocals, anticausatives, reflexiva tantum (or inherent reflexives (15)), middles (16), and antipassives (17).[6]

(15) Maria wyspała się.
Maria out.sleep$_{PST.3SG.F}$ REFL
'Maria had a good sleep.'
(Fehrmann, Junghanns, and Lenertová 2010: 207)

(16) Te samochody prowadzą się łatwo.
these cars drive$_{3PL}$ REFL easily
'These cars are easy to drive.' (Rivero and Sheppard 2003: 93)

(17) Chłopiec rzucał się kamnieniami.
boy throw$_{PST.3SG}$ REFL stones$_{INSTR}$
'The boy was throwing stones.' (Janic 2013: 161)

A restriction to the neuter preterite form immediately adjacent to the reflexive pronoun (e.g., *pracowało się* 'one worked') can considerably reduce the number of false positives at the expense of missing all the instances in the present, future, and conditional, and where *się* is separated from the verb by other words. Finally, it is virtually impossible to find the 3PL impersonal in a corpus because it is completely homonymous with the regular (personal) third-person plural with pro-drop, which of course is more frequent by several magnitudes.

Consequently, we decided to use a Polish-German parallel corpus and look for the impersonal pronoun *man* as well as passives in the German version in order to avoid any construction bias that a direct search for the Polish constructions would have. For this explorative study, the ParaSol corpus (von Waldenfels and Meyer 2006– ; cf. von Waldenfels 2006) seemed sufficient; furthermore, we restricted the analysis to the only two Polish original texts with a German translation in the corpus in order to exclude effects that foreign texts might have on the choice of construction by a Polish translator.[7]

In the corpus we found many constructions corresponding to German *man* or the German passive that are not ARBs—e.g., the personal passive, personal reflexive, infinitival constructions, verbal nouns, *można* 'one can', *trzeba*

[6] Cf. Rivero and Sheppard 2003: 99. Note, however, that in contrast to other Slavic (and Romance) languages there is no reflexive personal passive in Polish.

[7] The search terms used were "man" for the impersonal pronoun and "w[ieu]rd.*"[]* "(ge|abge|ange|aufge|ausge|be|beige|darge|einge|er|hinge|hinter|nachge|niederge|über|um|unter|ver|vorge|wegge|wider|zer|zuge|zusammenge).*(en|t)" within s for the passive. Only the texts *Kongres futurologiczny* (1971) by Stanisław Lem and *Opowieści galicyjskie* (1995) by Andrzej Stasiuk were selected.

'one must', *widać* 'one can see', *słychać* 'one can hear', *czuć* 'one can feel', *człowiek* 'human being, man', *ludzie* 'people', *kto* 'who(ever)', *ktoś* 'someone', *nikt* 'nobody', etc. However, since the aim of the study was not a contrastive comparison, we discarded these hits and included in our analysis only the three ARB constructions.

All in all, the little corpus search yielded 118 ARBs, of which roughly half (n = 61) were reflexive impersonals, a third (n = 36) were *-no/-to* constructions, and a sixth (n = 21) 3PL impersonals. Their distribution over tenses and aspects is illustrated in Figures 1 and 2 below. The diagrams show that the reflexive impersonal is mainly used in the present tense (at a ratio of 49 : 12 in relation to all other forms) and that the most frequent ARB in the present tense is the reflexive (49 : 9). In the perfective preterite, the vast majority of the instances is covered by the *-no/-to* construction (24 : 5), which in turn is used considerably more often in the perfective than in the imperfective aspect (25 : 11). The main field of competition between the three ARBs seems to be the imperfective preterite, and the 3PL impersonal is the one that does not show clear preferences for any tense (preterite 10 : 9 present) or aspect (IPFV 7 : 5 PFV, not counting the present tense).

As our sample of the corpus contains only written texts, it is hardly possible to get any significant data on register from the small number of ARBs we found. We therefore refrained from a corpus analysis of register. As to the question of universal vs. non-universal readings, in many cases it was hard to tell, even with the context given in the corpus, which reading is to be preferred, as in (18) on the opposite page. Therefore, we did not conduct a

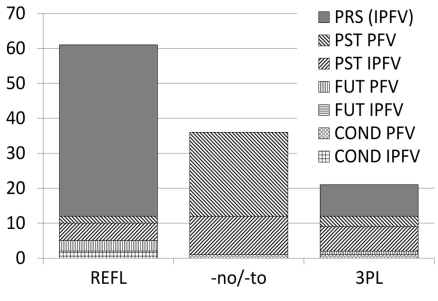


Figure 1. Tense and aspect against ARBs

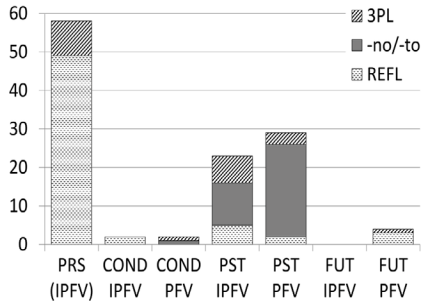


Figure 2. ARBs against tense and aspect

quantitative analysis. However, our qualitative analysis gives us the impression that in the imperfective preterite, which Figure 2 identifies as the main "battleground" of the constructions, the reflexive typically has a universal reading as in (19), whereas the *-no/-to* construction often has a non-universal reading as in (20).

(18) O polskiej mafii mówio**no** i pisa**no** już w połowie lat 90.
'Polish mafia was talked and written about as early as the mid-nineties.'

(19) Wtedy pisało **się** na maszynach do pisania.
'At that time people wrote on typewriters.'

(20) W 1980 roku zadłużenie Polski sięgało 11,5 mld zł, a nie 30 mld zł, jak pisa**no**.
'In 1980, the debt of Poland reached 11.5 billion złoty, and not 30 billion złoty as was written.'

5. Acceptability Judgment Test

The results of the corpus analysis made it seem worthwhile to verify all four factors—register, aspect, tense, and universal vs. non-universal reading—in an acceptability judgment test.

5.1. Test Items

As Bunčić (2018, 2019) has shown, at least the *-no/-to* construction and the reflexive impersonal are more acceptable with a more prominent agent than with a less prominent agent. We therefore selected 20 highly agentive Polish verbs[8]—i.e., verbs entailing Dowty's (1991) agentivity features [+volition], [+sentience], and [+movement]—thus keeping the level of agentivity effects across all test items constant. Apart from that, all verbs are transitive (which has a similar effect because subjects of transitive verbs are more prominent agents than subjects of intransitive verbs). These verbs were embedded in sentences with uniform syntactic behavior; they all start with a prepositional

[8] The tested verbs are the following: *dać* 'give'; *dodać* 'add'; *napisać* 'write'; *oddać* 'give back'; *odwrócić* 'turn'; *pić* 'drink'; *położyć* 'lay'; *przestawić* 'move'; *przynieść* 'bring'; *rozbić* 'break'; *rozwiązać* 'solve, untie'; *sciągnąć* 'take off, take down'; *stawić* 'put'; *ugotować* 'cook'; *wykopać* 'dig'; *wypić* 'drink up'; *wziąć* 'take'; *zbudować* 'build'; *złożyć* 'fold, put together'; *związać* 'tie (up)'.

phrase, followed by the verb in the particular impersonal construction and a direct object, and end with a subordinate clause, as in (21–23):

(21) W moim samochodzie **rozbito** (smash$_{PST.IMPRS}$) szyby, bo mam niemiecką rejestrację.

'The windows of my car were smashed in because I have a German license plate.'

(22) W meksykańskich restauracjach **gotuje się** (cook$_{3SG.REFL}$) dania, które są bardzo smaczne.

'In Mexican restaurants they cook dishes that are very tasty.'

(23) Na wfie w gimnazjum **rozwiązali** (untangle$_{3PL.PST}$) skakanki, ponieważ jakiś głuptas zrobił żart i je wszystkie poplątał.

'In PE at the high school, they were untangling the jumping ropes because some moron had made a practical joke and knotted them all up.'

We did not test any reflexive verbs, since in Polish the reflexive impersonal cannot be formed from inherently reflexive verbs (Siewierska 1988: 265).

The factors to be tested were varied as follows: The tense was either present or preterite (the future was not tested). The aspect of the verb was switched between perfective (only in the preterite) and imperfective. For register, some items were created to contain words marked as colloquial, like (23), whereas the rest was in a neutral register. For the universal/non-universal reading, we only created test items with a clearly generic reading, as in (22), or with a clearly specific reading, as in (21) or (23). (In the following we will therefore refer to this variable as *genericity*.) This set of variables allows for 6 × 3 different combinations in the preterite and 4 × 2 combinations in the present tense, since the *-no/-to* impersonal and the perfective aspect only occur in the preterite (see Table 1) and the generic reading does not readily agree with the perfective aspect.

5.2. Design of the Questionnaires

The test items were distributed over five different questionnaires, each containing 16–19 test items and an equal number of fillers (positive and negative control items), all of which were pseudo-randomized. Due to the multiplicity of factors to be examined, it was not possible to represent all factors equally in all the questionnaires. For example, if a questionnaire is balanced for tense (50% present, 50% preterite), it cannot at the same time also be balanced for aspect and include items in the imperfective preterite. Therefore, each of the

Table 1. Possible combinations of factors across ARBs

Tense	Reference	Aspect	Register	ARB
preterite	generic	imperfective	neutral	*-no/-to*, reflexive, 3PL
preterite	generic	imperfective	colloquial	*-no/-to*, reflexive, 3PL
preterite	specific	imperfective	neutral	*-no/-to*, reflexive, 3PL
preterite	specific	imperfective	colloquial	*-no/-to*, reflexive, 3PL
preterite	specific	perfective	neutral	*-no/-to*, reflexive, 3PL
preterite	specific	perfective	colloquial	*-no/-to*, reflexive, 3PL
present	generic	imperfective	neutral	reflexive, 3PL
present	generic	imperfective	colloquial	reflexive, 3PL
present	specific	imperfective	neutral	reflexive, 3PL
present	specific	imperfective	colloquial	reflexive, 3PL

questionnaires was created to examine a certain factor while keeping other factors constant, so that as a fallback strategy it would be possible to analyze each questionnaire separately. However, in the end we decided to evaluate all test items from all five questionnaires together and take care of confounding factors with statistical methods.

The participants were asked to rate the acceptability of each item on a six-point Likert scale ranging from "– – –" (completely unacceptable) to "+ + +" (completely acceptable). For the statistical analysis, the rating categories were converted into an equidistant numerical scale from –1 to +1 (i.e., –1.0, –0.6, –0.2, +0.2, +0.6, +1.0). We decided to choose an even number of rating categories to force participants to at least indicate a tendency in their rating and avoid indifferent "I don't know" answers. It was also possible to give a free-text comment on every test item.

Together with some questions about the sociolinguistic background of the test participants, the five questionnaires were made available online via soscisurvey.de. After the introductory questions were answered, an urn-drawing mechanism decided which of the five questionnaires would be shown. The link to the online test was sent to colleagues in Poland, Germany, and Austria with the request to distribute it among their students, friends, and colleagues.

5.3. Participants

The five questionnaires were filled out by a total of 298 people. Since six people responded in the negative regarding whether Polish was their native language, and another three did not give any answer, we only used the re-

maining 289 questionnaires. Each of the five individual questionnaires was answered by 53 to 65 people.

Of the test participants, 70% were female and 30% male, while 96% were between 19 and 60 years old. The 289 participants represent all 16 Polish voivodeships, with a 26% majority from Mazovia; 13% live outside Poland (and were not excluded from the analysis since they indicated Polish as their native language). An overwhelming 81% majority holds an academic degree, and another 16% are currently enrolled at a university, so that the level of education is clearly not representative of the population, which should be kept in mind because it might have an influence on the test results.

6. Results

Statistics were conducted in R, version 3.5.1, using the *lme4* library (Bates et al. 2015) to perform a linear mixed-effects model (LMEM) of the acceptability of Polish ARBs under the influence of tense, aspect, register, and genericity. Depending on which factor we tested for the acceptability of the construction, we entered the remaining factors as fixed effects. The variability of subjects and items was taken into account by including them as random intercepts. Throughout the paper, we present *p*-values that are considered significant at the $\alpha = 0.001$ level. For the visualization of the data, we use boxplots, which are based on the five-number summary of the dataset. The whiskers indicate the total value range of the dataset (which in our large datasets generally includes both the minimum and the maximum value). The box encompasses the second and third quartiles of the dataset. The black horizontal line, which separates the quartiles, denotes the median of the dataset, while the circle plots the arithmetic mean. In general, the larger the box, the greater the dispersion of the data.

6.1. Register

As is apparent from Figure 3, the reflexive impersonal scored notably better in informal items, but unexpectedly, *-no/-to* is not rated better in formal style, although in grammars it is usually treated as formal. A different picture emerges if we take a closer look at the ARBs in the present tense (see Figure 4). Here the 3PL impersonal is rated significantly better in informal style ($p < 0.001$), and there are no significant differences for the reflexive impersonal. These results confirm our prediction that the 3PL impersonal is marked as colloquial, whereas the reflexive impersonal is applicable in all styles, even though it was formulated independently from tense.

