volume 32, number 2

summer–fall 2024


# 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.

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 typically alternate between North America and Europe. For more information about SLS or to join, visit www.slaviclinguistics.org.

## 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 website, where submission guidelines are available, as well as online versions of volumes starting with volume 19 (2011).

*JSL*'s World Wide Web page: https://ojs.ung.si/index.php/JSL

The *Journal of Slavic Linguistics* publishes two regular issues per year, totaling approximately 360 pages: Winter–Spring and Summer–Fall. Each year *JSL* also publishes an extra open-access, online-only issue that contains the proceedings from the Formal Approaches to Slavic Linguistics (FASL) conference. For regular issues, 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 November 2025.

ISSN: 1068-2090

Technical Editors: Jordan Hussey-Andersen and Ellen Tamura

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@iu.edu**
**https://slavica.indiana.edu/**

journal of slavic linguistics

volume 32, number 2

summer–fall 2024



# articles



# reviews

# From the Editors

This is the second issue of volume 32, the 2024 volume. It has been JSL's practice to acknowledge and thank colleagues who serve as reviewers for JSL submissions approximately every two years, so here is a list of people who we wish to thank for contributing reviews between 1 January 2023 and 31 December 2024 (listed in alphabetical order):

Valentina Apresyan, Boban Arsenijević, Sacha Beniamine, Vladimír Benko, Loren Billings, Tatiana Bondarenko, Anna Bondaruk, Željko Bošković, Wayles Browne, Martin Březina, Irina Burukina, Pavel Caha, Anna Cardinaletti, Božena Cetnarowska, Nicholas Danis, Andriy Danylenko, Miloje Despić, Stephen M. Dickey, Tom Dickins, Mila Dimitrova Vulchanova, Mojmir Dočekal, Jakub Dotlačil, Imke Driemel, Antonio Fabregas, Jan Fellerer, Daniel Finer, Alberto Frasson, Jacopo Garzonio, Ljudmila Geist, Ekaterina Georgieva, Anna Grabovac, Marc Greenberg, Jirka Hana, Nicole Hockmann, Gašper Ilc, Tania Ionin, Pavel Iosad, Lukasz Jadrzejowski, Hakyung Jung, Uwe Junghanns, Edyta Jurkiewicz-Rohrbacher, Vsevolod Kapatsinski, Mikhail Knyazev, Alexei Kochetov, Iliana Krapova, Ivona Kučerová, Irenäus Kulik, Vladimir Kulikov, Masha Kyuseva, Keith Langston, Paulina Łęska-Bayraktar, Beccy Lewis, Jadwiga Linde-Usiekniewicz, Anna Lubowicz, Nerea Madariaga, Franc Marušič, Ora Matushansky, Natalia Meir, Igor Mel'čuk, Krzysztof Migdalski, Tanja Milićev, Petar Milin, Stefan Milosavljević, Petra Mišmaš, Marcin Morzycki, Olav Mueller-Reichau, Alexandre Nikolaev, Jaye Padgett, Jeffrey Parker, Matic Pavlič, Hagen Pitsch, Anna Poĺomská, Ljiljana Progovac, Zorica Puškar-Gallien, Milan Rezac, Tom Roberts, Kevin Roon, Jerzy Rubach, Marta Ruda, Catherine Rudin, Tobias Scheer, Marko Simonović, Roumyana Slabakova, Tomislav Sočanac, Sandra Stjepanović, Anastasia Stoops, Dejan Stosić, Adam Szczegielniak, Radek Šimík, Danko Šipka, Guy Tabachnick, Aida Talić, Lucie Taraldsen Medová, Neda Todorović, Ludmila Veselovska, Ekaterina Vostrikova, Bartosz Wiland, Magdalena Zawislawska, Rok Žaucer, Jelena Živojinović.

While we make every effort to ensure that this list is complete, we acknowledge the possibility of error—if you believe that the list is missing anyone, please write to us.

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

# Similar Place Avoidance in Slavic and Other Languages

**Aleš Bičan**

*Abstract*: The paper discusses a constraint on the distribution of homorganic CVC sequences known as Similar Place Avoidance (SPA). Though proposed as a statistical universal, it has been little considered in Slavic and other Indo-European languages. We evaluate the CVC distribution in 100 recorded and reconstructed varieties, of which 18 are Slavic, 44 are non-Slavic Indo-European, and 38 are non-Indo-European. The SPA principle has been formulated as pertaining to CVC sequences of two consonants sharing the same place, but it has also been suggested that coronals are dependent on sonorancy agreement for the constraint to take effect. This dependency is indeed observable but concerns dento-alveolars only, not coronals as a whole class. SPA weakly restricts combinations of dento-alveolar sonorants with palatal sonorants. Combinations of different-place coronal obstruents are disfavored, but this is instead due to sibilancy avoidance (a restriction of the co-occurrence of two sibilants in a CVC sequence, previously unreported). Finally, combinations of palatals (including post-alveolars) are less often subject to an SPA effect, and the Slavic languages virtually lack this kind of restriction.

*Keywords:* Obligatory Contour Principle, phonotactics, palatal, sonorant, coronal, sibilant, labial-coronal effect

## 1. Introduction*

One of the best-known phonotactic constraints operative in spoken languages is the avoidance of repetitions of similar consonants separated by a nuclear segment or appearing in other kinds of adjacency. This avoidance is at the core of several principles known as Obligatory Contour Principle (McCarthy 1986), Similarity Avoidance (Frisch et al. 2004), Identity Avoidance (Tang 2000), Repetition Avoidance (Walter 2007), and Similar Place Avoidance (Pozdniakov and Segerer 2007). The constraint primarily pertains to the place of articulation. However, other consonantal properties, such as the manner of articulation

* The paper has profited enormously from two anonymous inspirational and insightful reviews, especially from the second 22-page review. Needless to say, all remaining errors are our own.

or the state of the glottis, may contribute to the same-place consonant co-occurrence restriction, or the non-place features may have their own avoidance restrictions (MacEastern 1999; Pozdniakov 2010; see also Gallagher 2015 and references therein). There is also some evidence that vowel co-occurrence is restricted in a similar manner (Walter 2010; Doucette et al. 2024).

Since Cantineau's (1946) and Greenberg's (1950) pioneering studies of Classical Arabic, where the dislike for similar consonant repetition was first noted, the constraint has been observed in nearly every language studied. To date, the largest and most comprehensive treatment of the phenomenon is Pozdniakov and Segerer's seminal work (2007; henceforth P&S). Analyzing CVC sequences in large lexical databases for 31 languages from several families, they show that although these languages allow repetitions of consonants belonging to four basic place classes of labials, dentals/alveolars, palatals, and velars, the frequency of these homorganic combinations is much smaller than one would expect from the occurrence of the particular consonants. Their data also revealed that a comparable, though less strong, kind of avoidance may apply to consonants belonging to two place superclasses sharing some similarity, one grouping together labials and velars ("peripherals" from the perspective of the oral cavity), the other uniting dentals/alveolars and palatals ("medials"). These statistical cross-linguistic tendencies have become known as Similar Place Avoidance (SPA).[1]

Previously and subsequently, other studies have confirmed the cross-linguistic avoidance of homorganic CVC combinations of labials and velars, but the situation with "medial" or coronal consonants is more complicated. Researchers disagree on whether the avoidance pertains to coronals as a whole class (Cooper 2009; Mayer et al. 2010) or to dento-alveolars and palatals separately (P&S; Nikitina 2022), or whether palatals should be included in the class of coronals at all (Rousset 2004; Kinney 2005). The question of whether palatals pattern with dento-alveolars or with velars (the latter claimed by Padgett 1992 for Russian) is also unclear. In any manner of classification, however, the avoidance effect for "medial" place classes has turned out to be weaker than for the peripheral classes, or the effect manifests itself only in conjunction with non-place features, such as sonorancy, stricture/continuancy, or voicing (Pierrehumbert 1993; McCarthy 1994; Coetzee and Pater 2008).

Despite the uncertainty concerning the most appropriate formulation of the SPA principle, the fact that languages or their users tend to disfavor repetitions of similar consonants is undeniable. Besides the 31 languages in P&S's study, most of which are from Africa, the effect has been documented or noted

[1] As already pointed out, some writers subsume this phenomenon under other names and/or principles such as Obligatory Contour Principle (OCP, or OCP-Place when it concerns the place of articulation). We will speak about SPA irrespective of how the cited works label the phenomenon.

to exist in at least 60 other languages from Afro-Asiatic (including Semitic), Austronesian, Indo-European (IE), and other families, including creoles (see the references throughout this paper). The effect has further been confirmed for as many as 3,200 languages in Mayer et al. (2010), but this otherwise formidable study must be treated with some caution because the authors examined the limited vocabulary of the Swadesh list. SPA has also been considered in studies on the evolution of speech from babbling through first words to adult languages, especially in relation to the preference for Labial-Vowel-Coronal sequences (the Labial-Coronal effect; MacNeilage and Davis 1999, 2000; MacNeilage et al. 2000; Rousset 2004; Kinney 2005; Vallée et al. 2009; Carrissimo-Bertola 2010).

Support for the SPA principle is also found in areas other than statistical analyses of morpheme or word lists. Experiments have demonstrated that nonwords violating SPA are judged less well-formed by speakers of Hebrew (Berent and Shimron 1997), Arabic (Frisch and Zawaydeh 2001), and English (Davis 1991; Frisch 1997; Coetzee 2005). Other experiments indicate that SPA-violating sequences help Dutch speakers identify a word boundary or spot a word's presence in artificial utterances (Boll-Avetisyan and Kager 2014). There is, furthermore, evidence that loanwords resist adaptation to the recipient language when violating SPA (Frisch et al. 2004 on Maltese). Further support comes from Hulden (2017), who reports that an algorithm created for automatically detecting and separating vowels and consonants in a running text has proven to be more successful when SPA distributional restrictions have been implemented into the algorithm. Finally, some studies mention that SPA also limits consonant combinability in English consonant clusters (Pierrehumbert 1994). Explanations of why SPA is such a robust distributional restriction are usually grounded in language perception and processing, and confusability avoidance (Pierrehumbert 1993; MacNeilage et al. 2000; Frisch et al. 2004; Frisch 2004; Graff 2012; Boll-Avetisyan and Kager 2014; cf. also Wright 2004 and Zukoff 2017).