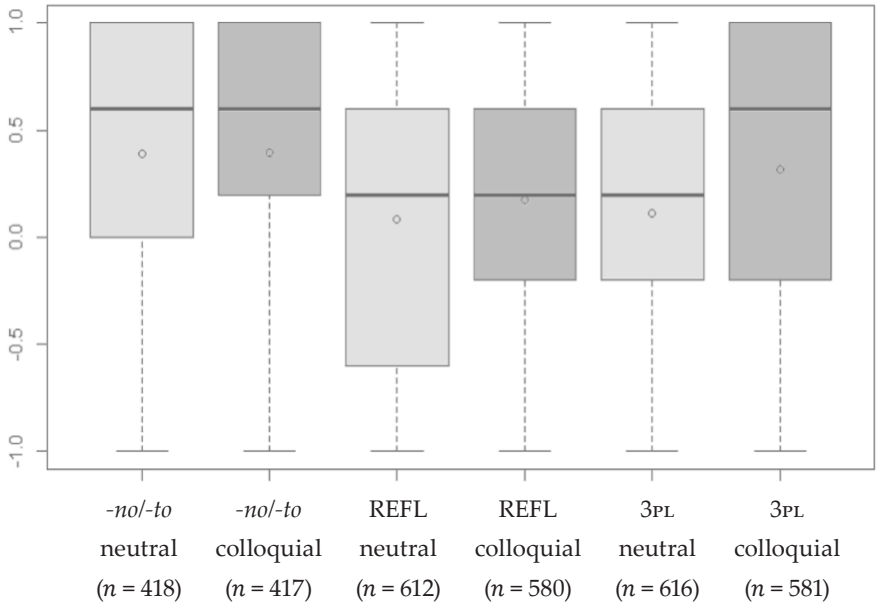


Figure 3. Colloquial vs. neutral register

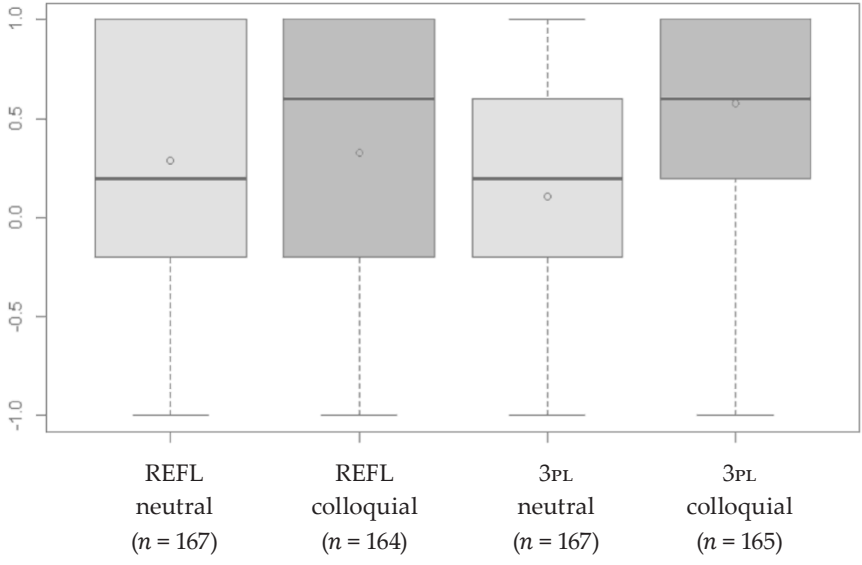


Figure 4. Colloquial vs. neutral register (present tense only)

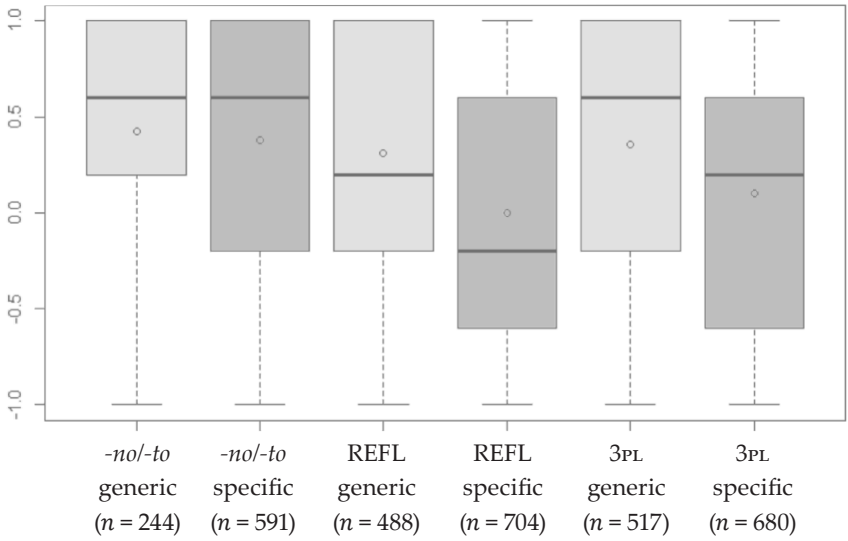


Figure 5. Generic vs. specific reading

6.2. Genericity

Visual inspection of Figure 5 above suggests that the reflexive impersonal is rated considerably better with generic than with specific readings. This is in line with our prediction, which was based on the results from the corpus query where there was a preference for *-no/-to* with an arbitrary reading versus the reflexive impersonal with a generic reading. However, after applying the mixed-effects model with confounding factors, the differences turned out to be statistically insignificant.

6.3. Aspect

Acceptability ratings for ARBs as a function of verbal aspect are shown in Figure 6. The statistical analysis proves that the reflexive impersonal is significantly more acceptable in the imperfective than in the perfective aspect ($p < 0.001$). This confirms our assumption that aspect might have an influence on the acceptability of particular ARBs.

However, the perfective aspect causes the situation to be read as specific, and Figure 5 has shown that specific readings render the reflexive impersonal (at least numerically) less acceptable. So, in order to isolate the effect of aspect from that of genericity, we excluded the items with generic reference. The results are shown in Figure 7 on the opposite page. Even with specific reference,

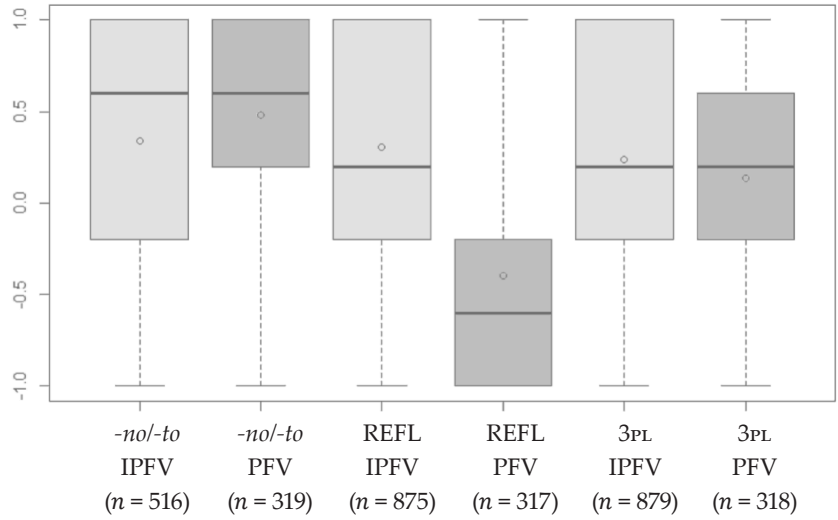


Figure 6. Imperfective vs. perfective aspect

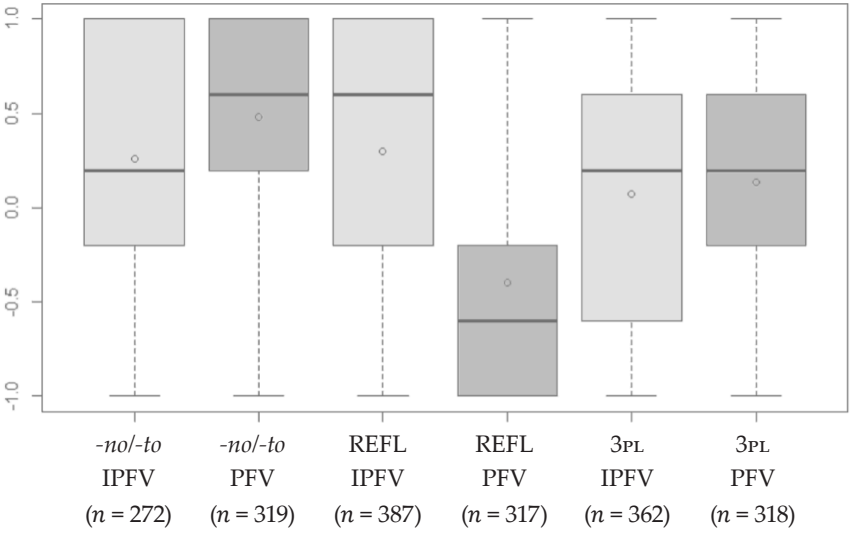


Figure 7. Imperfective vs. perfective aspect (specific only)

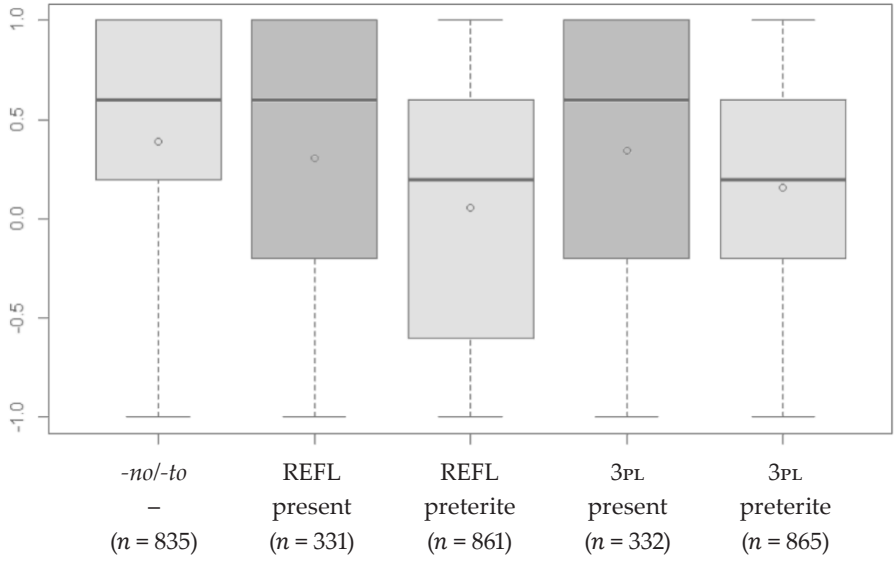


Figure 8. Present vs. preterite

the effect of aspect on the reflexive impersonal is still significant ($p < 0.001$). The reverse effect on the *-no/-to* construction, which is better in the perfective aspect, is more pronounced here, but still not significant.

6.4. Tense

The ratings for the test items depending on tense are shown in Figure 8 above. The reflexive and the 3PL impersonal are numerically more acceptable in the present tense than in the preterite. This would corroborate our prediction that these two constructions are less suitable in the preterite, where they are in competition with the *-no/-to* impersonal. However, the statistical analysis reveals no significant effects.

7. Conclusion

It is evident that the three competing Polish ARB constructions are selected on the basis of a rather complex multifactorial situation. While the four possible factors we examined render the situation almost too complex for an acceptability judgment test, one can easily think of further factors that might play a role: e.g., clusivity (cf. Krzek 2015), verb class (transitive, unaccusative, unerga-

tive, reflexive, etc.), negation, definiteness or specificity of the object, etc. However, we have found unambiguous evidence that the following factors play a role in the acceptability of the Polish ARBs. As expected, the 3PL impersonal is better in colloquial discourse, but the *-no/-to* construction shows no significant effect of register. The reflexive impersonal is much more acceptable with the imperfective than with the perfective aspect. The effects of specific vs. generic reading (where both reflexive and 3PL impersonal seem to prefer the generic reading) and of tense (where the same two ARBs seem to prefer present tense), as well as the slight tendency of the *-no/-to* construction towards the perfective aspect, are too small to be significant.

The curious fact that the *-no/-to* construction turned out to be insensitive to register might be explained by the unusually high level of education of our test participants. It is probably true that such highly educated speakers use the *-no/-to* impersonal even in everyday speech. In this case, the classification of this construction as "formal" or "bookish" by the grammars is actually based on a sociolect rather than a register.

Sources

[NKJP] *National corpus of Polish.* Available at: nkjp.pl. Last accessed 1 February 2022.

ParaSol, a corpus of Slavic and other languages. (2006–) Ruprecht von Waldenfels and Roland Meyer. Available at: www.parasolcorpus.org. Last accessed 30 November 2019.

References

Bartnicka, Barbara and Volkmar Lehmann. (2004) "Der Satz". Barbara Bartnicka, Björn Hansen, Wojtek Klemm, Volkmar Lehmann, and Hanna Satkiewicz, eds. *Grammatik des Polnischen*. Munich: Otto Sagner, 487–591.

Bates, Douglas, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. (2015) "Fitting linear mixed-effects models using lme4". *Journal of statistical software* 67(1): 1–48. DOI 10.18637/jss.v067.i01.

Bauer, Anastasia. (this volume) "Impersonalization strategies in Slavic: A contrastive corpus study based on the data from six Slavic languages". *Journal of Slavic linguistics* 29(2): 123–78.

Bunčić, Daniel. (2018) "Impersonal constructions in Slavic languages and the agentivity of the verb". Sebastian Kempgen, Monika Wingender, and Norbert Franz, eds. *Deutsche Beiträge zum 16. Internationalen Slavistenkongress, Belgrad 2018*. Wiesbaden: Harrassowitz, 101–10.

————. (2019) "Agent prominence in the Polish *-no/-to* construction". Anastasia Bauer and Daniel Bunčić, eds. *Linguistische Beiträge zur Slavistik.* Vol. 24.

JungslavistInnen-Treffen in Köln, 17.–19. September 2015. Frankfurt am Main: Peter Lang, 63–76.

————. (2020) "Agent prominence and movement in Russian third person plural impersonals ('indefinite-personal sentences')". *Die Welt der Slaven* 65(2): 297–311.

Cabredo Hofherr, Patricia. (2003) "Arbitrary readings of 3pl pronominals". Matthias Weisgerber, ed. *Proceedings of Sinn und Bedeutung (SuB) 7*. Konstanz: Universität Konstanz, 81–94.

Cinque, Guglielmo. (1988) "On *si* constructions and the theory of arb". *Linguistic inquiry* 19: 521–83.

Dowty, David R. (1991) "Thematic proto-roles and argument selection". *Language* 67(3): 547–619.

Egerland, Verner. (2003) "Impersonal pronouns in Scandinavian and Romance". *Working papers in Scandinavian syntax* 71: 75–102.

Fehrmann, Dorothea, Uwe Junghanns, and Denisa Lenertová. (2010) "Two reflexive markers in Slavic". *Russian linguistics* 34: 203–38.

Fellerer, Jan. (2008) "Kilka uwag o płaskiej koncepcji gramatyki języka polskiego jako obcego". Władysław T. Miodunka and Anna Seretny, eds. *W poszukiwaniu nowych rozwiązań: Dydaktyka języka polskiego jako obcego u progu XXI wieku*. Cracow: Wydawnictwo Uniwersytetu Jagiellońskiego, 91–101.

Gast, Volker and Johan van der Auwera. (2013) "Towards a distributional typology of human impersonal pronouns, based on data from European languages". Dik Bakker and Martin Haspelmath, eds. *Languages across boundaries: Studies in memory of Anna Siewierska*. Berlin: de Gruyter Mouton, 119–58.

Haider, Hubert. (2019) "Grammatical rules are discrete, not weighted, and not vulnerable". Ken Ramshøj Christensen, Henrik Jørgensen, and Johanna L. Wood, eds. *The sign of the V—Papers in honour of Sten Vikner*. Aarhus: Aarhus University, 205–26. DOI 10.7146/aul.348.96.

Himmelmann, Nikolaus P. and Beatrice Primus. (2015) "Prominence beyond prosody". Amedeo De Dominicis, ed. *pS-prominenceS: Prominences in Linguistics. Proceedings of the international conference*. Viterbo: DISUCOM Press, 38–58.

Janic, Katarzyna. (2013) *L'antipassif dans les langues accusatives*. Ph.D. thesis. Lyon: Université Lumière Lyon 2.

Kibort, Anna. (2008) "Impersonals in Polish: An LFG perspective". *Transactions of the Philological Society* 106(2): 246–89.

Krzek, Małgorzata. (2015) "Clusivity of the null subject pronoun in the Polish *-no/-to* construction" [slides of a presentation]. Available at: https://docgo.net/clusivity-of-the-null-subject-pronoun-in-the-polish-no-to-construction. Last accessed 1 February 2022.

Laskowski, Roman. (1984) "Kategorie morfologiczne języka polskiego—charakterystyka funkcjonalna". Renata Grzegorczykowa, Roman Laskowski, and Henryk Wróbel, eds. *Morfologia (Gramatyka współczesnego języka polskiego 2)*. Warsaw: PWN, 121–70.

Malamud, Sophia A. (2013) "(In)definiteness-driven typology of arbitrary items". *Lingua* 126: 1–31.

Malchukov, Andrej and Akio Ogawa. (2011) "Towards a typology of impersonal constructions: A semantic map approach". Andrej Malchukov and Anna Siewierska, eds. *Impersonal constructions: A cross-linguistic perspective*. Amsterdam: John Benjamins, 19–56.

Mel'čuk, Igor' A. (1974) "O sintaksičeskom nule". Aleksandr A. Xoldovič, ed. *Tipologija passivnyx konstrukcij: Diatezy i zalogi*. Leningrad: Nauka, 343–61.

Padučeva, Elena V. (2012) "Neopredelenno-ličnoe predloženie i ego podrazumevaemyj sub"ekt". *Voprosy jazykoznanija* 2012(1): 27–41.

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk, eds. (2012) *Narodowy korpus języka polskiego*. Warsaw: Wydawnictwo naukowe PWN.

Puzynina, Jadwiga. (1993) "Die sogenannten 'unbestimmt-persönlichen' Formen in der polnischen Sprache". Gerd Hentschel and Roman Laskowski, eds. *Studies in Polish morphology and syntax*. Munich: Otto Sagner, 31–61.

Rivero, María Luisa and Milena Milojević Sheppard. (2003) "Indefinite reflexive clitics in Slavic: Polish and Slovenian". *Natural language & linguistic theory* 21: 89–155.

Siewierska, Anna. (1988) "The passive in Slavic". Masayoshi Shibatani, ed. *Passive and voice*. Amsterdam: John Benjamins, 243–89.

————. (2008a) "Introduction: Impersonalization from a subject-centred vs. agent-centred perspective". *Transactions of the philological society* 106(2): 115–37.

————. (2008b) "Ways of impersonalizing: Pronominal vs. verbal strategies". María de los Ángeles Gómez González, J. Lachlan Mackenzie, and Elsa González Álvarez, eds. *Current trends in contrastive linguistics: Functional and cognitive perspectives*. Amsterdam: John Benjamins, 3–26.

Skibicki, Monika. (2016) *Polnische Grammatik*. 2nd edition. Hamburg: Buske.

Vinogradov, Viktor V. (1954) *Grammatika russkogo jazyka*. Vol. 2. Part 2. Moscow: Izdatel'stvo Akademii nauk SSSR.

von Waldenfels, Ruprecht. (2006) "Compiling a parallel corpus of Slavic languages: Text strategies, tools and the question of lemmatization in alignment". Bernhard Brehmer, Vladislava Ždanova, and Rafał Zimny, eds. *Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV).* Vol. 9. Munich: Sagner, 123–38.

Wiemer, Björn. (1995) "Zur Funktion von Sätzen ohne grammatisches Subjekt im Polnischen und Russischen (Auszüge aus einer vergleichenden Corpu-

sanalyse zu einer ausgewählten Textsorte)". *Wiener slawistischer Almanach* 35: 311–29.

Wiese, Eva. (1973) "Zur Analyse polnischer syntaktischer Konstruktionen aus dem Bereich der 'unpersönlichen' oder 'unbestimmt-persönlichen' Sätze". *Zeitschrift für Slawistik* 5: 625–33.

Maria Katarzyna Prenner
Slavic Department
Giessen University
Giessen, Germany
maria.prenner@slavistik.uni-giessen.de

Daniel Bunčić
Slavic Department
University of Cologne
Cologne, Germany
daniel.buncic@uni-koeln.de

Why *dwóch panów przyszło*, but *dwaj panowie przyszli* and *dwie kobiety przyszły*? Agreement with Quantified Subjects in Polish versus Russian and Bosnian/Croatian/Serbian*

Katrin Schlund

Abstract: In Russian, agreement with quantified subjects varies between plural (= semantic) and singular (= grammatical, default, impersonal) agreement, and there is ample evidence that this variation is governed by semantic and pragmatic factors (such as topicality and animacy of the subject). Although Bosnian/Croatian/Serbian follows stricter normative rules, variation does occur and is motivated similarly to Russian. Polish seems at odds with the paradigm of these languages. First, the grammar of contemporary Polish does not allow for variation in agreement with quantified subjects. Second, semantic agreement is available only with non-virile nouns in paucal numbers, while virile nouns require grammatical agreement (e.g., *dwie kobiety przyszły*$_{PL}$ 'two women came' but *dwóch mężczyzn przyszło*$_{SG}$ 'two men came'). This paper offers a way to integrate the Polish data into the Russian and Bosnian/Croatian/Serbian picture by drawing on historical and contemporary empirical evidence. Specifically, it offers a short analysis of variation between the nominative and oblique masculine forms of paucal numbers (*dwaj* vs. *dwóch*).

Keywords: semantic agreement, grammatical agreement, non-canonical subjects, differential subject marking, oblique subjects, quantified subjects, number phrases, Russian, Bosnian/Croatian/Serbian, Polish, *dwaj* vs. *dwóch*

1. Introduction

Numeral phrases are special in many languages, and this is particularly true of the Slavic language family. One phenomenon typical of languages all over the world is a grammatical distinction between the lowest few numbers and all higher numbers (e.g., Heine 1997: 32–34), as is characteristic of Slavic, where

* This paper profited immensely from the diligent work of two anonymous reviewers, whom I would like to thank here. Great thanks also go to my highly esteemed colleagues Daniel Bunčić (Cologne) and Maria Katarzyna Prenner (Giessen) for reading through earlier versions of this paper and giving valuable advice. Any remaining errors are my own.

the two sets of numbers evolved from different sources and show different morphosyntactic patterns up to the present day.

This paper is concerned with agreement resolution in phrases with quantified subjects (henceforth, QSs)—i.e., in phrases whose subjects include a noun quantified by a number. More precisely, the paper addresses the question of why variation between plural (semantic) agreement and singular (default, invariant, impersonal)[1] agreement is regularly available in Russian, marginally available in Bosnian/Croatian/Serbian (henceforth, BCS), and absent from Polish. The following examples illustrate the variation observable in Russian (1a–e)[2], BCS (2a–b), and Polish (3a–b):[3]

(1) Russian

a. Izvestno, skol'ko čelovek pogiblo na «Titanike»
known how.many people died$_{N.SG}$ on Titanic
'It is known how many people died on the Titanic.'
(*Kollekcija anekdotov*, 1970–2000)

b. [Š]est' čelovek pogibli, šestnadcat' propali
six people died$_{PL}$ sixteen disappeared$_{PL}$
bez vesti.
without news
'Six people have died, sixteen have disappeared without a trace.'
("Blogi", *Russkij reporter*, 2012)

c. Imeetsja dve kučki kamnej.
has$_{3SG.REFL}$ two piles stones$_{GEN}$
'There are two piles of stones.' (*Sbornik olimpiadnyx*…, 2016: 127)

d. V Tuve šest' otmorozkov ubili suprugov v
in Tuva six thugs$_{GEN}$ killed$_{3PL}$ couple$_{ACC}$ in
ix že kvartire.
their PART apartment
'Six thugs killed a couple in their apartment in Tuva.'
("V Tuve šest'…", *A42.RU*, 2019)

[1] As noted by Marušič and Nevins (2010: 315), grammatical agreement with QSs is actually "failure to agree at all".

[2] Examples (1a–b) were retrieved through the Russian National Corpus.

[3] Grammatical glosses follow the Leipzig Glossing Rules, with the addition of the following abbreviations: ACT—active; ENUM—enumerative (a special enumerative form, the so-called *brojna forma* of masculine nouns with paucal numbers in BCS); NONVIR—non-virile gender; VIR—virile gender.

e. Každyj god mnogo devušek i ženščin
every year many girls$_{GEN}$ and women$_{GEN}$
nadevajut ètot kostjum.
put.on$_{3PL}$ this costume
'Every year, a lot of girls and women wear this costume.'
(*Sneguročka + 1*, 2019)

(2) BCS

a. Došle (došla) su četiri mudre
came$_{FEM.PL}$ came$_{ENUM}$ AUX$_{3PL}$ four wise
žene (mudra muškarca).[4]
women$_{NOM.PL}$ wise$_{ENUM}$ men$_{ENUM}$
'Four wise women (men) came.'

b. Došlo je pet mudrih žena (muškaraca).
came$_{N.SG}$ AUX$_{3SG}$ five wise women$_{GEN.PL}$ men$_{GEN.PL}$
'Five wise women (men) came.'

(3) Polish

a. Przyszły trzy kobiety / psy
came$_{PL.NONVIR}$ three$_{NOM/ACC}$ women$_{NOM}$ dogs$_{NOM}$
'Three women/dogs came.'

b. Przyszło trzech mężczyzn
came$_{N.SG}$ three$_{GEN/ACC}$ men$_{GEN/ACC}$
'Three men came.'

Russian shows variation in grammatical versus semantic agreement with paucal (2, 3, 4) and non-paucal (≥ 5) numbers[5], with a tendency towards plural agreement with paucals and default singular agreement with non-paucals and unspecific quantifiers. The factors underlying this variation are understood quite well. By and large, the likelihood of plural agreement increases along with the subject-like semantic and pragmatic properties of the QS (such as animacy, topicality, etc.). BCS has a stricter set of rules for paucal versus non-paucal numbers, with semantic agreement in the paucal, and grammati-

[4] As in other Slavic languages, masculine nouns in BCS have a special form (*brojna forma*) with paucal numbers. Cf. also footnote 6.

[5] Marušič and Nevins (2010) refer to non-paucal numbers as "5&UP"—a label that has been taken up by other researchers (e.g., Willim 2015).

cal (impersonal) agreement in the non-paucal.[6] However, variation is observable in BCS as well, as will be shown in Section 3. Polish has the strictest set of rules with respect to the resolution of agreement with QSs. With paucal numbers, semantic agreement is available (and obligatory) for non-virile nouns, and for the rarely used nominative virile forms (*dwaj*, *trzej*, *czterej* 'two, three, four'), while grammatical agreement is obligatory for the oblique forms of paucal numbers (*dwóch*, *trzech*, *czterech* 'two, three, four') with virile nouns. Non-paucal numbers impose grammatical agreement by default, regardless of the gender of the quantified noun or of the semantic and pragmatic subject-like properties of the QS.

The choice of languages discussed here is not accidental. First of all, these languages represent the eastern, western, and southern groups of Slavic languages. Moreover, the three languages represent a continuum with respect to the resolution of agreement with QSs, with Russian (and all of East Slavic) showing the greatest variation between semantic and grammatical agreement (Corbett 2004: 215; Suprun 1969: 179), significantly less variation in BCS[7] (Corbett 1983: 147; Corbett 2004: 215), and almost no variation tolerated in Polish and other West Slavic languages (cf. Corbett 2004: 215; Suprun 1969: 200). The situation in Polish poses a challenge for functional analysis along the lines established for the analysis of QS agreement resolution in Russian (e.g., Corbett 1983: 136–56; Robblee 1993; Schlund 2018: 149–55) and, to a lesser extent, in BCS (Schlund 2019). This is because semantic agreement is available only for paucal numbers of non-virile nouns (3a) in Polish, but unavailable for the paucal numbers of virile nouns (3b). If anything, one would expect semantic agreement to combine more easily with virile nouns because virile nouns are semantically more salient and come closer to the subject prototype of a (male) human agent (Janda 1999: 201).

[6] Russian and BCS have a special morphological pattern for masculine nouns in the paucal that evolved from the dual but is synchronically identified as a genitive singular or as a special enumerative form. In Russian, the special form is restricted to the noun phrase, whereas it extends over the participle in the perfect tense in BCS:

(i) Russian

Pribyli	dva	učenika.
arrived$_{PL}$	two	pupils$_{ENUM}$

(ii) BCS

Stigla	su	dva	učenika.
arrived$_{ENUM}$	AUX$_{3PL}$	two	pupils$_{ENUM}$

Both: 'Two pupils arrived.'

[7] The South Slavic languages do not represent a homogeneous group in this respect. Bulgarian and Macedonian have generalized semantic/plural agreement over all numbers, whereas Slovene uses the dual and plural with paucal numbers and the singular with non-paucal ones (Corbett 2004: 215).

Sections 2 and 3 briefly revisit the factors governing variation between semantic and grammatical agreement with QSs in Russian (§2) and BCS (§3). Section 4 sketches the origins of Slavic numbers and the historical development of their agreement with QSs. Section 5 is dedicated to Polish and focuses on three questions: (1) Why is there no variation in QS agreement resolution in contemporary Polish? (2) Which factors govern the variation in morphological forms of virile paucal numbers between the oblique form requiring default agreement and the (now only rarely used) nominative form requiring plural agreement (e.g., *trzech panów przyszło* vs. *trzej panowie przyszli* 'three gentlemen came')? (3) What is the morphosyntactic status of the oblique virile forms—that is, what is their morphological case and what kinds of subjects are they (canonical, non-canonical, or oblique; cf. Seržant 2013; Schlund 2018)?

In search of answers to these questions, I will first give an overview of the historical resolution of agreement with QSs in general (§5.1) and then focus on the evolution of the virile numbers (§5.2). Section 5.3 explores the variation between oblique and non-oblique types of virile QSs in contemporary Polish. Section 5.4 argues for an analysis of the more frequent oblique type as oblique subjects, though with a full range of behavioral subject properties. This will help to explain why semantic agreement is unavailable precisely with virile QSs in paucal numbers, despite the fact that the semantic properties of virile nouns arguably resemble those of a prototypical subject more closely than those of non-virile nouns.