The great attention SPA has received suggests that the validity of this effect is firmly established, and any further confirmation is merely a matter of empirical support. However, as already mentioned, several unresolved theoretical issues remain concerning how the place classes subject to SPA restrictions are defined, and to what extent non-place features as well as segment distance, prosody, or morphology influence the effect. Since space prevents us from addressing all these issues, we will only pay attention to the role of sonorancy in SPA restrictions and to the way the class of coronals is delimited and subdivided. Also, we will look into the nature of palatals, which happen to be resistant to SPA in many languages. These problems will be subsumed under the main goal of our paper, which is to provide a comprehensive survey of the SPA effect using large lexical datasets for Slavic, IE, and other languages.

Many, though not all, previous SPA studies are potentially deficient in drawing generalizations about entire language families or subbranches without having investigated a sufficiently large sample. The sampling may lead to incorrect conclusions if, for instance, Russian and English are selected as representatives of Slavic and Germanic languages, respectively. Therefore, we will conduct an analysis of one language class (Slavic) supplemented by a similarly extensive treatment of its superordinate family (IE languages). The 38 non-IE languages widen the perspective but are not in the main focus. We have chosen Slavic to compensate for the neglect this family has received in SPA studies. To the best of our knowledge, only Russian has been analyzed from the perspective of SPA (Padgett 1992, 1995; Pozdniakov 2007, 2010, 2011). Croatian has been considered from a related perspective of segment repetition in general (MacKay 1970).

The rest of the paper is organized as follows. In the next section, we introduce our Slavic and non-Slavic material (§2.1, §2.2), explaining how it is phonologically classified (§2.3) and statistically evaluated (§2.4). Section 3 discusses the strength and diffusion of SPA in Slavic; the next section does the same for non-Slavic IE (§4.1) and non-IE languages (§4.2). Since many languages fail to avoid palatal homorganic CVC sequences, we will present a typology of palatal systems to find correlations between the nature of SPA and the nature of these systems. Section 5 is dedicated to the problem of coronals. One of the major findings of this paper is that two separate classes of coronals should be distinguished for the SPA effect to be adequately accounted for, namely dento-alveolars on the one hand and post-alveolars together with (alveolo-) palatals on the other. The previously described distributional restrictions on coronals are shown to be properties of dento-alveolars rather than coronals as a whole (§5.1). We also discuss the role of sonorancy agreement in SPA restrictions, showing that it is only dento-alveolars that strongly depend on this agreement, whereas palatals as well as labials and velars do not (§5.2). However, there is one cross-linguistically observable restriction on the combinability of coronal obstruents pertaining to sibilants, and this sibilancy avoidance is another major and previously unreported finding discussed in §5.3. Finally, §5.4 briefly addresses the problem of the labial-coronal effect and other combinations that are *favored* cross-linguistically instead of being avoided. In the final section of this paper, we evaluate various models of SPA, proposing our formulation, and briefly discuss the etymological nature of native Slavic roots with homorganic CVC sequences violating SPA.

## 2. Data and Methodology

It is generally agreed that SPA is a constraint on morpheme (root) structure (Booij 2011), and that SPA violations across morpheme boundaries are often

more readily tolerated (P&S). However, while word lists are easily obtainable for many languages, morpheme lists are either too small or simply lacking. Our analysis relies on word lists but considers only those $C_1VC_2$ sequences in which the V element is the first syllabic nucleus of a word, namely #$C_1VC_2$, #$CC_1VC_2$, #$CCC_1VC_2$, and more complex patterns, irrespective of whether $C_2$ belongs to the first or second syllable (# is a word boundary). Henceforth, we will just speak about CVC or $C_1VC_2$ sequences. The choice of the first syllable has been motivated by the fact that these syllables are less likely to contain a morpheme boundary, at least in Slavic and other IE languages, which prefer suffixation over prefixation. In this section, we describe our Slavic (§2.1), non-Slavic Indo-European, and non-IE language data (§2.2), the way phonemes of these languages are classified (§2.3), and the way CVC distribution is statistically evaluated with Yule's Q (§2.4).

## 2.1. The Slavic Material

Our analysis of the SPA effect rests on the Slavic vocabulary included in the indices of the etymological dictionary of the Old Church Slavonic (OCS) language, *Etymologický slovník jazyka staroslověnského* (Havlová et al. 1989–2022; henceforth *ESJS*), and on the OCS headwords in the dictionary. The dictionary painstakingly lists all known cognates of OCS words in every Slavic language, thanks to which the indices are, with some reservations to be mentioned presently, collections of etymologically interrelated vocabularies. *ESJS* also offers reconstructions of Proto-Slavic roots underlying the OCS lexemes, but the indices provide many more Proto-Slavic reconstructions than OCS headwords. This is because the dictionary evaluates all major etymological explanations, some of which are highly conjectural. Since we want to cover the Proto-Slavic stage, we have decided to rely instead on the list of Proto-Slavic reconstructions (headwords) from Derksen's (2008) dictionary.

Though largely etymologically related, the *ESJS* vocabularies are not equivalent. First, not all lexemes of all Slavic languages with a reconstructable Proto-Slavic etymon have an attested OCS cognate. Furthermore, the indices contain some other Slavic words mentioned in support of the etymological explanations of the OCS words, which are not necessarily cognates or whose relation to a particular OCS lexeme is disputed. Moreover, *ESJS* explains all OCS vocabulary, including loanwords. Their equivalents are selectively mentioned for the other Slavic languages and hence recorded in the indices. Finally, the wordlists consist of basic (lemma) forms, though occasionally orthographic and other variants are also included. Derivatives are not mentioned in the dictionary and its indices unless they are etymologically interesting.

Simply put, the *ESJS* indices provide a sample of Slavic cognate vocabulary that is fairly exhaustive regarding the languages covered and reasonably large for each of the languages considered. Besides Old Church

Slavonic and Proto-Slavic, the languages are Bulgarian, Macedonian, Serbo-Croatian, and Slovenian for the South branch; Slovak, Old Czech, Modern Czech, Upper Sorbian, Lower Sorbian, Polabian, Pomeranian, and Polish for the West branch; and Belarusian, Ukrainian, Old Russian, and Modern Russian for the East branch. The Pomeranian data omit Kashubian. Slovincian is also excluded because its vocabulary listed in *ESJS* is too small. For the same reason, we leave out all words from older or immediate stages of the Slavic languages except those of Old Russian and Old Czech, which are sufficiently represented in the dictionary. Polabian has the smallest lexical dataset (745 word-initial CVC sequences), and Modern Russian has the largest one (4,259 sequences). On average, there are 2,195 word-initial CVC sequences for every variety.

All the wordlists were transformed into phonological transcription by simple grapheme-to-phoneme substitutions. Though this must have produced some errors, the procedure can be regarded as essentially reliable thanks to the phonological spelling of the languages. Additionally, we will refer to other lexical databases of some Slavic languages that have been processed similarly or have a phonetic/phonological transcription.

## 2.2. The Non-Slavic IE and Non-IE Material

To enrich our analysis, we have collected phonologically or phonetically transcribed lexical material for 44 non-Slavic IE languages and 38 non-IE languages from various lexical sources.[2] Preference has been given to databases that include at least 1,000 lexical items and provide phonetic or phonological transcription for these items. When such a source is unavailable or unknown to us, we have chosen languages where the phonological form could be reliably inferred from the orthography (e.g., Lithuanian). The databases differ in size (ranging from 859 CVC sequences for Chukchi to 282,512 sequences for Norwegian; 26,960 sequences on average), contents (lemmas only or inflected forms included; loanwords included or excluded), detail (broad or narrow transcription), perspective (phonetic or phonological), and last but not least, reliability (probability of transcription errors). Thus, the data must be treated with some caution. The non-Slavic IE sample also includes a list of Proto-Indo-European reconstructions and some other reconstructed varieties.

While we have aimed for exhaustivity in the case of the IE languages, the selection of the non-IE languages has been motivated by data availability, their size, and the desire to cover as many families as possible. Consequently, the non-IE sample is not representative because some families are instantiated by

[2] The reader will find the list of sources for all the lexical data as well as the statistical evaluation of the data at https://ojs.ung.si/index.php/JSL/article/view/469/345. The datasets themselves are not provided because they are either publicly available or protected by copyright.

more than one language (especially the Austroasiatic family), while other families contain only one (e.g., the Dravidian family) or no representative (e.g., the Tupian family). The primary purpose of the non-IE languages is to illustrate the nature of SPA outside the Indo-European family.

## 2.3. Phonological Classification

We will describe the distribution of the non-syllabic consonants in CVC sequences/combinations (henceforth just sequences/combinations unless specified otherwise). The nuclear V segment includes syllabic consonants. Long and nasalized vowels are counted as singular Vs. Diphthongs are generally interpreted as single Vs (always in the Slavic languages), but they are sometimes treated as two-phoneme sequences depending on the particular database used or the nature of the diphthongs in a given language (e.g., in Danish; cf. Basbøll 2005). Stress, tone, and other prosodic features are ignored for the sake of simplicity, though they may affect the strength of SPA in some languages (see Frisch 1997 and Dmitrieva 2008 on English, and van Goch 2010 on Dutch).

The non-syllabic consonants (i.e., the Cs; henceforth just consonants) are divided into four major place classes for which we will use the labels labials, dentals, palatals, and dorsals, symbolized with letters P, T, Č, and K, respectively (we distinguish Č for palatals from C for consonants in general). Following P&S, we unite dentals and palatals on the one hand and labials and dorsals on the other. P&S call these superclasses medials and peripherals, respectively, but we prefer the more traditional labels coronals and non-coronals (see §5 for a further discussion of coronals).