2. Variation in Agreement with QSs in Russian

Russian QSs show variation in whether they prompt singular (= grammatical, impersonal, default) or plural (= semantic) agreement. A huge body of research[8] on this topic has established the following factors influencing QS agreement resolution in Russian (Corbett 1983: 142–56):

- Animacy of the QS favors semantic agreement.
- Preverbal position of the QS favors semantic agreement.
- Paucal numbers (< 5) tend towards semantic agreement
- Non-paucal numbers (≥ 5) and unspecific indications of quantity (e.g., *neskol'ko* 'some', *mnogo* 'many') tend towards grammatical agreement.

Schlund (2018: 149–55) argues that the factors favoring semantic agreement in Russian boil down to the semantic and pragmatic properties of prototypical subjects. Semantically, a prototypical subject is an agent, and a prototypical semantic agent is not only animate but also human, an actor, and in full

[8] Suffice it here to refer to some studies of Corbett (1983, 2004, 2010) and to the summary of factors given in Franks 2009: 358.

control of the event described by the predicate. Pragmatically, a prototypical subject is a topic, and topics typically occur in pre-verbal position. Example (1c) above is a case in point. The subject is non-topical and inanimate, which is why grammatical agreement occurs although the number of referents is only two. Example (1d), on the other hand, contains a highly agentive subject, which is why it can receive semantic agreement although the number of referents is five.

A prototypical subject is also a clearly distinguishable entity in that it has an individual referent. Paucal numbers themselves resemble the subject prototype in the sense that we can intuitively and immediately distinguish the number of entities we see (or imagine) only in numbers up to five, in general. Also, we easily construe two, three, or four actors as a group of individual entities. In other words, the actors are still more salient than their number. This saliency is directly reflected in the grammar: paucal numbers are modifiers, functioning like attributive adjectives. Our ability to conceive of a number of entities as individual entities decreases as the number increases. The human ability to intuitively discern the number of entities in a group (i.e., to subitize)[9] decreases with non-paucal numbers. This corresponds to the grammatical behavior of non-paucal numbers, which, having developed from nouns, originally functioned as heads of quantified phrases and governed the counted entities.

The variation between semantic and grammatical agreement with Russian QSs is also reflected in the syntactic behavior of these subjects. While QSs inducing semantic agreement usually display all behavioral subject properties, QSs inducing grammatical agreement are more restricted.[10] Although all kinds of Russian QSs can be syntactic subjects (i.e., they can answer the question of *kto*$_{NOM}$/*čto*$_{NOM}$ 'who/what'), some of them, particularly those with unspecific quantifiers and non-paucal numbers, do not bear nominative case. These Russian QSs can be classified as non-canonical subjects, differing from canonical ones in that they lack morphological subject properties (nominal

[9] Subitizing, or more precisely, perceptual subitizing, is the ability to know how many entities are in a small set without actually counting. There is also conceptual subitizing, which denotes the ability of "recognizing smaller groups within a larger set and adding those small groups together, such as two dots plus two dots equals four dots, or three dots and three dots makes six dots" (Levin, n.d.).

[10] An illustrative case in point is a classic example from Franks 1995: 121, in which the binding of a reflexive pronoun is possible only in combination with semantic agreement:

(i) Pjat' ženščin smotreli / ??smotrelo na sebja.
five women$_{GEN}$ looked$_{PL}$ / looked$_{N.SG}$ on self
'Five women looked at themselves.'

case marking and agreement with the verbal predicate[11]), while at the same time functioning as subjects syntactically (Seržant 2013: 320; Schlund 2018: 123).

There is one rule about QS agreement resolution in Russian that deserves special attention. When a QS is preceded by the demonstrative pronoun *èti* 'these', plural agreement is obligatory.[12] As shown by examples (4) and (5), the rule applies even when the subject is inanimate and occurs with a non-paucal number (the subject of (4) is also in non-topical position):

(4) Nedavno postroeny i èti sem' domov.
recently built$_{PL}$ and these seven houses$_{GEN.PL}$
'These seven houses were also built recently.'
(Rozental', Džandžakova, and Kabanova 2005: 465)

(5) I èti 20 minut okazalis' samye
and these 20 minutes$_{GEN.PL}$ turned.out$_{PL}$ only$_{NOM.PL}$
ščastlivye za stol' dlitel'noe vremja.
happy$_{NOM.PL}$ for so.much long$_{NOM}$ time$_{NOM}$
'And these 20 minutes turned out to be the happiest ones for a very long time.' (*Sneguročka + 1*, 2019)

The demonstrative pronoun *ètot* 'this' increases the referentiality of the counted noun. A QS within a determiner phrase with the demonstrative *ètot* thus gains in its resemblance to a prototypical subject. It is noteworthy, though, that the prototypical subject property of animacy (and, associated with it, agentivity) is not decisive here, since the rule applies even to the inanimate subjects of passives (4) and of unaccusative verbs (5).

[11] Of course, agreement resolution (semantic or grammatical) varies depending on the subject properties outlined above.

[12] Note that this rule does not apply when the demonstrative and the number appear in reverse order. In that case, the demonstrative receives genitive-case marking and both singular and plural agreement become available (also Franks 1995: 101):

(i) Sem' (iz) ètix domov postroeny/o nedavno.
seven (of) these$_{GEN.PL}$ houses$_{GEN.PL}$ built$_{PL/SG}$ recently
'Seven of these houses were built recently.'

This example also illustrates the point made above. The reverse order of the demonstrative and the number yields a slight decrease in referentiality since a specific number of items is singled out of a larger unit. This explains why singular default agreement becomes available. Note that the sentence is more acceptable when *iz* 'of' is included.

3. Less Variation in Agreement in BCS

In BCS the agreement behavior of a QS is normatively determined. The rule states that grammatical (= impersonal, default) agreement is obligatory for non-paucal numbers and unspecific quantifiers such as *nekoliko* 'some', *malo* 'few', and *mnogo* 'many'. Paucal numbers require either plural agreement or agreement in a specific masculine enumerative form that originated from the now lost dual (see example (2) on p. 223). Deviation from this rule is rare but does occur, and it is likewise influenced by animacy and word order (Corbett 1983: 146–50; Schlund 2019). Examples (6), (7b), and (8b) illustrate deviations from the rule, cases in which plural agreement occurs in combination with a non-paucal QS:

(6) Naših 5 momaka su bez problema
our$_{GEN.PL}$ 5 boys$_{GEN.PL}$ AUX$_{3PL}$ without problems

prošli u 3. k[rug] takmičenja.
passed$_{M.PL}$ in 3rd round competition$_{GEN}$

'Our five boys entered the third round of the competition without any problems.' (Schlund 2019: 177)

(7) a. Sedam boraca ISIL-a, šest žena i 12 djece
seven fighters of_ISIL six women and 12 kids

stiglo u BiH
arrived$_{N.SG}$ in BiH

'Seven IS-fighters, six women, and twelve children have arrived in Bosnia-Herzegovina.'

b. Šest žena i 12 djece, po dolasku na
six women and 12 kids after arrival at

Međunarodni aerodrom Sarajevo su prošli
international airport Sarajevo AUX$_{3PL}$ passed$_{M.PL}$

odgovarajuće provjere od strane Granične
according checks from side border$_{ADJ}$

policije BiH […].
police BiH

'Upon arrival at the Sarajevo International Airport, the six women and twelve children passed the appropriate checks conducted by the Herzegovinian border police.'

("Grupa bh. državljana vraćena iz Sirije", *Etleboro.org*, 2019)

(8) a. Najstarije na svijetu: Ovih pet žena
oldest$_{NOM.F.PL}$ on earth these$_{GEN.PL}$ five women$_{GEN.PL}$
rođeno je u 19. stoljeću
born$_{N.SG}$ AUX$_{3SG}$ in 19th century
'Oldest in the world: These five women were born in the 19th century' [headlines of a newspaper article]

b. Ovih pet dama znaju tajnu […].
these$_{GEN.PL}$ five ladies$_{GEN.PL}$ know$_{3PL}$ secret
'These five ladies know a secret….'
("Najstarije na svijetu", *Klix*, 2015)

According to the rule, there should be default singular agreement in all these examples because the numbers involved are greater than or equal to five. However, semantic (i.e., plural) agreement occurs in (6), (7b), and (8b). All of the QSs in examples (6–8) are animate and function as the topics of their clauses, but only the subjects of (6) and (8b) can be assigned the semantic role of agent. Therefore, agentivity cannot be the crucial factor allowing for semantic agreement in (7b), which is lower in agentivity. What seems to be important, though, is that the subjects in (7b) and (8b) are definite. The definiteness of the subjects in (7b) and (8b) anaphorically relate to their first mention in the text. The definiteness of the QS in (7b) is clarified in the English translation, which obligatorily includes the definite article. Definite and possessive contexts increase referentiality, and it seems that this increase is crucial for the possibility of semantic agreement with QSs in BCS. This finds a correlate in the rule that QSs preceded by the definite pronoun *ètot* require semantic agreement in Russian (cf. §2). QSs in BCS can also be classified as non-canonical subjects in the sense of Seržant 2013 and Schlund 2018; non-canonical subjects function as syntactic subjects and display behavioral subject properties (e.g., binding reflexive pronouns, raising) but may or may not induce agreement of the predicate.

4. A Glance at the Diachrony of Agreement with QSs

Slavic paucal numbers originally functioned as attributes, agreeing with the nouns they quantified (Suprun 1969: 172). In Proto-Slavic, QSs with *dъva* (m.) / *dъvě* (f./n.) 'two' and *oba* (m.) / *obě* (f./n.) 'both' required dual agreement, whereas QS with *trьje* (m.) / *tri* (f./n.) 'three' and *četyre* (m.) / *četyri* (f./n.) 'four' required plural agreement (Suprun 1969: 61, 142; Gvozdanović 1999: 187f.).

Non-paucal numbers were historically feminine *i*-stem nouns[13] (Suprun 1969: 171), governing the case of the noun they quantified. As in modern Russian, BCS, and Polish, non-paucal numbers imposed a (partitive) genitive plural on the nouns they quantified. The fact that the non-paucal numbers originated from nouns also explains why they grammatically required singular agreement. However, for the impersonal (singular neuter) form of the finite predicate to combine with non-paucal numbers as syntactic subjects, it was necessary for the non-paucal numbers to gradually lose their morphological status as (feminine) nouns.[14]

The difference in the morphosyntax of smaller versus larger numbers is grounded in human perception. The larger a set of counted entities, the less it is possible (or necessary) to focus on the individual entities included in the set (see Siuciak 2008: 16; Garncarek 2018: 79). With large numbers, our focus naturally switches to the numerical value itself, which is why larger numbers are more "noun-like" than lower numbers (see Corbett 2004: 217). Very large numbers still behave like nouns even in the modern Slavic languages.[15] Paucal numbers, on the other hand, are modifiers of the quantified noun, with the noun remaining the focus of attention.

Suprun (1969: 171f.) assumes that there was fluctuation in agreement resolution, particularly of non-paucal QSs, in Slavic early on. However, in Old Church Slavonic, our oldest available written evidence, agreement with paucal numbers was consistently in the plural and dual, respectively, whereas singu-

[13] An exception was *desętъ* 'ten,' originally a masculine noun that later adapted to the feminine *i*-stem declension of the number nouns *pętъ* 'five' through *devętъ* 'nine' (Suprun 1969: 61f.). Moreover, big numbers belonged to other declension classes (e.g., *sъto* 'hundred').

[14] It is therefore more adequate to distinguish three possible kinds of agreement with QSs in Slavic: semantic agreement (*congruentia ad sensum*) in the plural; truly grammatical agreement (*congruentia ad formam*; referred to as syntactic agreement by Miechowicz-Mathiasen 2013), that is, singular feminine agreement reflecting the gender of the noun denoting the non-paucal number; and default agreement (3N.SG) (cf. Miechowicz-Mathiasen 2013: 81). However, since the original grammatical agreement vanished very early, I use the terms grammatical agreement, singular agreement, and default agreement synonymously. Example (i)—from *Biblia Leopolity* (1561)—illustrates one of the rare instances of original grammatical agreement from historical Polish data:

(i) siedem kłosów wyrastała
seven$_{NOM.SG}$ ears.of.wheat$_{GEN.PL}$ grew$_{F.SG}$
'seven ears of wheat have grown' (Łoś 1928: 98)

[15] For example, Polish *tysiąc*$_M$ 'thousand' tolerates both singular masculine and default—that is, singular neuter—agreement (Miechowicz-Mathiasen 2013: 97).

lar agreement prevailed with non-paucal numbers (Corbett 2004: 215).[16] It is therefore more likely that singular agreement was dominant with non-paucal numbers also in Proto-Slavic.

5. Polish: No Variation, No Motivation?

Polish, like the other West Slavic languages, shows very little variation in the agreement resolution of QSs (Suprun 1969: 200; Corbett 2004: 215). Non-paucal numbers and unspecific quantifiers impose default agreement, regardless of the gender of the quantified noun, or of any semantic or pragmatic factors. See, for example, (9–10) below:

(9) Siedem psów zaatakowało trzy wilki.
seven dogs$_{GEN.PL}$ attacked$_{N.SG}$ three$_{ACC}$ wolves$_{ACC}$
'Seven dogs attacked three wolves.'

(10) w tym pokoju śpi osiem kobiet/
in this room sleep$_{3.SG}$ eight$_{NONVIR}$ women$_{GEN.PL}$
ośmiu mężczyzn.
eight$_{VIR}$ men$_{GEN.PL}$
'There are eight women/eight men sleeping in this room.'

QSs with paucal numbers behave differently, depending on the gender of the quantified noun. Contemporary Polish has two genders in the plural: virile and non-virile. QSs with non-virile nouns require semantic agreement, as in (11). Importantly, and somewhat unexpectedly, virile nouns do not induce semantic agreement. Instead, the default neuter singular is obligatory, as in (12). This is because paucal virile QSs carry accusative-genitive morphology although they are syntactically subjects. The question is thus not merely why singular agreement is obligatory with a virile QS in the paucal, but also where the oblique case marking originates:

(11) Trzy psy zaatakowały wilka.
three dogs$_{NOM.PL}$ attacked$_{PL.NONVIR}$ wolf$_{ACC}$
'Three dogs attacked a wolf.'

(12) Trzech chuliganów zdemolowało samochody.
three$_{GEN/ACC}$ hooligans$_{GEN.PL}$ demolished$_{3N.SG}$ cars$_{ACC}$
'Three hooligans demolished cars.'

[16] Corbett's (2004: 215) data do not indicate whether singular agreement was in the (original) feminine gender or in the neuter.