The labial class consists of the consonants in which at least one lip is the primary moving articulator, including the labio-velar approximant /w/, which agrees with the classification of P&S and Mayer et al. (2010) as well as with the findings of Kumagai (2020). The dorsal (short for dorso-guttural) class gathers the consonants articulated in the back of the oral cavity or in the throat: velars, uvulars, pharyngeals, epiglottals, and glottals. In some languages (such as the Semitic ones; see, e.g., McCarthy 1994), the throat consonants may behave as a separate class in the consonantal distribution. Since the Slavic languages virtually lack these segments, we refrain from positing a separate place class here. The velar nasal [ŋ] is classified as an allophone of the dental /n/ before velars in the Slavic languages. In the non-Slavic ones, it is usually treated as a separate dorsal segment (depending on the language or the database analyzed).

The palatal class comprises consonants produced with a raised tongue middle in a post-alveolar or palatal articulatory region (Ladefoged and Maddieson 1996: 14–15), namely, post-alveolars like /ʃ/, alveolo-palatals like /ɕ/, and true palatals like /c/, /ɟ/, /ɲ/, /j/ and like /ç/, /ʝ/ (Hall 1997 separates the latter

two from the former four; cf. also Rocasens 1990). We discuss PALATAL systems in more detail in §4.1. Finally, the DENTAL (short for dento-alveolar) class contains (inter)dentals, (pre-)alveolars, and retroflexes.[3] Retracted sibilants are also classed here. The classification of retroflexes requires a few notes. Although retroflexion is often understood as a separate place of articulation (which is reflected in the IPA chart), some linguists view it as an articulatory gesture (Hamann 2002). The exact target of this gesture ("place of articulation") varies across languages and speakers but tends to be in the post-alveolar region. It would thus make sense to classify retroflexes as PALATALS. However, unlike post-alveolars, alveolo-palatals, and true palatals, retroflexes do not have a raised (domed) tongue middle. Moreover, retroflexes may be interpreted as allophones of dentals/alveolars in some languages, such as Norwegian, Swedish, or Kukatja. We, therefore, place retroflexes into the DENTAL class.

It has been argued that Polish, Russian, Lower Sorbian, and Serbo-Croatian also contain retroflex consonants (Zygis 2003), namely those affricates and fricatives that are usually transcribed in Slavic literature with the symbols /č/, /š/, /ž/ or /t͡ʃ/, /ʃ/, /ʒ/ and that are generally described as post-alveolar sibilants. Articulatorily and perceptually, these segments differ slightly from "true" retroflexes, for example, in Indo-Aryan languages. We will demonstrate in §5.3 that their distribution parallels the distribution of other PALATALS in Polish. Consequently, the Slavic retroflex/post-alveolar sibilants are included in the PALATAL class.

Another problem concerns the classification of palatalized consonants (such as /sʲ/ in Russian) and other consonants with secondary articulation. Since secondary articulation is often contextually determined (allophonic), the standard treatment is to group them with their non-palatalized counterparts (see Padgett 1992 and Pozdniakov 2007 on Russian and Pierrehumbert 1993 on Arabic). The same approach will be followed here, though the problem needs to be considered in more detail.

In the rest of the paper, the terms LABIAL, DENTAL, PALATAL, DORSAL, CORONAL, and NON-CORONAL stand for the phonological classes just defined. They

[3] In contrast to P&S, we place /s/ as well as /z/, /t͡s/, and /d͡z/ among DENTALS, not PALATALS (P&S: 312, fn 4; see also Pozdniakov 2007). P&S support this choice by stating that the /s/ fricative often appears in the /ʃ/ slot in African languages, further noting that "[the reclassification of /s/ as a DENTAL] would have slightly changed the figures, but not the tendencies [for avoidance]". However, elsewhere (p. 341) they admit that the classification of /s/ is the most problematic one, which might have been a source of some discrepancy. Note also that Mayer et al. (2010) classify /ʃ/ and /ʒ/ as DENTALS but /t͡ʃ/ and /d͡ʒ/ as PALATALS without any explanation. For P&S and us, all these four consonants are PALATALS. Another case of disagreement concerns liquids; Mayer et al. (2010) treat /r/ and /l/ as PALATALS, whereas P&S treat them as DENTALS. The latter is what we also do (together with most other studies).

are written with small capitals, while regular letters indicate that a more traditional phonetic meaning of these terms is implied or that we refer to other phonological conceptions (the latter especially in the case of coronals). Thus, when we state that some language's PALATAL class contains true palatal obstruents like /c/, /ɟ/ and post-alveolar sibilants like /ʃ/, /ʒ/, we mean that the consonants produced at the two articulatory targets (the hard palate and the back of the alveolar ridge) are grouped into one SPA identity class.

Following Yip (1989), we understand an SPA identity place class as a well-defined set of consonants for which the following hypotheses are assumed.

(1) Hypotheses concerning SPA identity classes

a. CVC sequences consisting of consonants from the same identity class (henceforth: homorganic CVC sequences) are statistically underrepresented (avoided).

b. CVC sequences consisting of consonants from different identity classes (henceforth: heterorganic CVC sequences) are not statistically underrepresented; they are either overrepresented (favored) or show no statistically significant distributional anomalies.

The corroboration and/or rejection of the hypotheses in (1) depends on how identity classes are defined, to which problem we will return in §5. Another question is how statistical underrepresentation is determined.

## 2.4. Statistical Evaluation

Linguists have put forth several ways to determine whether the distribution of some elements differs from some norm, and whether this difference is not just a random distortion in the data population. One standard measure is a comparison of the observed frequency (O) and the expected frequency (E). The former is the actual frequency of some combination of elements, whereas the latter is a theoretical frequency the combination would have had if it were derived from the frequencies of the elements in the combination. An easy comparison of O and E is obtained by dividing these two frequencies, and this simple method has been used in most SPA studies, at least since Pierrehumbert (1993), though it was already employed in other analyses (e.g., Janson 1986). An O/E ratio greater than one means overrepresentation (a combination occurs more frequently than expected); a ratio of less than one implies underrepresentation (it occurs less frequently than expected).

Since it is unlikely that the observed frequency of a combination would equal its expected frequency in real data, the O/E ratio always indicates that a combination occurs more or less often than expected. We must have criteria

for deciding what discrepancy from the equation is statistically significant and linguistically interesting. This can be determined by statistical measures such as the $\chi^2$ test, often used in this connection, or we can set up some arbitrary and reasonable limits. The latter was done by P&S, who regarded the situation when O was different from E by at least 15% as interesting; an O/E ratio below this threshold was interpreted as falling within the range of natural distortions in the data, that is, as no evidence for overrepresentation or underrepresentation.

Although the O/E ratio is a simple and intuitively interpretable test, it has been subject to criticism. For example, Wilson and Obdeyn (2009) demonstrated that the measure was mathematically flawed because it confounded co-occurrence restrictions with positional probabilities. They argued for an alternative loglinear model, but it was proven to suffer from the same weaknesses (Stanton and Stanton 2022). Other, linguistically more profound arguments against the O/E ratio are laid out in Grotberg's (2022) dissertation, which discusses several statistical measures used for the analysis and evaluation of SPA.

Most previous studies have shown that CORONAL (DENTAL) homorganic CVC sequences are less often underrepresented than LABIAL or DORSAL ones, or that their underrepresentation effect is less strong than that for LABIALS and DORSALS. At the same time, it is well known that CORONALS are present in nearly every phoneme inventory (Maddieson 1984). Our material confirms this omnipresence and further testifies that the average frequency of nearly all combinations containing a DENTAL is greater than the average frequency of any other combination type. Now, Grotberg (2022: 54ff.) demonstrates that the O/E test is sensitive to this skewed distribution of CORONALS as opposed to the NON-CORONALS so that it underestimates the amount and force of avoidance among CORONALS (see the next paragraph). To ward off this bias, Grotberg proposes another statistic known as Yule's Q or Yule's coefficient of association, which is common in psychological or sociological studies, and which does not suffer from the aforementioned dependence.

Yule's Q belongs to the odds ratio family that quantifies the strength of association between two events, which for us is the occurrence of a consonant $C_1$ with a consonant $C_2$ in a $C_1VC_2$ sequence. Both tests are calculated from 2×2 contingency tables such as Table 1 (on the following page), with four values: A, B, C, and D (the shaded area). Should we want to calculate Yule's Q for combinations of two consonants $C_1$ and $C_2$ (e.g., LABIALS co-occurring with other LABIALS in a CVC sequence), then value A is the observed frequency of these combinations. Value B is the observed frequency of $C_1$ occurring with all other consonants but $C_2$ (LABIALS with NON-LABIALS). Value C is the frequency of the reverse combinations, that is, of all consonants other than $C_1$ occurring with $C_2$ (NON-LABIALS with LABIALS). Finally, value D corresponds to the frequency of all remaining combinations (NON-LABIALS with NON-LABIALS). The sum of A,

B, C, and D equals the frequency of all CVC combinations (Total). The sums of the particular rows and columns are marginal totals (e.g., the total frequency of labials). Since dentals are so frequent cross-linguistically, the marginal totals for any combination involving a dental will be greater than the other marginal totals. The O/E ratio and the $\chi^2$ test are sensitive to this discrepancy, but the odds ratio and Yule's Q are not.

**Table 1**. A 2 × 2 contingency table for the calculation of Yule's Q of a $C_1VC_2$ combination

| | $C_2$ | $\neg C_2$ | **Marginal totals** |
|---|---|---|---|
| $C_1$ | A | B | A + B |
| $\neg C_1$ | C | D | C + D |
| **Marginal totals** | A + C | B + D | Total (A + B + C + D) |

The odds ratio (OR) is the ratio between two odds, A to B and C to D, that is, the ratio of the frequency of $C_1$ occurring with $C_2$ to the frequency of $C_1$ occurring with consonants other than $C_2$, and of the frequency of consonants other than $C_1$ occurring with $C_2$. Two equivalent formulas for calculating OR are given under (2a). Yule's Q is a transformation of OR that can be obtained either from OR or directly from values A, B, C, and D (2b).