Table 1 gives an overview of QS agreement resolution in contemporary Polish:

Table 1. QS agreement resolution in contemporary Polish

	Paucal	**Non-Paucal**
Non-Virile Gender	plural agreement	default agreement
Virile Gender	default agreement	default agreement

This fact of Polish grammar is difficult to integrate into the analysis developed on the basis of Russian and BCS and the historical information outlined above. After all, there is no reason to assume that QSs with non-virile nouns make semantically more prototypical subjects than QSs with virile nouns. Rather, one would assume that it would be the other way around, with virile nouns more closely resembling prototypical subjects than non-virile nouns (cf. Janda 1999: 201).[17]

Interestingly, there exists an alternative paucal form for virile nouns that is in the nominative case and induces semantic agreement, namely the forms *dwaj, trzej, czterej* mentioned in the introductory part of this paper. From a purely formal perspective, it is possible to change every instance of an oblique paucal subject with a noun of the virile subgender into its non-oblique counterpart. Predicate agreement, then, obligatorily switches from neuter singular to virile plural, as in (13) as compared to (12):

(13) Trzej chuligani zdemolowali samochody
three$_{VIR.NOM}$ hooligans$_{NOM}$ demolished$_{VIR.PL}$ cars$_{ACC}$
'Three hooligans demolished cars.'

The nominative form is hardly discussed in studies on QS agreement resolution in contemporary Polish. This is probably due to the somewhat obsolete character of these forms (cf. Długosz-Kurczabowa 2003, s.v. *dwa*). However, and as will be shown in more detail in Section 5.3, the nominative paucal numbers in the virile gender do occur in contemporary texts as well. The question is when the nominative virile paucal forms may still be favored over their (now dominating) oblique counterparts and how the two parallel sets of forms developed. Before addressing these questions, we will first take a look at QS agreement resolution in earlier stages of Polish and at the historical development of the virile paucal numbers in particular.

[17] This appears to be the case in Slovak, where virility triggers plural agreement also with numbers higher than five (Suprun 1969: 179).

5.1. Historical Agreement with Non-Paucal QSs in Polish

Miechowicz-Mathiasen (2013: 82–87) provides empirical evidence that agreement with non-paucal QSs fluctuated in Polish between plural and (neuter) singular agreement between the 15th and 17th centuries, with a vanishingly low proportion of singular feminine agreement (one instance out of 777 QSs) in the oldest sources she analyzes.[18] Her data also indicate an increase in singular neuter agreement over the investigated time period, from 1455 to 1632 (ibid. 87).

Since Miechowicz-Mathiasen (2013) is not concerned with factors underlying variation between singular and plural agreement with QSs, she does not distinguish the QSs found in her data with respect to animacy or agentivity, word order, or other factors. The following examples illustrate the range of variation. They were either taken from Miechowicz-Mathiasen 2013 or, when the historical text source is given, directly from the texts.[19] In these latter cases, the examples were not collected through a structured and exhaustive corpus analysis but rather retrieved by cursory queries for numbers in the respective texts.

Some examples (14–18)[20] suggest that animacy and preverbal (topical) position played a similar role as in contemporary Russian and, to a minor extent, contemporary BCS:

(14) dziesięć żon piec będą w jednem piecu chleby
ten wives$_{GEN.PL}$ bake$_{INF}$ will$_{3PL}$ in one stove bread$_{ACC.PL}$
'ten wives will bake bread in one stove'
(Miechowicz-Mathiasen 2013: 83)

(15) sześć set mężow […] stali przed drzwiami
six hundred$_{GEN.PL}$ men$_{GEN}$ stood$_{M/VIR.PL}$ in.front door$_{INST.PL}$
'six hundred men stood before the door(s)[21]'
(Miechowicz-Mathiasen 2013: 86)

[18] These sources are the *Biblia królowej Zofii* (Bible of Queen Sofija; 1455) and the *Rozmyślanie przemyskie* (Przemyśl meditation; ~1450).

[19] These texts were accessed through the *Korpus tekstów staropolskich* (corpus of Old Polish texts, *KTS*).

[20] Example (14) appears in the *Biblia królowej Zofii*, while (15) and (17) are taken from the *Biblia Gdańska* (1632).

[21] *Drzwi* 'door' is a plurale tantum also in modern Polish.

(16) pięć mądrych nabrały oleju w swoje lampy, ale
five wise$_{GEN.PL}$ took$_{NONVIR.PL}$ oil in their lamps but
pięć szalonych […] nie wzięły oleju z sobą
five foolish$_{GEN.PL}$ […] not took$_{NONVIR.PL}$ oil with them
'the five wise virgins put oil in their lamps, but the five foolish ones did not take oil with them' (*Rozmyślanie przemyskie*: 487, via *KTS*)

(17) z rzeki wychodziło siedem krów
from river stepped.out$_{N.SG}$ seven cows$_{GEN.PL}$
'seven cows were stepping out of the river'
(Miechowicz-Mathiasen 2013: 85)

(18) Szło też s nim piędziesiąt poczesnych mężow…
went$_{3N.SG}$ also with him fifty honorable$_{GEN.PL}$ men$_{GEN.PL}$
'With him went fifty honorable men…'
(*Rozmyślanie przemyskie*: 115, via *KTS*)

All of the QSs above are animate. Those in preverbal position induce plural agreement (14–16); the two examples in postverbal position display singular agreement (17, 18). However, not all illustrative examples given by Miechowicz-Mathiasen (2013) adhere to this pattern. For instance, in the following two almost identical examples from the *Biblia Brzeska* (1563), one displays singular and the other one plural agreement without obvious reason:

(19) było siedem braciej
was$_{N.SG}$ seven brothers$_{GEN.PL}$
'there were seven brothers' (Miechowicz-Mathiasen 2013: 86)

(20) byli tedy siedem braciej
were$_{M/VIR.PL}$ then seven brothers$_{GEN.PL}$
'then there were seven brothers' (Miechowicz-Mathiasen 2013: 86)

Variation was also available with inanimate QSs. Typically, inanimate QSs showed default agreement in post-verbal position but could also trigger plural agreement:

(21) było tamo sześć sędow kamiennych
were$_{N.SG}$ there six courts$_{GEN.PL}$ stone$_{ADJ.GEN.PL}$
'there were six courts made of stone'
(*Rozmyślanie przemyskie*: 208, via *KTS*)

(22) były tam sześć stągiew kamiennych
were$_{NONVIR.PL}$ there six waterpots$_{GEN.PL}$ stone$_{ADJ.GEN.PL}$
'there were six waterpots made of stone'
(Miechowicz-Mathiasen 2013: 83)[22]

The following examples, however, seem to be motivated:

(23) potem kiedy czternaście niedziel minęło
then when fourteen weeks$_{GEN.PL}$ passed$_{N.SG}$
'then, when fourteen weeks had passed'
(*Rozmyślanie przemyskie*: 12, via *KTS*)

(24) sześć dni są, w ktore mamy robić
six days are$_{3PL}$ in which have$_{1PL}$ work$_{INF}$
'there are six days in which to work'
(*Rozmyślanie przemyskie*: 376, via *KTS*)

Time data typically trigger singular agreement. This is because time data are bad candidates for prototypical subjects (even in preverbal position) and, as in the above case, tend to convey background information (note that example (23) is a subordinate clause). In (24), on the other hand, the QS is the center of attention and does not function as a specification of time but as the subject of an existential sentence.

Siuciak (2008: 175–88) provides converging evidence for these observations, noting that singular default agreement with non-paucal numbers was becoming ever more frequent from the 16th century onwards. Importantly, she also notes that semantic agreement was more persistent with virile nouns (ibid. 186).

With the paucal numbers *trzy* 'three' and *cztery* 'four' and non-virile nouns, plural agreement was the norm, as illustrated in (25):

(25) a na tem drągu były czaszki cztyrzy
and on that stick were$_{NONVIR}$ sculls$_{NOM.PL}$ four
'and there were four skulls on that stick'
(*Biblia królowej Zofii*, Exodus 37:20, via *KTS*)

[22] This example comes from the *Biblia Brzeska*.

5.2. The Rise of the Paucal Virile Numbers

5.2.1. Late Old Polish[23]

In Late Old Polish the virile subgender had not yet developed. Importantly, the genitive-accusative syncretism of virile nouns in the plural began precisely with numeral phrases (Miechowicz-Mathiasen and Dziubała-Szrejbrowska 2012: 2; Miechowicz-Mathiasen 2013: 89).[24] The first instances of genitive-accusative syncretism are attested in the dual[25] (15th c.), and they expanded into the plural in the course of the 16th and 17th centuries (cf. Miechowicz-Mathiasen and Dziubała-Szrejbrowska 2012: 6, and the references therein).

To find out when the specific virile, nominal, and oblique forms of the paucal numbers emerged, we searched the *Korpus tekstów staropolskich* (corpus of Old Polish texts, *KTS*). By the middle of the 15th century, we still find the masculine form *dwa* also with virile nouns, and when in syntactic subject po-

[23] I assume the traditional periodization of Polish: Old Polish: ca. 1150–ca. 1500; Middle Polish: ca. 1500–ca. 1780; Modern Polish: since 1780 (Klemensiewicz 2002).

[24] One reason why the new accusative syncretic with the genitive entered into the plural domain precisely with numeral phrases could be that the genitive was already present here with non-paucal numbers early on (Miechowicz-Mathiasen and Dziubała-Szrejbrowska 2012: 6–8). Examples such as (i), in which the counted entity precedes the paucal number, might have been particularly favorable for the dissemination of the genitive-accusative syncretism into the plural:

(i) Wybierzcie wy mężow lepszych a mocniejszych
choose$_{2PL}$ you$_{2PL}$ men$_{GEN/ACC}$ best$_{GEN/ACC}$ and strongest$_{GEN/ACC}$
dwanaście.
twelve
'You choose twelve of the best and strongest men.'
(*Rozmyślanie przemyskie*: 756, via *KTS*)

Typically, the number precedes the counted noun in numeral phrases, e.g., *dwanaście mężow* 'twelve men'. However, and particularly in spoken language, the counted noun may also precede the number—e.g., *widzę mężow dwanaście* lit. 'I see men twelve'. This inverse structure may arise due to human perception and focus of attention. Sometimes, we first perceive men (particularly when they appear in high numbers) and only then try to estimate their number. Therefore, we add the number only after the noun to be counted. This is also why in Russian the pre-position of the counted noun yields an estimate, not an exact number (e.g., *desjat' sobak* 'ten dogs' vs. *sobak desjat'* 'roughly ten dogs'). It is not implausible that instances of post-position of the number favored the spread of the genitive-accusative syncretism in numeral phrases with virile nouns.

[25] That the dual was affected by genitive-accusative syncretism before the plural has been explained by the fact that the intrusion of genitive-accusative syncretism started in the singular (Siuciak 2008: 85, and the references therein).

sition, with a predicate agreeing in the dual or plural (cf. (26) and (27)). At the same time, the old accusative ending for non-virile and virile nouns that is syncretic with the nominative (e.g., *miecze* 'swords$_{NOM/ACC}$' and *wszytki męże* 'all$_{NOM/ACC}$ men$_{NOM/ACC}$') was largely preserved, as illustrated in (27):

(26) aczby wadziła sie [*sic*] męża dwa
when quarrel$_{M.DU}$ REFL men$_{NOM.DU}$ two$_{NOM.M}$
'when two men quarrel' (*Biblia królowej Zofii*, Exodus 21:22, via *KTS*)

(27) dwa syny Jakobowa Symeon a Lewi […]
two$_{NOM.M}$ sons$_{NOM/ACC.DU}$ Jakob$_{POSS}$ Symeon and Lewi […]
wziąwszy miecze szli do miasta a
take$_{PST.PTCP.ACT}$ swords$_{ACC}$ went$_{M/VIR.PL}$ to town and
zbiłasta wszytki męże
killed$_{M.DU}$ all$_{ACC}$ men$_{ACC}$
'two of Jakob's sons, Symeon and Lewi, having taken their swords, went to the town and killed all the men'
(*Biblia królowej Zofii*, Genesis 34:25, via *KTS*)

When virile nouns combined with the other paucal numbers, three and four, they consistently imposed plural agreement on the predicate:

(28) wtenczas, cztyrze krolowie w żydowskiej ziemi
at that time four$_{M}$ kings$_{NOM.PL}$ in Jewish land
krolowali
ruled$_{M/VIR.PL}$
'at that time, four kings ruled in the Jewish land'
(*Rozmyślanie przemyskie*: 401, via *KTS*)

The form *cztyrze* used in (28) is a relic of the inherited masculine form **četyre*. Interestingly, the sentence following directly after example (28), given here as (29), renders the number four already as *cztyrzy*:

(29) cztyrzy krolowie w żydowskie [*sic*] ziemi byli
four kings$_{NOM.PL}$ in Jewish land were$_{M/VIR.PL}$
'there were four kings in the Jewish land'
(*Rozmyślanie przemyskie*: 401, via *KTS*)

Historically, *cztyrzy* derives from the syncretic form of the feminine and neuter genders (**četyri*), which gradually entered the masculine paradigm. In the

15th century, *cztyrzy* was available for nouns of all genders (e.g., *cztyrzy modlitwy/dni/miesiące/krolowie* 'four prayers/days/months/kings').[26]

There is also evidence of the virile number forms *czterej* and *trzej*, which existed alongside the older forms:[27]

(30) przyszli trzej krolowie
came$_{M/VIR.PL}$ three kings$_{NOM.PL}$
'three kings came' (*Rozmyślanie przemyskie*: 203, via *KTS*)

(31) [wiele s opuścił wypisania umęczenia bożego, co]
ini trzej ewanjeliści napełnili
others three$_{NOM.VIR}$ evangelists$_{NOM.PL}$ filled$_{VIR.PL}$
'[Saint John left out much about the sufferings of Christ that] the other three evangelists filled out' (*Rozmyślanie przemyskie*: 790, via *KTS*)

(32) a nieśli ji k niemu czterzej
and brought$_{M/VIR.PL}$ him$_{ACC}$ towards him$_{DAT}$ four$_{NOM.VIR}$
'and four (men) brought him to him (Jesus)'
(*Rozmyślanie przemyskie*: 292, via *KTS*)

There is no evidence of the nominative virile form *dwaj* in the entire corpus of Old Polish (*KTS*). Instead, *dwa* is used alongside the virile forms *trzej* and *czterej*:[28]

(33) W tych przytczach nie tylko dwa, trzej alibo
in these cases not only two$_{M}$ three$_{VIR}$ or
cztyrzej kmiecie wynić mogą
four$_{VIR}$ peasants go.out$_{INF}$ can$_{3PL}$
'In these cases, not only two, three or four peasants can go out'
(*Kodeks Świętosławów* 44:10–11; 1448–50, via *KTS*)

[26] There are two more instances of the old masculine form *cztyrze* used with a QS in the *Rozmyślanie przemyskie*, one of them referring to the four evangelists, the other to the masculine noun *tysiąc* 'thousand'. *Tysiąc* occurs many times as a subject in the *Rozmyślanie przemyskie*, and its agreement fluctuates greatly between singular masculine, singular neuter, and plural agreement.

[27] Siuciak (2008: 190) argues that the new virile forms *trzej* and *czterej* first occurred in mere determinations of quantities ("zdania o charakterze konstatacji liczbowych"). She also points out that the nominative forms were not restricted to virile nouns initially but were available also for animate nouns referring to males (ibid. 99).

[28] *Trzej* evolved from the masculine form *trьje* > *trzē* > *trzej* by means of contraction and diphthongization (Janda 1999: 218). The same applies to *czterej*.

There are no virile QSs with the oblique forms *dwóch*, *trzech*, and *czterech* in the Old Polish corpus. This means that the oblique forms must have penetrated into the subject domain later than the 15th century.

5.2.2. The 17th and 18th Centuries

There is no digital corpus comparable to the *KTS* available for the 16th century. Therefore, to find out when the oblique virile forms started creeping into the subject domain, research was conducted in the *Elektroniczny korpus tekstów polskich z XVII. i XVIII. wieku (do 1772 roku)* (electronic corpus of Polish texts from the 17th and 18th centuries (until 1772)), also referred to as the "Baroque corpus", covering the years 1601–1772 (Kieraś and Woliński 2018: 3854).

For the purposes of this study, only the manually annotated part of the corpus (containing around 511,000 segments[29]) turned out useful, as the automatically annotated corpus contained too many false positive cases of virile QSs. Fortunately, the oblique forms in subject position are tagged as nominatives in the corpus, which is why it has been possible to find all the paucal virile QSs tagged as such in the manually annotated corpus by means of a simple morphological query[30]. In this way, 49 instances of virile QSs in paucal numbers were retrieved. Table 2 summarizes the absolute frequencies of the relevant forms.

Table 2. Frequencies and earliest attestations of nominative and oblique virile QSs with paucal numbers in the manually annotated part of the *Elektroniczny korpus tekstów polskich z XVII. i XVIII. wieku*

Type	Token Frequency
dwaj	16
trzej	15
czterej	3
dwóch	6
trzech	4
czterech	5

[29] As of 6 July 2020.

[30] The exact syntax of the query was [case = "(nom)" & orth = "exact form of the number", e.g. "dwóch"].

The earliest attestation of *dwaj* dates to the first year of the corpus, 1601. This means that the form must have emerged in the course of the 16th century, which is plausible in light of the fact that the forms *trzej* and *czterej* are attested already in the 15th century (cf. §5.2.1).

The earliest attestations of virile paucal QSs in the oblique case date from the second half of the 17th century, with *dwóch* being the oldest attestation in the corpus. The source of this first instance of a virile QS with *dwóch* was written between 1656 and 1688, so the exact date of its occurrence is unclear. The example reads as follows:

(34) Trzej tylko dragani strzelili spadło
three$_{NOM.VIR}$ only dragoons$_{NOM}$ shoot$_{PST.VIR.PL}$ fall$_{PST.N.SG}$
tamtych z koni dwóch naszego też
those$_{GEN.PL}$ from horse two$_{GEN/ACC.M}$ our$_{ACC}$ also
jednego postrzelono w szyję.
one$_{ACC.M}$ shot$_{PST.PTCP}$ in neck
'Only three dragoons were shooting: two of theirs fell from their horses, one of ours was also shot in the neck.'

The frequencies given in Table 3 on the opposite page show that agreement resolution was not as clear-cut as it is today. There were hybrid cases with a number in the nominative combining with a noun in the genitive plural and yielding grammatical agreement, as in (35), or with an oblique form imposing semantic agreement, as in (36):[31]

(35) Byłem przy jednej nad którą czterej Doktorów siedziało
was$_{1SG.M}$ with one on who four$_{NOM}$ doctors$_{GEN.PL}$ sit$_{N.PST}$
'I was at a woman's who had three doctors sitting with her' (1680)

(36) Obrazek na którym Sodoma gorejąca, nad którą
picture$_{NOM}$ on which Sodom burning on which
aniołów dwóch ogień spuszczają, w ramkach
angels$_{GEN.PL}$ two$_{GEN.VIR.PL}$ fire$_{ACC}$ let.down$_{3PL}$ in frames
gładkich, złocistych
smooth golden
'A picture in a smooth golden frame showing the burning Sodom, on which two angels are dropping fire' (1696)

[31] Siuciak (2008: 191) also gives some examples of such hybrid constructions.

Table 3. Agreement resolution of virile quantified subjects with paucal numbers in the manually annotated part of the *Elektroniczny korpus tekstów polskich z XVII. i XVIII. wieku*

Type	Singular Agreement	Plural Agreement	No Finite Predicate Available
dwóch	4	2	0
trzech	1	1	2
czterech	2	0	3
dwaj	0	14	2
trzej	2	9	5
czterej	1	2	0

It could be that the first instances of virile QSs in the newly emerging oblique forms were less prototypical subjects in the sense outlined in Section 2. The first example of an oblique form functioning as subject, given in (34) above, is the subject of an unaccusative verb (*dwóch spadło* 'two fell') and is thus low in agentivity. Interestingly, this first oblique virile QS retrieved from the corpus follows a highly agentive paucal QS occurring in the nominative and with semantic agreement (*trzej*$_{NOM}$ *dragani*$_{NOM}$ *strzelili*$_{PL}$).

Indeed, most of the 15 instances of oblique virile QSs found in the data are intransitive and some even semantically unaccusative, but the number of examples is just too small to draw any further-reaching conclusion from this. There is also an instance of an agentive, referential QS in preverbal position with the oblique form *dwóch* (although with plural agreement, which may be interpreted as reflecting its subject-like properties). This example dates, however, from the middle of the 18th century and is thus decidedly younger than example (34):

(37) [MACEDONES waleczny Naród w Macedonii, których]

dwóch	Królów,	Filip	i	Aleksander	ad	summum
two$_{ACC}$	kings$_{GEN}$	Filip	and	Aleksander	to	highest
Sławy	culmen	podnieśli[32]				
glory	summit	raised$_{VIR.PL}$				

'[The Macedonians are a bellicose people in Macedonia, whom] two kings, Filip and Aleksander, raised to the crest of glory'

[32] The insertion of Latin words (*ad summum culmen*) is typical of this text (*Nowe Ateny* 4, 1756) and other historical Polish texts.

Given the small number of examples, I must leave the question of whether less typical QSs introduced the oblique virile forms earlier than more prototypical QSs to future studies, which should also take into account sources from the 16th century.

There is no continuous diachronic corpus of Polish available at present (Król et al. 2019). In particular, there is no corpus available covering the 58 years between the end of the Baroque corpus (1772) and the *Korpus tekstów polskich z lat 1830–1918* (Kieraś and Woliński 2018: 3854). As of June 2020, the *Korpus tekstów polskich z lat 1830–1918* includes varied samples of 1,000 texts. The automatically tagged part includes 1.3 million words; its manually annotated part includes 600,000 words (Witold Kieraś, p.c.).

The search for nominative and oblique virile QSs in this corpus did not yield any more hybrid examples—that is, instances of oblique QSs with plural agreement and nominative QSs with singular agreement, as attested occasionally in the Baroque corpus. This finding is indicative of a consolidation of the system.

Table 4 below summarizes the results of the search for paucal virile QSs in nominative and oblique forms in the *Korpus tekstów polskich z lat 1830–1918*. Due to the small size of the corpus, the informative value of these numbers is limited. However, as compared to the previously investigated stage in Table 3, in Table 4 oblique forms are in an almost even distribution with the nominative forms, already slightly outnumbering the latter.[33]

Table 4. Agreement resolution of quantified virile subjects with paucal numbers in the *Korpus tekstów polskich z lat 1830–1918*

Type	Token Frequency
dwóch	36
trzech	8
czterech	7
dwaj	35
trzej	7
czterej	5

[33] Note that the numbers provided in Siuciak 2008: 193, to which I got access only recently, are very much in line with the data given here.

5.3. *Dwaj* vs. *dwóch* in Contemporary Polish

Having traced the history of the two virile paucal forms, we will now turn to the variation attested between these forms in contemporary Polish. Grammars typically do not mention that the two sets of virile paucal forms take part in a paradigmatic relationship (e.g., Bielec 2015: 71f.). Or else, the oblique forms are simply described as a "formale Variante mit abweichender Syntax" 'formal variant with deviant syntax' (Bartnicka et al. 2004: 280) without giving any information about the use of these variants. Bielec (1999: 243) notes in a footnote that "[i]n colloquial speech[,] **dwaj** is replaced by **dwóch** or **dwu**" [boldface in the original].[34] This is an important clue, and it corroborates the intuitions of native speakers that the nominative set sounds somewhat outdated and belongs to an elaborate style.

A look into authentic data of contemporary Polish shows that the nominative forms are still well established. This raises the question of whether their use is really just a matter of style. Examples collected randomly by means of simple Google queries include the following:

(38) a. Grupa wyrostków skatowała dwóch mężczyzn
group$_{NOM}$ adolescents$_{GEN.PL}$ attacked$_{F.SG}$ two$_{GEN/ACC}$ men$_{GEN/ACC}$
'A group of adolescents attacked two men'
[headlines of a newspaper article]

b. Trzej chuligani w ciągu kilku minut, zaatakowali na ulicy dwóch przypadkowych mężczyzn
three$_{NOM}$ hooligans in course some minutes attacked$_{VIR.PL}$ on street two$_{ACC}$ random$_{ACC}$ men$_{ACC}$
'Three hooligans attacked two random men on the street within a few minutes.' ("Grupa wyrostków…", *24Opole*, 2012)

c. Jeśli wszyscy trzej byli poniżej 17 r.ż skończy się na rozprawie przed s[ą]dem rodzinnym.
if all$_{NOM}$ three$_{NOM}$ were$_{VIR.PL}$ under 17 years end.up$_{3PL}$ REFL on trial before court family$_{ADJ}$
'If all three of them were under 17 years old, they'll end up on trial before the family court.' [a user's commentary]

[34] Cf. also the frequencies of *dwaj* across various styles in contemporary Polish given in Linde-Usiekniewicz and Rutkowski 2003 (esp. p. 138) and in Linde-Usiekniewicz and Rutkowski 2007: 184.

(39) a. Dwóch policjantów z Tych jest
Two$_{GEN/ACC}$ police.officers$_{GEN/ACC}$ from Tychy is
podejrzanych o pobicie bezdomnego.
suspected$_{GEN/ACC}$ PREP beating homeless$_{ACC}$
'Two police officers from Tychy are suspected of beating a homeless man.'

b. Dzisiaj rano dwaj policjanci z
today morning two$_{NOM}$ policemen$_{NOM}$ from
Komendy Miejskiej w Tychach zostali$_{VIR.PL}$
police.station municipal in Tychy became
zatrzymani
detained$_{VIR.PL}$
'This morning the two policemen from the Tychy Municipal Police Station were detained […].'
("Na Śląsku zatrzymano dwóch policjantów", *naTemat*, 2020)

In both examples, the oblique forms occur in the first mention of the QS in the text, that is, in (38b) and (39a). The nominative forms are used only in subsequent mentions—in (38c) and (39b). This is reminiscent of QSs in Russian and BCS, where an increase in referentiality (for instance, by the introduction of a definite pronoun) makes semantic agreement available.

To establish whether contexts of increased referentiality—as signaled by the occurrence of a definite pronoun in the QS—favor the use of the nominative forms, a search in the corpus of modern Polish (National Corpus of Polish, NKJP; cf. Przepiórkowski et al. 2012) was conducted. The queries included three supposedly frequent virile nouns (*pan* 'mister'; *mężczyzna* 'man'; *chłopak* 'young man, boy'). Each of these nouns was once combined with the nominative and once with the oblique paucal form in the query. Importantly, the queries included the definite pronoun *ten* 'this' in the corresponding forms (i.e., *ci*$_{NOM.VIR.PL}$ and *tych*$_{ACC/GEN.VIR.PL}$). For example, for the paucal number two, one search was conducted for each of the nouns with *ci dwaj panowie/mężczyźni/chłopaki*, and one search for *tych dwóch panów/mężczyzn/chłopaków*. All the results were checked manually to make sure only syntactic subjects were included. Figure 1 (opposite) summarizes the absolute frequencies of nominative and oblique forms from all 18 queries.[35] With all three paucal numbers, the nominative forms are clearly favored in noun phrases with the demonstrative pronoun *ten*. To make sure that this result was not due to an over-representation of nominative forms in the data, it was compared with the overall frequencies of

[35] Three nominative and three oblique forms for each noun = (3 + 3) x 3 = 18 queries altogether.

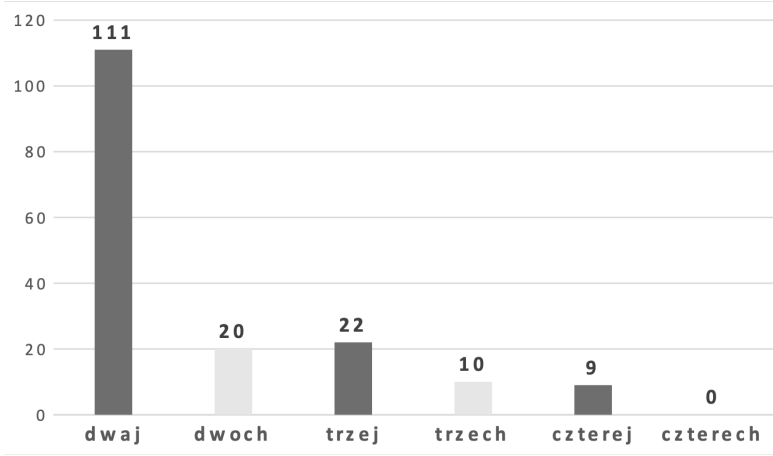


Figure 1. Token frequencies of nominative and oblique virile paucal numbers in definite noun phrases with three virile nouns in the NKJP (syntactic subjects only, 1,800M segments)

the nominative and oblique paucal forms in the manually annotated sub-corpus of the NKJP. Table 5 below shows that oblique paucal forms are generally more frequent in the manually annotated part of the corpus. This reflects the fact that the oblique forms are ousting the nominative forms in contemporary Polish. The overall dominance of the oblique forms underlines the meaningfulness of the finding that the nominative forms still dominate precisely when combined with a demonstrative pronoun, i.e., in contexts of increased referentiality. Against the background of this finding, the variation between *dwaj* and *dwóch* can be interpreted as a case of pragmatically motivated differential subject marking.