(2) a. Odds Ratio

$$OR = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{AD}{BC}$$

b. Yule's Q

$$Q = \frac{OR - 1}{OR + 1} = \frac{AD - BC}{AD + BC}$$

The range of possible results is zero to infinity for OR and −1 to +1 for Yule's Q. It is this finite and symmetrical range that makes Yule's Q a more helpful tool for expressing the strength of association between two consonants (O/E also ranges from zero to infinity). When there is no association, the difference between the frequency of $C_1VC_2$ (labials with labials) multiplied by the frequency of $\neg C_1V\neg C_2$ (non-labials with non-labials) equals the frequency of $C_1V\neg C_2$ (labials with non-labials) multiplied by the frequency of $\neg C_1VC_2$ (non-labials with labials). It means that $C_1$ is equally associated with $C_2$ and all other consonants. In actual data, this zero Yule's Q is probably only a theoretical possibility; there will always be some deviation from zero in either direction.

Positive Yule's Q expresses the strength of association between $C_1$ and $C_2$, whereas negative Q expresses the strength of disassociation. In the first case,

$C_1$ is more likely to occur with $C_2$ than other consonants; in the latter, it is less likely. If $C_1$ occurs with $C_2$ (i.e., A is non-zero) but does not combine with other consonants (i.e., either B or C or both are zero), and at the same time there exists at least one combination of $\neg C_1$ and $\neg C_2$ (D is non-zero), Yule's Q equals 1. This is the complete association of $C_1$ with $C_2$. On the other hand, if $C_1$ does not combine with $C_2$ (A is zero) but combines with at least one other consonant (B or C or both are non-zero), and at the same time there are combinations of $\neg C_1$ and $\neg C_2$ (D is non-zero), Yule's Q is −1. It is the complete disassociation of $C_1$ with $C_2$ or the absence of $C_1VC_2$ combinations.[4]

Again, in real frequencies of CVC sequences, it is rare to obtain the values of +1 and −1, though both cases appear in our data, the latter being more frequent. For example, Otomi has one combination of a PALATAL obstruent with another PALATAL obstruent but no combination of a PALATAL obstruent with any other consonant; hence Q = 1. Afrikaans has no combination of two PALATALS, but PALATALS combine with other consonants; hence Q = −1. However, in most other cases, Q is a non-zero value between these two extremes. Table 2 on the following page gives an example of a positive Q calculated for CVC combinations of DENTALS with LABIALS in Polish. Table 3 on the following page is an example of a negative Q for combinations of DENTALS with DENTALS in the same language. The former value speaks for the association between DENTALS and LABIALS, and the latter speaks for the lack of association between DENTALS and DENTALS.

Since it is practically impossible to determine Yule's Q for all possible events (for all CVC sequences in all words), we do not know the true value of the strength of association between two variables. We can only determine an estimate of Yule's Q based on a particular sample. These estimates will vary from sample to sample, but the Q values themselves do not tell us how good they are. One way to evaluate the reliability of an estimate is to determine its confidence interval (CI). A 95% CI means that, were we to calculate Yule's Q for a theoretically infinite number of other samples, 95% of those samples would be expected to have a value in the reported range. It is thus highly probable that the true value of Yule's Q lies within the range reported as the 95% confidence interval.[5]

[4] If both A and B or both A and C are zero, the Yule's Q formula fails because we divide by zero. The standard procedure is to add 0.5 or even 0.1 to each value (Hollander et al. 2014: 512). Although this practice produces a meaningful Q value, it shadows the real linguistic significance of the underlying CVC distribution. A zero value means that either $C_1$ does not combine with any consonant or that $C_2$ does not combine with any other consonant, or both. These situations cannot be evidence for SPA because languages of this kind restrict the occurrence of both homorganic and heterorganic CVC sequences involving a given consonant.

[5] Yule's Q is essentially a normal distribution (Grotberg 2022: 94, fn 4). The lower and upper limits of CI are calculated from the lower and upper limits for Odds Ratio (OR),

**Table 2**. Calculation of Yule's Q and its confidence interval for combinations of DENTALS with DORSALS in Polish

| | | $C_2$ | |
|---|---|---|---|
| | | DORSAL | NON-DORSAL |
| $C_1$ | DENTAL | 83 | 566 |
| | NON-DENTAL | 191 | 1660 |

$$Q = \frac{83\times1660 - 566\times191}{83\times1660+566\times191} = 0.12$$

95% CI = [–0.02, 0.25]

**Table 3**. Calculation of Yule's Q and its confidence interval for combinations of DENTALS with DENTALS in Polish

| | | $C_2$ | |
|---|---|---|---|
| | | DENTAL | NON-DENTAL |
| $C_1$ | DENTAL | 362 | 533 |
| | NON-DENTAL | 759 | 846 |

$$Q = \frac{362\times846 - 533\times759}{362\times846+533\times759} = -0.14$$

95% CI = [–0.22, –0.06]

θL and θU, respectively (Hollander et al. 2014: 521):

$$Q_{Lower} = \frac{\theta L - 1}{\theta L + 1}, \text{ where } \theta_L = e^{\ln(OR) - Z \times SE\ln(OR)}$$

$$Q_{Upper} = \frac{\theta U - 1}{\theta U + 1}, \text{ where } \theta_U = e^{\ln(OR) + Z \times SE\ln(OR)}$$

The Z value defines the appropriate confidence limits, such as 2.58 for the 99% CI or 1.96 for the 95% CI. The constant *e* is Euler's number. The $SE_{ln(OR)}$ value is the standard error calculated from the following formula in which A, B, C, and D are the frequency values in a contingency table (see Table 1 on p. 185):

$$SE_{ln(OR)} = \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}$$

Thus, the 95% CI upper and lower values for Table 2 are calculated as follows:

$$SE_{ln(OR)} = \sqrt{\frac{1}{83} + \frac{1}{566} + \frac{1}{191} + \frac{1}{1660}} = 0.14 \qquad OR = \frac{AD}{BC} = \frac{83\times1660}{566\times191} = 1.27$$

$$\theta_L = e^{\ln(OR) - Z \times SE\ln(OR)} = e^{\ln(1.27) - 1.96 \times 0.14} = 0.97, \text{ hence } Q_{Lower} = \frac{\theta L - 1}{\theta L + 1} = \frac{0.97-1}{0.97+1} = -0.02$$

$$\theta_U = e^{\ln(OR) + Z \times SE\ln(OR)} = e^{\ln(1.27) + 1.96 \times 0.14} = 1.68, \text{ hence } Q_{Upper} = \frac{\theta U - 1}{\theta U + 1} = \frac{1.68-1}{1.68+1} = 0.25$$

The wider the CI, the greater the chance that it overlaps zero and includes both positive and negative values, which means that neither genuine overrepresentation (expressed by positive values) nor genuine underrepresentation (negative values) can be excluded. Returning to tables 2 and 3, we see that both Q values for Polish (0.12 and −0.14, respectively) are almost equally distant from zero (the strength of association and disassociation is very similar for both combination types), but we also witness that the CI for the Q value for the combinations of DENTALS with DORSALS includes zero, while the CI for the combinations of two DENTALS does not. Consequently, we lack reliable evidence for an association between DENTALS and DORSALS in this language. On the other hand, we can claim with some confidence that combinations of DENTALS with other DENTALS are avoided there.

The CI width is not only dependent on our confidence level but also on the size of our sample and the distribution of certain CVC combinations. Small samples will produce wider CIs than larger samples at the same confidence level. Similarly, CVC sequences containing very infrequent consonants will likely produce greater CIs. Considering our Slavic data, we have decided to work with 95% CIs. In Slovenian, for example, wider intervals make the Yule's Q value for DORSAL homorganic sequences insignificant, but the value is significant for a more extensive alternative dataset with the same CI.

Let us return to Yule's Q proper. The more it deviates from zero, the greater the (dis)association there is between $C_1$ and $C_2$. Following Rosenthal (1996: 51), Grotberg (2022: 92) distinguishes several bands of the strength of effect size. The values of $|Q| < 0.2$ correspond to a very weak (dis)association, and those of $0.2 < |Q| < 0.43$ to a weak (dis)association. We may conflate all stronger effects into one strong (dis)association category. Although Yule's Q for homorganic CVC sequences exceeds |0.43| in many languages of our sample, this value is rarely reached for non-homorganic sequences. In the tables to come, we parenthesize the Q coefficients whose 95% CIs do not make them reliable estimates. These values are included in the counts of average Yule's Q for language classes and subclasses.

When there is a significant disassociation, that is, negative Yule's Q with a CI not containing zero, we will speak for convenience and stylistic diversity about a CVC sequence being disfavored, dispreferred, underrepresented, or avoided. On the other hand, a CVC sequence will be described as being favored, preferred, overrepresented, or not avoided when its Yule's Q is positive and the CI does not contain zero (significant association).

Sometimes a combination of two consonants turns out not to be attested at all in our data even though some languages in question combine these phonemes with other consonants. Afrikaans was mentioned above as an example of such a language, one lacking homorganic PALATAL sequences. Although Yule's Q for these cases is −1, which equals the Odds Ratio of 0, its confidence interval cannot be calculated by the standard procedure outlined in footnote 5

because the logarithm of 0 is undefined. The measure mentioned in footnote 4 (adding 0.1 to values A, B, C, and D) makes the CI calculable, but in most cases the interval is too wide to include zero. It is either due to the small size of some of our data or due to the nature of the Yule's Q statistic. Unless additional data are examined, we cannot confidently conclude that such combinations are completely avoided. We have therefore decided not to use languages that lack a certain CVC sequence (we will call them –1 languages) in the calculations of average Yule's Qs. Yet we mention them, as they are obviously of linguistic interest.

In what follows and in the online supplementary materials, Yule's Q values and CIs are calculated from the frequencies of the CVC sequences in which V is the first word nucleus as recorded in the lexical databases (dictionaries) used. A CVC sequence may thus occur in several words containing the same morpheme.

## 3. Pure SPA in Slavic Languages

Our default hypothesis is that CVC sequences containing consonants from the same identity class are avoided in most or all Slavic languages. Evidence for this avoidance will be negative Yule's Q whose CI does not include zero. We assume four identity classes, labials, dentals, palatals, and dorsals, as defined in §2.3. This is what P&S call Pure SPA, in which only place features are involved. The role of non-place features will be considered in §5. A language will be said to exhibit a pure SPA effect if at least three of the four identity classes are avoided (i.e., the majority).