Table 5. Overall frequencies of nominative and oblique paucal virile forms in the manually annotated sub-corpus of the NKJP (approx. 511,000,000 segments)[36]

Type	Token Frequency	Type	Token Frequency
dwóch	63	*dwaj*	39
trzech	39	*trzej*	11
czeterech	9	*czeterej*	1

[36] I owe thanks to Witold Kieraś from the Institute of Computer Science of the Polish Academy of Sciences, Warsaw, for providing this information.

Another domain where the nominative forms seem more resistant are pure nominations, as, for instance, in titles of novels or movies—*Trzej muszkieterowie* (the three musketeers) and *Dwaj bracia* (two brothers)—or pieces of art—*Dwaj mężczyźni kontemplujący księżyc* (two men contemplating the moon), the title of a drawing by Caspar David Friedrich. This use of the nominative form is reminiscent of Jakobson's (1936/1971) classic account of the nominative case as representing the pure naming function ("Träger der reinen Nennfunktion", emphasis in original, ibid. 33).[37] Importantly, though, this can be no more than a tendency as the oblique forms are also available in these contexts (cf. the Polish translation of the movie *Three Men and a Baby*: *Trzech mężczyzn i dziecko*).

5.4. What Case is *dwóch/trzech/czterech*, and What Kinds of Subjects are QSs in Polish?

So far, the forms *dwóch/trzech/czterech* have been referred to simply as "oblique forms" as opposed to the nowadays less frequent nominative forms. It remains to determine the morphological case of these forms. The discussion of this question has a long tradition in Polish linguistics but will be summarized here with relative brevity.[38]

One possibility is simply analyzing the oblique forms as nominative when they occur in subject position. This is the stance taken in the annotations of Polish corpora. While this practice has been very convenient for the purposes of this study, it is unsatisfying in theoretical terms to treat the nominative and oblique forms identically. There are two other options available. One is to analyze the virile oblique forms as genitives and the non-virile non-oblique forms as nominatives. This has become known as the nominative-genitive hypothesis (*hipoteza mianownikowo-dopełniaczowa*; Przepiórkowski 2004). The other option is to analyze the oblique virile forms as accusatives. This has the advantage of a unified treatment of virile and non-virile paucal numbers in subject function, namely as accusatives, which is why this position is also referred to as the accusative hypothesis (*hipoteza biernikowa*; Przepiórkowski 2004).

[37] As mentioned in footnote 27, Siuciak (2008: 190) notes that the first attestations of virile forms in the nominative were in contexts of mere nomination of quantities ("zdania o charakterze konstatacji liczbowych"). Against this background, the functional retreat of the nominative forms is at the same time a retreat to the domain where they originated.

[38] See also overviews given by Przepiórkowski (2004) and Klockmann (2012: 37–42; 2017: 137–40).

The accusative analysis seems to have the greatest number of adherents, not only recently, but also across time (e.g., Łoś 1928; Suprun 1969: 84[39]; Franks 1995: 131–35; Rutkowski 2000; Przepiórkowski 2004; Miechowicz-Mathiasen 2012; Citko, Germain, and Witkoś 2018: 33–37; Witkoś et al. 2018; Witkoś 2020).[40] It is also the position adopted in this paper. Formalist and synchronically oriented papers have put great effort into the derivation of this accusative, for instance, by assuming a silent preposition assigning accusative case (see Lyskawa 2020). Such a solution is, however, unsatisfying in both theoretical and diachronic terms. First, the assumption of zero elements should be a last resort—a fact acknowledged in most theoretical frameworks, particularly within the vast body of functionally oriented frameworks with which this study is associated. Second, it is totally unclear how a silent, accusative-assigning preposition might have arisen diachronically.

Works including a diachronic perspective have assumed that the origin of the accusative is an accusative of measure (*biernik miary*) that was extended from quantifiers of nominal origin to numbers (e.g., Łoś 1928; Przepiórkowski 2004; Miechowicz-Mathiasen 2013). An example in point is *trochę* 'a bit', which is a fossilized accusative of the now vanished noun *trocha* 'small piece' (e.g., Herda 2019: 28). Other instances are the adverbials *odrobinę* 'a bit' and *masę* 'a lot', which can still be used as nouns (*odrobina* 'a small piece'; *masa* 'mass') up to the present day. This idea receives further plausibility in light of the fact that the accusative case is particularly frequent with expressions of quantity (we usually say how much of something we have or want; both contexts imply that the possessed or desired object or substance occurs as a direct object). Accordingly, Łoś (1928: 101) assumes the evolution of *trochę* 'a bit' illustrated in (40) on the following page:

[39] Suprun (1969: 84) only suspects that *dwóch* might be an accusative rather than a genitive, but he makes the interesting point that replacement of a nominative by an accusative is not restricted to number expressions (prominent examples are the Russian feminine nouns *svekrov'* 'mother-in-law' and *cerkov'* 'church', which were originally accusatives of *ū*-stem nouns). Cf. also examples (41–43) below.

[40] It should be noted that some authors assume the accusative analysis only with respect to non-paucal numbers in both genders, leaving out the question of what is the case of the virile non-paucal oblique forms (e.g., Miechowicz-Mathiasen 2013; Lyskawa 2020). Although we are concerned here mainly with the non-paucal oblique virile forms, the accusative analysis has the advantage of allowing for a unified analysis of both plural genders (virile and non-virile), and of all numbers, be they paucal or non-paucal.

(40) dał mi trochę cukru >
gave$_{M.SG}$ me$_{DAT}$ a.bit/a.small.piece$_{ACC}$ sugar$_{GEN}$
było trochę cukru
was$_{N.SG}$ a.bit$_{ADV}$ sugar$_{GEN}$
'he gave me a bit (= a small amount) of sugar' > 'there was a bit of sugar' (Łoś 1928: 101)

The question is whether the grammaticalization of nominal quantifiers corresponds to the chronology established for the first attestations of virile oblique non-paucal QSs in Section 5.2. More precisely, the grammaticalization of nominal quantifiers must already have been in process when the first attestations of the virile oblique non-paucal QS occurred.

Herda (2019) traces the origin of the three indefinite quantifiers of nominal origin *trochę*, *odrobinę*, and *masę*. The oldest of these quantifiers is *trochę*, whose earliest attestation as an adverbial extent modifier[41] dates from the 15th century (Herda 2019: 28–30). As established in Section 5.2.2, the oldest attestation of a virile oblique QS found in this study dates from the second half of the 17th century, which makes the scenario assumed here possible.

It is understandable that the penetration of the accusative into the domain of QSs did not have the same consequences for paucal QSs with non-virile nouns, or for QSs with non-paucal numbers. As for the paucal QS with non-virile nouns, the accusative is syncretic with the nominative. Therefore, it has been possible for the non-virile paucal QS to resist reinterpretation as an accusative. Interestingly, Łoś (1928: 10) notes that there was indeed a tendency of non-virile paucal QSs towards default agreement (e.g., *było*$_{N.SG}$ *trzy gwiazdy* 'there were three stars'), which is indicative of the fact that an "accusative interpretation" of the number was also taking place with non-virile QSs. This development was obviously stopped, possibly also due to normative intervention, since default agreement is not admissible (anymore) with non-virile paucal QSs in contemporary Polish. The same analysis also applies to non-paucal non-virile QSs, as the nominative form does not differ from the accusative either. The virile forms of non-paucal numbers, e.g., *pięciu/sześciu/siedmiu* 'five$_{VIR}$/six$_{VIR}$/seven$_{VIR}$', in QSs can also be accounted for as accusatives (cf. the overview given in Siuciak 2008: 192).[42]

[41] Extent modifiers (e.g., *a bit* in *to wait a bit*) develop earlier than degree modifiers (e.g., *a bit* in *to worry a bit*) in the grammaticalization of denominal adverbial modifiers cross-linguistically (cf. Herda 2019 and the references therein).

[42] For a sketch of the history of the ending *-u* in virile numbers, see Siuciak 2008: 80–82.

The assumption of quantitative subjects in the accusative receives further plausibility in light of the following examples from contemporary Polish (41) and BCS (42–43), which have other QSs in the accusative as well:

(41) Kupę czasu minęło od tamtego wydarzenia.
heap$_{ACC}$ time$_{GEN}$ passed$_{N.SG}$ from that incident
'A lot of time has passed since that incident.' (Herda 2019: 24)

(42) Polovinu roditelja zabrinuto za budućnost,
half$_{ACC}$ parents$_{GEN}$ worried$_{N.SG}$ for future
a DODATNO SU OPTEREĆENI ON LINE ŠKOLOVANJEM SVOJE DJECE [emphasis in original]
'Half of the parents are worried about the future, and additionally, they are stressed from homeschooling their kids'
("Polovinu roditelja…", *IstraIN*, 2020)

(43) Prošlo je godinu dana.
passed$_{N.SG}$ AUX$_{3SG}$ year$_{ACC}$ day$_{GEN}$
'One year passed.'

One more indication that QSs are in the accusative is that they can function as the direct objects of so-called Adversity Impersonals (Witkoś 2020: 260), which require a direct object in the accusative:

(44) Dwóch marynarzy zabiło po ich wachcie.
two$_{ACC}$ sailors$_{ACC}$ killed$_{N.SG}$ after their watch
'Two sailors were killed after their watch.'
(example adapted from Citko, Germain, and Witkoś 2018: 24)

Klockmann (2012: 58–106, 141f.; 2017: 138–40) is a recent adherent of the nominative-genitive hypothesis, which assumes nominative case for numbers 2–4 with non-virile nouns, but genitive for the oblique numbers with virile nouns. Klockmann puts great effort into deriving the genitive of paucal numbers with virile nouns, including the assumption of a "cyclic Agree" mechanism and an appeal to the partitive genitive assigned to nouns of either gender when combined with non-paucal numbers. However, the accusative hypothesis as outlined above is a simpler, more inclusive and more plausible analysis in light of historical and comparative data.

It remains to answer the question of what kind of subjects Polish QSs actually are. Although Polish QSs receive non-canonical (non-nominative) case marking, they show all behavioral subject properties, such as, for instance, binding of reflexive pronouns (45) and control into gerunds (46):

(45) Ośmiu rabotników$_i$ oddało swoje$_i$ klucze.
eight$_{VIR}$ workers$_{GEN}$ handed.over$_{N.SG}$ their$_{ACC}$ keys$_{ACC}$
'The eight workers handed over their keys.'
(Citko, Germain, and Witkoś 2018: 33)

(46) Wracając$_i$ do domu, dwóch mężczyzn$_i$ usłyszało krzyk.
returning to home two$_{ACC}$ men$_{GEN}$ heard$_{N.SG}$ scream$_{ACC}$
'Returning home, two men heard a scream.'

Polish QSs thus display behavioral properties like those of canonical subjects. Unlike the latter, however, QSs do not impose agreement on predicates and do not take the nominative case (see also Dziwirek 1994: 214–17; Witkoś et al. 2018: 101–14; Witkoś 2020: 259).

The only exceptions to this rule—that is, the only QSs inducing semantic agreement in contemporary Polish—include numbers whose forms allow for an interpretation as attributive modifiers agreeing with the head noun in the nominative. These are QSs with non-virile nouns and paucal numbers, and of course, the clearly nominative virile forms *dwaj/trzej/czterej*.

All other QSs in contemporary Polish can be analyzed as subject-like obliques in the sense of Seržant 2013 and Schlund 2018. Like all subject-like obliques, Polish QSs cannot induce agreement in predicates. Their capacity to display all behavioral properties of subjects, however, brings Polish QSs closer to the category of non-canonical subjects than other subject-like obliques (such as, for instance, the dative experiencers in various impersonal constructions of Slavic, whose behavioral subject properties are typically much more restricted; e.g., Schlund 2018: 140–46 for Russian).

6. Conclusion

Slavic QSs form a heterogeneous but cognitively well-motivated category. This is because their heterogeneity is inspired by differences in human perception and construal of low numbers on the one hand, and of high or unspecific numbers on the other. These differences help explain both the heterogeneous origins of paucal and non-paucal numbers and their sometimes perplexing agreement resolution patterns and morphological forms up to the present day.

Quantified subjects typically deviate from the subject prototype, particularly with respect to the subject property of referentiality. In Russian deviation from the subject prototype is reflected in the agreement resolution of QSs, with semantic agreement indicating greater prototypicality of the QS than grammatical agreement. In BCS normative rules determine agreement resolution to a great extent, but a tendency towards semantic agreement even with non-paucal numbers is observable precisely when semantic and pragmatic

subject properties increase. As in Russian, this holds particularly for contexts of greater referentiality.

Diachronically, agreement resolution of Polish QSs was originally motivated in ways very similar to the cases in Russian and BCS. The fact that virile paucal QSs in modern Polish require default neuter singular agreement and are marked for the oblique (non-nominative) case likewise received a plausible explanation against historical data. In line with a number of previous accounts, it has been argued here that this case is an accusative of measure (*accusativus mensurae*) that was extended from nominal expressions of quantity to QSs with numerals. The accusative differs from the nominative only with virile nouns (that is, *dwóch/trzech/czterech*), which is why paucal virile QSs function like oblique subjects (or subject-like obliques, in Seržant's 2013 terminology). This analysis of virile paucal QSs applies also under the genitive hypothesis, but there is greater evidence for the accusative hypothesis. Oblique subjects never induce semantic agreement in the predicate, which is why default agreement occurs. The present paper provides some evidence that the usage of nominative and oblique forms of virile paucal numbers in QSs in contemporary Polish is influenced by the subject property of referentiality, but further studies will be necessary to test this claim. As noted by Siuciak (2008: 190), the first attestations of the virile paucal numbers *trzej* and *czterej* (and later, *dwaj*) occur in contexts of mere nomination, and there is some evidence that they are being reduced to this initial function in contemporary Polish (§5.3).

It remains to be investigated in more detail why contexts of increased referentiality can make semantic agreement obligatory (as in Russian) or more likely (as in BCS), or influence the choice between two alternative sets of numbers (namely, the nominative and oblique virile forms of paucal numbers in Polish). After all, the influence of increased agentivity on semantic agreement resolution appears weaker across all three languages. To answer this question, a more refined concept of referentiality will be necessary—one which makes a clear distinction between referentiality as a semantic category and the formal means of a language to signal different degrees or aspects of referentiality. Demonstrative pronouns are but one indicator of the referentiality of a noun phrase; there are other categories to keep in mind (such as, for instance, possessive pronouns or anaphoric reference). What is more, the range and usages of demonstrative pronouns vary considerably across Slavic. While demonstratives serve mainly deictic functions in Russian, anaphoric use of demonstrative pronouns is more elaborate in Polish (see Bunčić 2014: 81). Any effort to determine the role of deixis or anaphor in the agreement resolution of Slavic QSs will have to take these differences into account.[43]

[43] I owe thanks to Hagen Pitsch (Göttingen) for pointing out to me that issues of normativity might also play a role here. The fact that the agreement resolution of Russian

Leaving these questions to further studies, the diachronic and synchronic evidence interpreted in this paper reveals that agreement resolution of QSs in Slavic is a prime example of converging formal, semantic, and cognitive forces in grammar.