Table 4 (on p. 191) provides Q coefficients for the four identity classes in the 18 Slavic languages.[6] Since the lexemes for which we quantify the CVC distribution are essentially cognates, the differences between the Q values can be understood as an approximation of the development of SPA from Proto-Slavic to its individual offspring.[7] Recall that the statistics have been calculated only for the CVC sequences where V is the first word nucleus. The default hypothesis is corroborated for labials, dentals, and dorsals, but not palatals (see the parenthesized values, which are statistically non-significant). Unless the latter class is wrongly delimited, which possibility we address in §5, the Slavic languages are characterized by the absence of palatal SPA.

[6] For brevity, we will use the following abbreviations in this section: PSl = Proto-Slavic, OCS = Old Church Slavonic, Bul = Bulgarian, Mac = Macedonian, SCr = Serbo-Croatian, Sln = Slovenian, Slk = Slovak, OCz = Old Czech, MCz = Modern Czech, USo = Upper Sorbian, LSo = Lower Sorbian, Plb = Polabian, Pom = Pomeranian, Pol = Polish, Bel = Belarusian, Ukr = Ukrainian, ORu = Old Russian, MRu = Modern Russian.

[7] See also Pystynen (2014) and Cathcart (2023) on diachronic views on SPA.

Labial avoidance is encountered in all Slavic varieties (as well as most other languages; see below). On average, it is stronger in the South and East varieties (average Yule's Q is −0.60 in both branches) than in the West branch (−0.39). PSl also displays a strong avoidance effect here (−0.70). The decreased strength within the West languages may be a function of the development of the original PSl palatalized labials, which have been dissolved into two-phoneme sequences such as /pj/ in Czech or Polish. Dental avoidance is also observable in all varieties, and if we compare the PSl state with the descendant languages, we could say that the development has been toward a decrease in the strength of this type of SPA. The weakening is again most obvious in the West languages; it may relate to the development of palatals out of dentals/alveolars.

The dorsal homorganic sequences are also widely disfavored, but the Yule's Q test does not provide reliable evidence for an avoidance effect in ORu, OCz, Plb, and USo. For the first three varieties, this result looks suspicious and must be a consequence of the small size of our data. It seems unlikely that ORu and OCz would fail to avoid K_K when these sequences are disfavored both in the previous and subsequent developmental stages (in PSl, and MRu and MCz, respectively). Our datasets for ORu and OCz are among the smallest compared to the other Slavic languages (see Table 4 on the following page). Likewise, the small data size is probably why Plb does not avoid K_K. On the other hand, the absence of dorsal SPA in USo appears to be a genuine feature of this language independent of the data size. USo is the only Slavic language having dorsal sonorants, namely the uvular rhotic /ʁ/ (and its palatalized variant) corresponding to the dental/alveolar rhotic /r/ in the other languages (Howson 2017). When USo /ʁ/ is reclassified as a dental, the Yule's Q statistic produces a high negative significant value for K_K (−0.56), and the value for T_T increases, too (−0.28 instead of −0.16). Thus, though the USo rhotics may be classified as dorsals based on their articulation, phonotactically they do not behave like these consonants, which may be due to them being dentals in origin.

The attitude of the Slavic languages toward palatal homorganic sequences requires more attention. Only three languages significantly disfavor these sequences (Bul, Sln, and SCr); two significantly favor them (LSo and Pol). We lack reliable statistical evidence for this avoidance (or preference) in the remaining 13 languages, though. This diversity may be explained by both the common development within particular Slavic subbranches and the nature of the palatal systems, though the problem is more complex, as is suggested by our alternative data discussed further below.

Within Slavic, we observe five kinds of palatal systems, the breakdown of which is reproduced in Table 5 (on the following page): types P2, P3, P4, P6, and P9 (see the next section and Table 9 on p. 199 for the other types). Each type is characterized by the presence of the consonant classes we place into

**Table 4**. Yule's Q values for homorganic CVC sequences in the Slavic languages

| | Bul | Sln | SCr | PSl | Mac | OCS | Bel | Ukr | MRu |
|---|---|---|---|---|---|---|---|---|---|
| **P_P** | −0.69 | −0.62 | −0.52 | −0.70 | −0.62 | −0.57 | −0.65 | −0.59 | −0.56 |
| **T_T** | −0.44 | −0.25 | −0.34 | −0.47 | −0.32 | −0.34 | −0.43 | −0.42 | −0.35 |
| **Č_Č** | −0.33 | −0.27 | −0.22 | (−0.42) | (−0.26) | (−0.10) | (−0.25) | (−0.17) | (0.04) |
| **K_K** | −0.41 | −0.22 | −0.35 | −0.33 | −0.31 | −0.49 | −0.50 | −0.53 | −0.49 |
| **N** | 2,809 | 2,899 | 3,831 | 1,259 | 1,584 | 2,073 | 2,490 | 2,814 | 4,259 |

| | Slk | MCz | Pom | LSo | Pol | ORu | OCz | Plb | USo |
|---|---|---|---|---|---|---|---|---|---|
| **P_P** | −0.59 | −0.42 | −0.29 | −0.22 | −0.16 | −0.61 | −0.61 | −0.51 | −0.29 |
| **T_T** | −0.30 | −0.25 | −0.14 | −0.19 | −0.14 | −0.40 | −0.20 | −0.30 | −0.16 |
| **Č_Č** | (−0.08) | (−0.01) | (−0.09) | 0.17 | 0.11 | (−0.10) | (−0.05) | (−0.27) | (0.01) |
| **K_K** | −0.47 | −0.38 | −0.56 | −0.40 | −0.56 | (−0.18) | (−0.40) | (0.12) | (−0.07) |
| **N** | 2,354 | 3,189 | 1075 | 1,701 | 2,500 | 1,292 | 1,101 | 745 | 1,679 |

(P_P = labial sequences, T_T = dental sequences, Č_Č = palatal sequences, K_K = dorsal sequences. Statistically non-significant values are parenthesized. Shading highlights significant positive values. The row N gives the total number of CVC sequences in a particular language.)

**Table 5**. A partial typology of palatal systems, applied to the Slavic languages

| Type | /j/ | /ʃ/ | /ɲ/ | /c/ | /ç/ | /ɕ/ | O | Languages |
|---|---|---|---|---|---|---|---|---|
| P2 | • | • | | | | | | PSl, **Bul**, **Sln**, Bel, Ukr, ORu, MRu, USo, Plb |
| P3 | • | • | • | | | | | Pom |
| P4 | • | • | • | • | | | | Mac, OCS, OCz, MCz, Slk |
| P6 | • | • | • | | | • | | **SCr**, *Pol* |
| P9 | • | • | | | | • | | *LSo* |

(/j/ = palatal approximants, /ʃ/ = post-alveolar sibilants, /ɲ/ = palatal sonorants (nasals and liquids), /c/ = palatal plosives, /ç/ = palatal fricatives, /ɕ/ = alveolo-palatal sibilants, O = other segments (such as clicks). The languages in bold avoid palatal sequences; those in italics favor them; the rest show no distributional anomaly in either direction.)

the palatal identity class, namely the palatal approximant /j/, post-alveolar sibilants (/ʃ/, which includes both affricates and fricatives), palatal sonorants other than /j/ (/ɲ/, typically the nasal but also /ʎ/ in Slk or the /ř/ phoneme in MCz),[8] true palatal plosives (/c/), true palatal fricatives (/ç/, of the kind found in German but not attested in Slavic), alveolo-palatal sibilants (/ɕ/, e.g., in Pol, again including both affricates and sibilants), and some other segments (O, not found in Slavic).

Type P2 is the most common palatal system, limited to post-alveolar sibilants and /j/. The Č_Č avoidance has been evidenced in two of the nine languages belonging here (Bul and Sln). The other seven instances are all four East languages, PSl, USo, and Plb, each of them lacking palatal avoidance. This absence is shared with type P4, which differs from P2 in containing true palatal plosives (and sonorants). P4 consists of two South Slavic languages, Mac and OCS, and three West languages, OCz, MCz, and Slk (the Czech-Slovak subgroup).

All these facts support our assumption that the presence of palatal SPA is both related to the nature of palatal systems and to the common development of Slavic subbranches. First, palatal avoidance occurs in the South Slavic varieties that lack true palatal plosives; the non-South languages without true palatal plosives fail to disfavor Č_Č.[9] Second, palatal avoidance is also absent in languages with true palatal plosives irrespective of their affiliation. Finally, it is also absent in Pom, which has true palatal sonorants but not plosives. However, it is not obvious whether the lack of palatal SPA is really due to the sonorants because the non-Slavic languages belonging in P3 tend not to underrepresent Č_Č, while those belonging in P4 do (see Table 9 on p. 199).

Let us move to the remaining three varieties, SCr, Pol, and LSo. The typology in Table 5 does not explain the nature of palatal SPA in these languages. SCr and Pol belong in P6, and LSo belongs in P9, but SCr avoids Č_Č, whereas Pol and LSo favor these sequences. All three of these languages have two sibilant series within the palatal class, the post-alveolars and alveolo-palatals.

[8] In a more minute typology, the palatal nasals and liquids could be separated, but since the purpose of this typology is to find out correlations between SPA and palatal systems, we have grouped them together. We have not found any such correlation in OCz, MCz, Slk, OCS, and SCr, which all possess palatal liquids. Neither has any such correlation been detected in 17 other non-Slavic languages that possess these sonorants. The interpretations of the phoneme systems are in accord with the illustrations of the IPA published in the *Journal of the International Phonetic Association*, which are available for most Slavic languages.

[9] PSl transcends the classification into West, East, and South branches but belongs in P2 and does not avoid Č_Č. It is worth noting that Yule's Q for Č_Č in PSl reaches the greatest negative value among the Slavic languages (−0.42), and its confidence interval, which is rather large due to the small size of the PSl data, mostly occupies the negative area ([−0.72, 0.01]). This suggests avoidance.

In SCr, however, alveolo-palatals are limited to affricates, whereas the other two languages have both affricates and fricatives. Thus, it seems that the preference of Č_Č encountered in Pol and LSo is a West Slavic innovation related to the development of a rich system of alveolo-palatal sibilants in addition to post-alveolar sibilants.

The unique nature of Č_Č in Pol can further be illustrated with the help of a larger set of lexical data taken from WikiPron dictionaries extracted from Wiktionary (107,924 first-syllable CVC sequences).[10] Table 6 on the following page provides Q values for the combinations of the three sibilant series Pol possesses, the dento-alveolars /t͡s/, /d͡z/, /s/, /z/; the post-alveolars/retroflexes /t͡ʂ/, /d͡ʐ/, /ʂ/, /ʐ/; and the alveolo-palatals /t͡ɕ/, /d͡ʑ/, /ɕ/, /ʑ/. All combinations of the dento-alveolar sibilants with any other sibilants in any order are avoided (including repetitions of dento-alveolar sibilants). In contrast, three of the four combinations of the post-alveolar and alveolo-palatal sibilants (the shaded cells) are favored; it is only sequences of an alveolo-palatal plus a post-alveolar that are avoided.[11] These facts suggest that the post-alveolar and alveolo-palatal sibilants phonotactically belong to one class distinct from that of the dento-alveolar sibilants. This conclusion supports our decision to place Pol and LSo post-alveolar sibilants into the palatal class, even though some linguists (Hamann 2002; Zygis 2003) have interpreted these sibilants as retroflexes (which are placed in the dental class in the other languages of our sample).

Having access to some other lexical databases, we can examine the nature of SPA even further. Our primary datasets of the words drawn from the *ESJS* indices are essentially cognates, native to the Slavic family, and basic, uninflected, and underived forms. However, the Slavic languages contain more vocabulary, including inflected and derived forms and many words imported from other languages. All these items may influence the SPA effect and its strength. Table 7 on the following page reproduces the Q values for alternative lexical databases for nine Slavic languages. The databases are not necessarily related to each other or unbiased. The MCz data contain both native and borrowed words extracted from several dictionaries; the MRu and Sln data are derived from headwords of two etymological dictionaries; the material for the remaining languages is again taken from WikiPron dictionaries, which may contain a rather large number of loanwords.

[10] See the online appendix for the sources of these additional databases, https://ojs.ung.si/index.php/JSL/article/view/469/345.

[11] The negative Q value for the CVC sequences of alveolo-palatals followed by post-alveolars may be due to a more general avoidance of the CVC sequences that begin with an alveolo-palatal. First, notice that the combinations of alveolo-palatals with alveolars and post-alveolars reach the greatest negative Q values. Second, the $C_1VC_2$ sequences in which $C_1$ is an alveolo-palatal are less common than the ones in which $C_2$ is such a consonant.

**Table 6**. Yule's Q values for combinations of sibilants in the Polish alternative data (107,924 CVC sequences)

| | Dento-alveol. | Post-alveolar | Alveolo-pal. |
|---|---|---|---|
| Dento-alveolar | −0.20 | −0.31 | −0.29 |
| Post-alveolar | −0.23 | 0.12 | 0.14 |
| Alveolo-palatal | −0.53 | −0.34 | 0.20 |

(Shading marks homorganic combinations of the sibilants belonging to the palatal class.)

**Table 7**. Yule's Q values for alternative data of selected Slavic languages

| | Bul | Sln | SCr | Mac | Bel | Ukr | MRu | MCz | Pol |
|---|---|---|---|---|---|---|---|---|---|
| **P_P** | −0.49 | −0.56 | −0.59 | −0.53 | −0.29 | −0.57 | −0.66 | −0.47 | −0.22 |
| **T_T** | −0.40 | −0.31 | −0.33 | −0.29 | −0.34 | −0.36 | −0.33 | −0.28 | −0.17 |
| **Č_Č** | −0.49 | (−0.06) | (0.07) | (−0.03) | −0.35 | −0.09 | −0.11 | +0.09 | 0.09 |
| **K_K** | −0.65 | −0.45 | −0.44 | −0.30 | −0.38 | −0.45 | −0.43 | −0.49 | −0.37 |
| **N** | 30,527 | 8,460 | 19,643 | 49,091 | 29,748 | 23,441 | 16,360 | 62,107 | 104,924 |

(The shaded values are markedly different from those reproduced in Table 4 on p. 191.)

The individual Q values for P_P, T_T, and K_K reproduced in Table 7 differ from those in Table 4 on p. 191 (= our primary data), but they are all significantly negative. Hence, the avoidance of these sequences is a stable feature of the Slavic languages, independent of the type of lexical data. However, what interests us most is Č_Č, and here we observe substantial differences (shaded in the table). We have previously seen that no other Slavic languages but Bul, Sln, and SCr avoid palatal homorganic sequences. The alternative data produce a slightly different picture. Sln and SCr do not disfavor these sequences, in which they agree with Mac. On the other hand, all three East Slavic languages show a palatal avoidance effect, though its strength is relatively weak. The *ESJS* data provide no evidence for this avoidance here. To complicate the situation even further, we have calculated Yule's Q for 3,750 native roots in MRu, using the quantitative data from Pozdniakov (2007), and these data do not indicate that Č_Č is avoided. Thus, the MRu situation recorded in Table 7 may be due to the presence of loanwords. Finally, the MCz data manifest a slight but statistically insignificant degree of *preference* for

palatal homorganic sequences. When we calculated Yule's Q for 7,609 native roots in MCz from a different source (see the online appendix), these data provided no evidence for preference or dispreference of Č_Č. All these alternative datasets suggest that loanwords, inflected words, and derivates influence the nature of palatal avoidance.

Despite the variability, one obvious conclusion is inferable from the facts presented in this section: the Slavic languages are not characterized by any strong or stable tendency to avoid homorganic combinations of palatals.

## 4. Pure SPA in Non-Slavic Languages

To see whether the state observed for Slavic is specific to this language group, we have evaluated lexicons for other languages, of which 44 belong to the IE family, and 38 are recruited from several non-IE families. We will not provide individual Yule's Q values for these languages; they can be found in the online appendix (see fn 2). The IE languages will be discussed in more detail because they are closely related to the Slavic languages, and because no other family is represented by so many members in our data. We believe the more comprehensive and exhaustive the analyzed language data are, the more useful information they can provide. We reject the idea that one language represents the whole family and that limited language representatives lead to meaningful, cross-linguistically valid generalizations ("universals"). In short, the non-IE languages only supplement our analysis, providing tentative evidence about the situation outside the IE world.

### 4.1. Non-Slavic IE Languages

Table 8 on the following page reproduces average Q values for Proto-Indo-European (PIE) and all IE branches, some of which are represented by one language only.[12] The table includes the Slavic branch too. The parenthesized numbers express how many languages have produced a reliable Q value; only these languages are included in the calculation of the averages. The Hellenic languages (Ancient and Modern Greek) do not have any palatals, at least in our data. Modern Greek palatal consonants are allophones of velars and partly

[12] Our analysis confirms, in essence though not in detail, the findings of previous SPA studies. See P&S, Cooper (2009), and Sandell (2015) on PIE; Berkley (2000) and Grotberg (2022) on Latin; Grotberg (2022) on Medieval Castilian; Plénat (1996), Berkley (2000), Rousset (2004), and Carrissimo-Bertola (2010) on French; MacNeilage et al. (2000) on Spanish; Frisch et al. (2004) on Italian; Carrissimo-Bertola (2010) on Portuguese; Berkley (1994, 2000), Frisch (1997), MacNeilage et al. (2000), and Dmitrieva (2008) on English; van Goch (2010) and Boll-Avetisyan and Kager (2014) on Dutch; Khan (2007) on East Bengali.

**Table 8.** Average Yule's Q values for homorganic CVC sequences in particular IE branches

| | PIE (n = 1) | Hellenic (n = 2) | Italic + Romance (n = 8) | Anatolian (n = 1) | Celtic (n = 4) |
|---|---|---|---|---|---|
| **P_P** | −0.58 (1) | −0.71 (2) | −0.65 (8) | −0.63 (1) | −0.60 (4) |
| **T_T** | −0.51 (1) | −0.42 (2) | −0.38 (8) | −0.44 (1) | −0.36 (4) |
| **Č_Č** | 0 (0) [1] | × | −0.25 (3) [1] | N.S. | 0 [1] |
| **K_K** | −0.32 (1) | −0.46 (2) | −0.49 (8) | −0.63 (1) | −0.42 (3) |

| | Baltic (n = 2) | Germanic (n = 15) | Albanian (n = 1) | Indo-Aryan (n = 8) | Armenian (n = 1) |
|---|---|---|---|---|---|
| **P_P** | −0.57 (2) | −0.56 (15) | −0.56 (1) | −0.53 (8) | −0.52 (1) |
| **T_T** | −0.43 (2) | −0.35 (15) | −0.27 (1) | −0.31 (7) | −0.19 (1) |
| **Č_Č** | −0.25 (2) | −0.50 (10) [4] | −0.13 (1) | −0.30 (6) [1] | N.S. (×) |
| **K_K** | −0.54 (2) | −0.33 (14) | −0.14 (1) | −0.52 (8) | −0.52 (1) |

| | Tocharian (n = 1) | Slavic (n = 18) | All non-Slavic (n = 44) | All IE (n = 62) |
|---|---|---|---|---|
| **P_P** | −0.32 (1) | −0.51 (18) | −0.58 (44) | −0.56 (62) |
| **T_T** | −0.20 (1) | −0.30 (18) | −0.36 (43) | −0.34 (61) |
| **Č_Č** | N.S. (×) | −0.27 (3) | −0.37 (22) [8] | −0.36 (25) [8] |
| **K_K** | −0.58 (1) | −0.43 (14) | −0.43 (42) | −0.40 (56) |

(The numbers in round brackets, or parenthesis, indicate how many languages with significant *negative* Q values have been used for the calculation. The numbers in square brackets express how many languages lack certain combinations altogether (the −1 languages); they have not been included in the average calculation. Square-bracketed values are not given when they equal 0. N.S. = not significant.)

of alveolars (Arvaniti 2007). Although Table 8 naturally invites us to compare the strength of the SPA effect across various IE branches and its development from PIE, we should keep in mind that the comparison is only approximate.[13] First, we compare average Yule's Q values calculated from samples of various sizes and contents. The Q values tend to be very similar to each other, and

[13] We would like to thank one of the reviewers for pointing this out.

their confidence intervals, moreover, largely overlap. Second, since Yule's Q is more or less normally distributed, we can say that one Q is larger or smaller than another, but the absolute difference between these two values does not really tell us how different the values are (Grotberg 2022: 93ff.).