Sources

"Blogi". (2012) *Russkij reporter*. Retrieved through the Russian National Corpus.

Elektroniczny korpus tekstów polskich z XVII. i XVIII. wieku (do 1772 roku). (2013–) Polish Academy of Sciences. Available at: https://korba.edu.pl/query_corpus/. Last accessed 18 December 2021.

"Grupa bh. državljana vraćena iz Sirije". (20 December 2019) *Etleboro.org*. Available at: https://etleboro.org/g/1cf8a2fb569b0fa6b4cedf0de2b80283bs/iz-sirije-u-bih-deportovano-sedam-mu%C5%A1karaca-%C4%8Dlanova-idil-a-%C5%A1est-%C5%BEena-i-12-djece. Last accessed 8 December 2021.

"Grupa wyrostków skatowała dwóch mężczyzn". (22 April 2012) Seeb. *24Opole* [Opole informational service]. Available at: https://www.24opole.pl/8998,-Grupa_wyrostkow_skatowala_dwoch_mezczyzn,wiadomosc.html. Last accessed 29 April 2020.

Kollekcija anekdotov: Blondinki. (1970–2000) G. B. Xasanov. Retrieved through the Russian National Corpus.

Korpus tekstów polskich z lat 1830–1918. Polish Academy of Sciences. Available at: http://korpus19.nlp.ipipan.waw.pl/query_corpus/. Last accessed 16 December 2021.

[*KTS*] *Korpus tekstów staropolskich*. Polish Academy of Sciences. Available at: https://ijp.pan.pl/publikacje-i-materialy/zasoby/korpus-tekstow-staropolskich/. Last accessed 18 December 2021.

"Na Śląsku zatrzymano dwóch policjantów. Mieli dotkliwie pobić bezdomnego". (16 December 2020) *naTemat*. Available at: https://natemat.pl/330875,tychy-dwaj-policjanci-zatrzymani-w-zwiazku-z-pobiciem-bezdomnego. Last accessed 22 January 2022.

"Najstarije na svijetu: Ovih pet žena rođeno je u 19. stoljeću". (8 January 2015) *Klix*. Available at: https://www.klix.ba/magazin/zanimljivosti/najstarije-na-svijetu-ovih-pet-zena-rodjeno-je-u-19-stoljecu/150108013. Last accessed 22 January 2022.

[NKJP] *National corpus of Polish*. Available at: http://nkjp.pl/. Last accessed 18 December 2021.

QSs is mostly determined semantically finds a correlate in the alternation between nominative and instrumental case in Russian copular sentences, which may likewise reflect semantic nuances. In Polish, agreement with QSs and case selection in copular sentences are determined structurally (cf. Pitsch 2018).

"Polovinu roditelja zabrinuto za budućnost, a DODATNO SU OPTEREĆENI ON LINE ŠKOLOVANJEM SVOJE DJECE". (18 April 2020) *IstraIN* [Istrian news portal]. Available at: https://istrain.hr/index.php/ostalo-arhiva/114-zanimljivosti/8728-polovinu-roditelja-zabrinuto-za-buducnost-a-dodatno-su-optereceni-on-line-skolovanjem-svoje-djece. Last accessed 11 August 2020.

Russian national corpus. (2003–) Available at: http://ruscorpora.ru/new/. Last accessed 18 December 2021.

Sbornik olimpiadnyx zadač po matematike. (2016) Nikolaj Gorbačëv. 4th edition. Moscow: MCNMO. Retrieved 3 April 2020 from https://books.google.de/

Sneguročka + 1: Skazka dlja vzroslyx. (2019) Elena Martynova. Litres. Retrieved 17 April 2020 from Google Books.

"V Tuve šest' otmorozkov ubili suprugov v ix že kvartire". (28 June 2019) Ksenia Sušina. *A42.RU*. Available at: https://gazeta.a42.ru/lenta/news/56326-v-tuve-shest-otmorozkov-ubili-suprugov-v-ih-zhe-kvartire. Last accessed 8 December 2021.

References

Bartnicka, Barbara, Björn Hansen, Wojtek Klemm, Volkmar Lehmann, and Halina Satkiewicz. (2004) *Grammatik des Polnischen*. Munich: Sagner. [Slavolinguistica, 5.]

Bielec, Dana. (1999) *Polish: An essential grammar*. London and New York: Routledge. [Routledge Grammar Workbooks.]

————. (2015) *Basic Polish: A grammar and workbook*. 2nd edition. London and New York: Routledge.

Bunčić, Daniel. (2014) "Definitheit als ‚verborgene Kategorie' im Russischen?" Hagen Pitsch, ed. *Linguistische Beiträge zur Slavistik*. Vol. 21. *JungslavistInnen-Treffen in Göttingen, 13–15. September 2012*. Munich: Otto Sagner, 75–96. [Specimina philologiae Slavicae, 180.]

Citko, Barbara, Allison Germain, and Jacek Witkoś. (2018) "If you cannot agree, move on! On labels and non-nominative subjects". *Glossa: A journal of general linguistics* 3(1): 28. 1–46. DOI 10.5334/gjgl.399

Corbett, Greville G. (1983) *Hierarchies, targets and controllers: Agreement patterns in Slavic*. London and Canberra: Croom Helm. [Croom Helm Linguistic Series.]

————. (2004) *Number*. Cambridge: Cambridge University Press. [Cambridge Textbooks in Linguistics.]

————. (2010) "Agreement in Slavonic". *Glossos* 10: 1–61.

Długosz-Kurczabowa, Krystyna. (2003) *Nowy słownik etymologiczny języka polskiego*. Warsaw: Wydawnictwo Naukowe.

Dziwirek, Katarzyna. (1994) *Polish subjects*. New York and London: Garland Publishing, Inc. [Outstanding Dissertations in Linguistics.]

Franks, Steven. (1995) *Parameters of Slavic morphosyntax*. New York and Oxford: Oxford University Press. [Oxford Studies in Comparative Syntax.]

————. (2009) "Case assignment in quantified phrases". Tilman Berger, Sebastian Kempgen, Peter Kosta, and Karl Gutschmidt, eds. *Die slavischen Sprachen/The Slavic languages*. Berlin: de Gruyter, 355–68. [Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science, 32.1.]

Garncarek, Piotr. (2018) "Jak liczą Polacy—Lingwakulturowe aspekty gramatyki opisowej i historycznej liczebnika". *Acta Universitatis Lodziensis: Kształcenie Polonistyczne Cudzoziemców* 25: 77–85.

Gvozdanović, Jadranka. (1999) "Some remarks on number morphosyntax in Slavic". Jadranka Gvozdanović, ed. *Numeral types and changes worldwide*. Berlin and New York: Mouton de Gruyter, 187–96.

Heine, Bernd. (1997) *Cognitive foundations of grammar*. New York and Oxford: Oxford University Press.

Herda, Damian. (2019) "On the adverbialization of Polish indefinite quantifiers of nominal origin: A diachronic study of *trochę* 'a bit', *odrobinę* 'a bit', and *masę* 'a lot' ". *Studies in Polish linguistics* 14(1): 19–42.

Jakobson, Roman. (1936/1971) *Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutung der russischen Kasus. Roman Jakobson: Selected writings.* Vol. 2. *Word and language*. The Hague: Mouton, 23–71.

Janda, Laura A. (1999) "Whence virility? The rise of a new gender distinction in the history of Slavic". Margaret H. Mills, ed. *Slavic gender linguistics*. Amsterdam and Philadelphia: John Benjamins, 201–28.

Kieraś, Witold and Marcin Woliński. (2018) "Manually annotated corpus of Polish texts published between 1830 and 1918". Nicoletta Calzolari et al., eds. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyasaki, Japan: European Language Resources Association, 3854–59.

Klemensiewicz, Zenon. (2002) *Historia języka polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Klockmann, Heidi. (2012) *Polish numbers and quantifiers: A syntactic analysis of subject-verb agreement mismatches*. M.A. thesis, Utrecht University.

————. (2017) *The design of semi-lexicality: Evidence from case and agreement in the nominal domain*. Ph.D. dissertation, Utrecht University.

Król, Magdalena et al. (2019) "Narodowy Korpus Diachroniczny Polszczyzny: Projekt". *Język Polski* 99(1): 92–101.

Levin, Vanessa. "What is subitizing? Activities for preschool and pre-K". *Pre-K pages*. Available at: https://www.pre-kpages.com/subitizing/. Last accessed 2 December 2019.

Linde-Usiekniewicz, Jadwiga and Paweł Rutkowski. (2003) "O zależności między właściwościami składniowymi i frekwencyjnymi liczebników głównych w polszczyźnie". Romuald Huszcza and Jadwiga Linde-Usiek-

niewicz, eds. *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*. Warsaw: Wydział Polonistyki Uniwersytetu Warszawskiego, 125–40.

————. (2007) "Frequency of use of Polish numbers does not influence their syntax". *Proceedings of the 22nd Annual Northwest Linguistics Conference*. Burnaby, B.C., Canada: FU Linguistics Graduate Student Association (LGSA), 175–86. [Simon Fraser University Working Papers in Linguistics, 1.] Available at: https://www.sfu.ca/gradlings/working-papers/wp_1.html.

Łoś, Jan. (1928) "'Trzy piękne córki było nas u matki', czyli formy podmiotu i orzeczenia w zdaniach z podmiotem logicznym, określonym przydawką liczebnikową". *Język Polski* 13(4): 97–112.

Lyskawa, Paulina. (2020) "The structure of Polish numerically-quantified expressions". *Glossa: A journal of general linguistics* 5(1): 31. 1–37. DOI 10.5334/gjgl.880.

Malchukov, Andrej and Akio Ogawa. (2011) "Towards a typology of impersonal constructions. A semantic map approach". Andrej Malchukov and Anna Siewierska, eds. *Impersonal constructions: A cross-linguistic perspective*. Amsterdam and Philadelphia: John Benjamins, 19–56. [Studies in Language Companion Series, 124.]

Marušič, Lanko and Andrew Nevins. (2010) "Two types of neuter: Closest-conjunct agreement in the presence of '5 and up'". Wayles Browne et al., eds. *Annual Workshop on Formal Approaches to Slavic Linguistics: The Second Cornell Meeting 2009*. Ann Arbor: Michigan Slavic Publications, 301–17. Available at: https://ling.auf.net/lingbuzz/000919

Miechowicz-Mathiasen, Katarzyna. (2012) "Licensing Polish higher numbers: An account of the accusative hypothesis". Joanna Błaszczak, Bożena Rozwadowska, and Wojciech Witkowski, eds. *Current issues in generative linguistics: Syntax, semantics and phonology*. Wrocław: Wrocław University Press, 58–75. [Generative Linguistics in Wrocław, 2.]

————. (2013) "On agreement with number phrase subjects including ≥ 5 in Old and Middle Polish". *Scando-Slavica* 59(1): 80–107.

Miechowicz-Mathiasen, Katarzyna and Dominika Dziubała-Szrejbrowska. (2012) "The role of gender in the rise of numerals as a separate category". *Journal of historical syntax* 1: 1–39.

Pitsch, Hagen. (2018) *Kopula, Auxiliar, Prädikativ: Zur Satzstruktur in nordslavischen Sprachen*. Berlin and Bern: Peter Lang. [Specimina philologiae Slavicae, 195.]

Przepiórkowski, Adam. (2004) "O wartości przypadka podmiotów liczebnikowych". *Bulletin de la Société Polonaise de Linguistique* 60: 133–43.

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk, eds. (2012) *Narodowy korpus języka polskiego*. Warsaw: Wydawnictwo naukowe PWN.

Robblee, Karen E. (1993) "Individuation and Russian agreement". *Slavic and East European journal* 37(4): 423–41.

Rozental', D. Ė., E. V. Džandžakova, and N. P. Kabanova. (2005) *Spravočnik po russkomu jazyku: Pravopisanie, proiznošenie, literaturnoe redaktirovanie*. Moscow: AYRIS PRESS.

Rutkowski, Paweł. (2000) "Składnia polskich grup liczebnikowych: Próba opisu formalnego". *Poradnik Językowy* 8: 10–28.

Schlund, Katrin. (2018) "A unifying approach to impersonality in Russian". *Zeitschrift für Slawistik* 63(1): 120–68.

————. (2019) "Grammatische versus semantische Kongruenz von numberphrasen im Bosnischen, Kroatischen und Serbischen". Alexander Bierich, Thomas Bruns, and Henrieke Stahl, eds. *Gedächtnisraum Literatur—Gedächtnisraum Sprache: Europäische Dimensionen slavischer Geschichte und Kultur*. Festschrift for Svetlana and Gerhard Ressel. Berlin: Peter Lang, 173–89. [Trierer Studien zur Slavistik, 5.]

Seržant, Ilja. (2013) "The diachronic typology of non-canonical subjects and subject-like obliques". Ilja Seržant and Leonid Kulikov, eds. *The diachronic typology of non-canonical subjects*. Amsterdam and Philadelphia: John Benjamins, 313–60. [Studies in Language Companion Series, 140.]

Siuciak, Mirosława. (2008) *Kształtowanie się kategorii gramatycznej liczebnika w języku polskim*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Suprun, Adam E. (1969) *Slavjanskie čislitel'nye: Stanovlenie čislitel'nyx kak osoboj časti reči*. Minsk: Izdat. BGU im. V. I. Lenina.

Willim, Ewa. (2015) "Case distribution and ϕ-agreement with Polish Genitive of Quantification in the feature sharing theory of Agree". *Poznań studies in contemporary linguistics* 51(2): 315–57.

Witkoś, Jacek. (2020) "On accusative number subjects in Polish". Teodora Radeva-Bork and Peter Kosta, eds. *Current developments in Slavic linguistics: Twenty years after (based on selected papers from FDSL 11)*. Berlin: Peter Lang, 259–72. [Potsdam Linguistic Investigations, 29.]

Witkoś, Jacek, Dominika Dziubała-Szrejbrowska, Piotr Cegłowski, and Paulina Łęska. (2018) *The syntax of number noun constructions: A view from Polish*. Berlin: Peter Lang. [Polish Studies in English Language and Literature, 40.]

Katrin Schlund
Department of Slavic Studies
Martin Luther University Halle-Wittenberg
Halle, Germany
katrin.schlund@slavistik.uni-halle.de