Looking at the last three columns in Table 8, we witness that the non-Slavic IE languages are similar to the Slavic ones both in the extent and strength of the avoidance of P_P, T_T, and K_K as well as in the fact that palatals are again exceptional, though less obviously. All non-Slavic IE languages avoid P_P just like the Slavic ones do, and this type of avoidance is the strongest in nearly every branch. The exceptions are Anatolian (i.e., Hittite), Armenian, and Tocharian (B), in which dorsals equal or exceed the avoidance strength of labials. The average Q value for P_P is a little greater in the non-Slavic languages, which is due to the consistently higher values across the non-Slavic branches and due to the comparatively weaker avoidance of P_P in the West Slavic languages. Curiously, the average Q value for the non-Slavic languages agrees with the value of PIE (see the first column). It is also worth pointing out that four IE branches have developed a stronger avoidance effect for P_P than PIE. They are placed in the upper part of Table 8 (Hellenic, Italic + Romance, Anatolian, and Celtic).[14] All the remaining branches, including the Slavic one, are characterized by a decrease in the strength of labial avoidance.

As mentioned, K_K sequences show the strongest avoidance in Hittite, Armenian, and Tocharian B. On average, however, the strength of dorsal avoidance is virtually identical in the Slavic and non-Slavic languages. If we accept PIE and its Q values as an approximation of the original state of the IE languages, we notice that all the IE branches have evolved into having stronger avoidance of the dorsal homorganic sequences than PIE originally had. This may result from the development of the three PIE velar series (plain, palatalized, and labialized). Dorsal avoidance is not evidenced for Breton and Plautdietsch, though. While in Breton the lack of evidence may be a function of the small lexical sample (n = 888), the size of the Plautdietsch data is high enough to be reliable (n = 13,510). In fact, the absence of dorsal avoidance in this language may not be surprising at all because Germanic languages generally show weaker avoidance here compared to the other languages.

The dental CVC sequences are underrepresented in nearly all non-Slavic IE languages except for Pashto. However, this may only be a consequence of the small size of the Pashto data, which may also be the reason why this language fails to avoid palatal sequences (see below). At the same time, T_T sequences tend to be less strongly avoided than the non-coronal sequences, which we have already observed in the Slavic languages, and which is also true for the non-IE ones (see the next section). And if the K_K avoidance strength appears to have been increased in all IE branches except for the Germanic one, the

[14] Individually, only the Portuguese Q value for P_P does not exceed that for PIE.

dental avoidance strength has instead been relaxed in every IE branch and, in fact, in nearly every IE language. The languages in which Yule's Q is greater than that for PIE (−0.51) are Latin (−0.61) and Modern Greek (−0.55).[15] This change is likely a consequence of the rise of palatals other than /j/ (including true palatals and post-alveolars) in the particular IE varieties. Other factors must have played a role, too, because some IE languages have a rather limited palatal inventory.

This brings us to the last place class, palatals, whose homorganic sequences are again avoided in fewer languages than sequences of the other places. However, the portion of the avoidant languages is considerably larger than in the Slavic group. The avoidance is observed in 30 (68.18%) non-Slavic IE languages, of which eight lack Č_Č sequences altogether (Afrikaans, Gothic, Old Frisian, Proto-Germanic, Ossetian, Proto-Celtic, Latin, and PIE). In the Slavic family, there are only three such languages (16.67%).

The presence or absence of palatal avoidance thus apparently depends both on the languages' genealogy and the nature of their palatal inventory. In Table 9 on the following page, we present the full typology of the palatal systems across all three language groups, which was already partly reproduced in Table 5 on p. 191 for Slavic. The columns with IPA symbols stand for the consonant classes we group into the palatal SPA class.[16] The last four columns provide the ratio between the languages that significantly avoid Č_Č and all languages of a particular type and a particular language group.

[15] We have also calculated the Q value for T_T from the quantitative data for PIE used in P&S, which relied on a different material. The value is −0.49, which is still relatively high. Yule's Q for P_P is −0.66, for K_K, it is −0.46, and the value for Č_Č is not statistically significant. Note also that P&S are silent about their exact classification of the PIE consonants, but it is likely that they treat palatovelars as palatals, which is what Sandell (2015: 11, fn 16) does as well. Classifying our data in the same way, we get similar results to those of P&S. In contrast, Cooper (2009: 60) views palatovelars and labiovelars as dorsals with secondary articulations, which has also been suggested by one of the reviewers (see also Weiss 2016). We have followed this classification here, but it is worth noting that when the palatovelars are excluded from the dorsal class, Yule's Q for this class increases (−0.43 for our data).

[16] The sibilant classes /ʃ/ and /ɕ/ include both affricates and fricatives (post-alveolar and alveolo-palatal ones, respectively; see also Kokkelmans 2021 on sibilants in general). On the other hand, the class of true palatal fricatives /ɕ/ excludes affricates because they are either non-occurrent in our data or interpreted as variants of palatal plosives, with which they vary in some languages (e.g., in Albanian, Kolgjini 2004; or Hungarian, Siptár and Törkenczy 2000). Note also that the typology reflects the phonological databases we use and the phonological classification they employ. Thus, for example, Finnish is classified as a language with post-alveolar sibilants (contrary to Kokkelmans 2021) because our Finnish database contains a number of loanwords with these segments.

**Table 9**. Full typology of PALATAL systems

| Type | /j/ | /ʃ/ | /ɲ/ | /c/ | /ç/ | /ɕ/ | O | Slavic | Other IE | Non-IE | All |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P0 | | | | | | | | × | ×/2 | × | ×/2 |
| P1 | • | | | | | | | × | 6/7 | 0/1 | 6/8 |
| P2 | • | • | | | | | | 2/9 | 14/17 | 3/5 | 19/31 |
| P3 | • | • | • | | | | | 0/1 | 3/7 | 2/7 | 5/15 |
| P4 | • | • | • | • | | | | 0/5 | 2/3 | 3/4 | 5/12 |
| P5 | • | • | • | • | • | | | × | 0/3 | 1/1 | 1/4 |
| P6 | • | • | • | | | • | | 1/2 | × | 1/1 | 2/3 |
| P7 | • | • | • | | | | • | × | × | 0/1 | 0/1 |
| P8 | • | • | | • | | | | × | × | 1/1 | 1/1 |
| P9 | • | • | | | | • | | 0/1 | 1/1 | 1/3 | 2/5 |
| P10 | • | | • | • | • | | | × | 1/1 | 1/1 | 2/2 |
| P11 | • | | • | • | | | | × | × | 4/7 | 4/7 |
| P12 | • | | • | • | | • | | × | × | 3/3 | 3/3 |
| P13 | • | | • | | | • | | × | 1/1 | 1/1 | 2/2 |
| P14 | | • | | • | | | | × | 1/1 | × | 1/1 |
| P15 | | • | | | | | | × | × | 0/1 | 0/1 |
| P16 | | | | | | • | | × | 1/1 | 0/1 | 1/2 |
| Total | | | | | | | | 3/18 | 30/44 | 21/38 | 54/100 |

(/j/ = palatal approximants, /ʃ/ = post-alveolar sibilants, /ɲ/ = palatal sonorants (nasals and liquids), /c/ = palatal plosives, /ç/ = palatal fricatives, /ɕ/ = alveolo-palatal sibilants, O = other segments (such as clicks). The last four columns provide the ratio between the languages that avoid Č_Č and all languages of a particular type. The statistics include languages that lack PALATAL sequences.)

For the sake of completeness, the typology includes type P0 for two languages without any PALATALS (Ancient and Modern Greek).

The simplest PALATAL inventory among the IE languages is the one that just contains /j/ (P1), although this system is not found in the Slavic group. Likewise, its representation is too small in the non-IE language group to draw any conclusion (only Greenlandic belongs here, lacking PALATAL SPA). All seven non-Slavic IE languages with this system avoid Č_Č. Proto-Celtic is of this kind, thereby differing from the three non-reconstructed Celtic languages in

our sample (Welsh, Irish, and Breton) that have different systems and that lack palatal avoidance.

Another simple type is P2, with just /j/ and post-alveolar sibilants. It is by far the most frequent palatal system cross-linguistically. It is represented by 17 non-Slavic IE varieties, of which as many as 14 avoid Č_Č. This tendency conspicuously contrasts with the situation in the Slavic group, where type P2 is also the most common system but where only Bulgarian and Slovenian show a palatal SPA effect. The ratio for the non-IE languages is also suggestive of the tendency, but our sample is again too small to draw any conclusion.

Of the non-Slavic IE languages belonging in P2, it is only Armenian, Welsh, and Pashto for which we lack clear evidence for palatal avoidance. Since Armenian is the only representative of its branch, it is hard to judge to what extent the absence is an anomaly. In contrast, the Welsh case may be a feature characterizing the whole Celtic branch (see above). However, the absence of palatal avoidance in Pashto is very likely due to the limited data, which is also a reasonable explanation for the lack of dental avoidance in this language, too (see above). Moreover, the Č_Č sequences are underrepresented in the other three Iranian languages, all belonging in P2 as well (Ossetian, Kurmaji, and Persian).

Let us move to P3 (/j/, post-alveolar sibilants, and other palatal sonorants, usually the nasal /ɲ/). It is another very common system encountered among the non-Slavic languages (both IE and non-IE), but not in the Slavic group, where only Pomeranian belongs here, lacking palatal SPA. The ratios in Table 9 indicate that P3 languages tend *not to avoid* Č_Č, which appears to relate to the presence of palatal sonorants other than /j/. Recall that P2 languages, without these sonorants, instead tend *to avoid* palatal sequences. This conclusion is further supported by the non-IE group, where five out of seven P3 languages lack palatal avoidance and where three out of five P2 languages have this avoidance.

The non-Slavic IE languages belonging in P3 and not avoiding Č_Č are three Romance varieties: Italian, Catalan, and French (the last one significantly favors these sequences; see below). In contrast, another two Romance languages belonging in P3 have palatal avoidance, namely Spanish and Portuguese; Yule's Q for the former is relatively high, −0.43, while the latter attests only very weak avoidance (−0.15; so does Romanian, −0.17).[17] Finally, type P3 also includes Breton (no palatal avoidance; see above on the Celtic languages) and Hindi (avoidance). The Hindi case should perhaps be reconsidered be-

[17] The contents of the Spanish palatal system may be a matter of discussion. First, the phonemic status of [j] is uncertain because some analysts interpret it as a non-syllabic allophone of /i/ (Hualde 2005). Secondly, the obstruent /j/ has merged in Spanish dialects with the sonorant /ʎ/, which is reflected in the transcription of the lexical database that is the source of our data (i.e., Spanish is interpreted as a language lacking palatal plosives). This might have influenced the nature of palatal SPA.

cause the palatal nasal occurs here (a) in clusters before another palatal consonant, in which case it might be allophonic, and (b) in several loanwords, in which case the occurrence is marginal (Ohala 1999).

A fairly common system among the Slavic languages, though not so frequent elsewhere, is type P4, differing from P3 by containing true palatal plosives. All five Slavic languages of this kind lack palatal avoidance, which contrasts again with the other language groups. The Č_Č sequences are avoided in two out of three non-Slavic IE languages (Albanian and Latvian, but not Tocharian B) and in three out of four non-IE languages (Nancowry, Khasi, and Hungarian, but not Basque). This again supports our conclusion that the increased absence of palatal avoidance is a peculiarity of the Slavic family.

The last type instantiated by more than one IE language but absent in Slavic is P5, in which we find both true palatal plosives and fricatives. All three non-Slavic IE languages belonging here do not underrepresent Č_Č (Galician, Irish, and Plautdietsch). However, Turkish, as the only non-IE representative, avoids these sequences. There are two other languages that contain true palatal fricatives but lack post-alveolar sibilants, and both avoid palatal sequences, too (Icelandic and Nyah Kur). More languages are needed to evaluate the possibility that the nature of palatal avoidance is dependent on the presence of true palatal fricatives like /ç/. Note that Hall (1997) presents arguments for not regarding these fricatives as coronals (we have classified all palatal consonants as coronals).

There is not a lot to say about the remaining types. Many are attested only in the non-IE group, and those found among the two IE groups are largely represented by one language only. The only tentative generalization we can make here is that the Č_Č sequences are inclined to be avoided in languages without post-alveolar sibilants (types P10–13 including the only discussed type P1). To conclude our discussion of the non-Slavic IE languages, let us mention that there is one language that favors Č_Č, namely French. We have related the preference of palatal homorganic sequences to the presence of two palatal sibilant series (post-alveolars and alveolo-palatals) in Polish and Lower Sorbian. However, the same explanation does not hold for French, which has a fairly standard palatal system (post-alveolar sibilants, /j/, and /ɲ/). The non-Slavic languages, whether from the IE family or not, cannot be helpful in corroborating our hypothesis that the preference in Polish and Lower Sorbian is due to the presence of the two sibilant series. Four non-Slavic languages have two sibilant series within the palatals. Norwegian and Tatar avoid Č_Č but limit alveolo-palatals to fricatives, whereas Polish and Lower Sorbian also have alveolo-palatal affricates. Alveolo-palatal affricates and fricatives are present in Lezgian and Chukchi, but their palatal Q values are not significant.

### 4.2. Non-IE Languages

Our sample comprises 38 non-IE languages selected from several macrofamilies, most of which are represented by one variety only. An exception is the Austroasiatic family, with nine members from the Mon-Khmer group, which has been motivated by the availability of phonologically transcribed lexical data. Since P&S provided statistical information on the distribution of all CVC sequence types, we take advantage of their data for the 28 non-IE languages in their sample, using Yule's Q to gain a broader perspective.[18] Table 10 on the following page reproduces average Yule's Q values for both non-IE language samples, comparing them with the values for the Slavic and other IE languages repeated from the previous sections.

The non-IE languages confirm the trends already spotted in the other groups. Labials are most strongly and widely avoided, though there are languages without such avoidance. Dorsals show the second strongest avoidance effect; several languages are without such avoidance within each group. Dentals are also very widely avoided but less strongly. Finally, palatals are avoided by markedly fewer languages than the other places of articulation, and the non-IE languages are similar in avoidance strength to the Slavic languages, while the non-Slavic IE languages are not. There are also several other differences in detail to mention.

All IE languages have been found to have a labial avoidance effect. In contrast, we lack reliable evidence for this avoidance in one non-IE language in our sample (Otomi) and two in P&S's sample (Bullom and Sua). What is more, two non-IE languages in our sample significantly *prefer* P_P (Chinese and Mazahua). The weaker degree of labial avoidance seen in P&S's bundle is probably a trait of African languages, which make up most of the sample. In particular, Niger-Congo languages tend to have a weaker avoidance effect (including Xitsonga in our sample). On the other hand, the decreased negative Q value for labials in our non-IE sample can be related to the nine Mon-Khmer languages, of which eight have labial Q values below the average.

[18] Another re-evaluation using Yule's Q was already done by Grotberg (2022), though she did not provide actual Q values. From P&S's sample of non-IE languages, we have left out Basque because it is already part of our material, though the source of the data is different. The re-calculated statistic using Yule's Q is also available at https://ojs.ung.si/index.php/JSL/article/view/469/344. We should also note that there is some uncertainty about the statistics of P&S's languages. Having recalculated their numerical values from the paper's appendix, we have found out that there are probably some errors either in the values or in the tables reproduced in the paper proper. For example, in their Table 15, Mpongwe is marked as not significantly avoiding P_P, even though the O/E deviation value for this pattern is −25.8, which they otherwise regard as significant avoidance. Also, the sum of the occurrences of particular combinations for Mongolian and Kiga-Nkore does not equal the sum provided in the appendix.

**Table 10**. Average Yule's Q values for homorganic CVC sequences and the numbers of languages avoiding these sequences

| | Slavic (n = 18) | Non-Slavic IE (n = 44) | Non-IE ours (n = 38) | Non-IE P&S (n = 28) |
|---|---|---|---|---|
| **P_P** | −0.51 (18) | −0.58 (44) | −0.49 (35) | −0.41 (26) |
| **T_T** | −0.30 (18) | −0.36 (43) | −0.26 (38) | −0.30 (25) |
| **Č_Č** | −0.27 (3) | −0.37 (22) [8] | −0.30 (21) | −0.28 (13) |
| **K_K** | −0.43 (14) | −0.43 (42) | −0.38 (35) | −0.45 (25) |

(See Table 8, p. 196, regarding the round and square brackets.)

The already-mentioned Chinese and Mazahua are exceptional within the whole language bundle. They are the only varieties (including P&S's sample) that prefer labial as well as dorsal homorganic sequences. Admittedly, the association effect is very weak, though still significant (0.06 and 0.14 for labials and dorsals in Chinese, 0.08 and 0.16 for labials and dorsals in Mazahua, respectively). It is yet to be determined to what extent this is due to the nature of our lexical data and/or the statistical measure used, but we do not rule out the possibility that there are really languages preferring two non-coronal sequences. The preference for labials has been previously reported for Yup'ik (Rousset 2004: 196), though it concerned sequences of bilabials only (we are unaware of any mention of languages preferring dorsals).

There is little to say about dentals (but see the next section). The strength and extent of dental avoidance in the non-IE languages are comparable to the situation in the other families. This avoidance is observed in all non-IE languages of our sample but is, curiously, absent in three of the P&S sample, namely Kiga-Nkore, Quechua, and Sue. There is only one language in our sample for which we lack evidence for dorsal avoidance, Otomi (see below). In P&S's sample there are three languages without this avoidance (Fula, Manjaku, and Sua); since all are from the Atlantic family, the absence may be an areal/genetic feature.

The avoidance of palatal homorganic sequences is again somewhat anomalous because many non-IE languages fail to avoid them, just as the IE languages do. In the previous subsection, when discussing the palatal system typology, we argued that the presence or absence of this avoidance might relate to the contents of the palatal inventory. However, since most types we introduced are generally represented by eight or fewer non-IE languages, we

need more comprehensive data to draw reasonable conclusions. Therefore, our discussion will be limited here to a couple of general observations.

To begin with, there is agreement among all four bundles of languages on the fact that Č_Č sequences are avoided by the smallest number of languages in contrast to the other places. The Slavic family clearly differs from the rest because the portion of the avoidant languages is the smallest here (only three languages avoid Č_Č, i.e., 16.67%). The portion is larger in both non-IE bundles (55.26% for our own sample, 46.43% for P&S's one), but by far the greatest number of the languages avoiding Č_Č is found in the non-Slavic IE group (61.36%, including the −1 languages that lack these sequences altogether).

Another language-specific feature appears to be the absence of PALATAL avoidance in all nine non-Atlantic Niger-Congo languages included in our and P&S's sample (the Atlantic languages proper do not have this feature). Three of the non-Atlantic languages even prefer Č_Č (Proto-Bantu, Bemba, and Kiga-Nkore). The preference is also observable in languages from other families, namely, Korean, Tagalog, Maliseet-Passamaquoddy, Cebuano, and Semai (all included in our own sample). Above we argued that the preference of Č_Č in Polish and Lower Sorbian may relate to the presence of alveolo-palatal sibilants, but this cannot explain the same kind of preference within the non-IE languages mentioned because only Korean has alveolo-palatal sibilants.

One final remark on the non-IE bundle: We find two languages here for which Yule's Q statistic provides no evidence for the SPA effect. One is Otomi (our sample), which significantly avoids DENTALS only; the other is Sua (P&S's sample), where no place is significantly underrepresented. The latter case is suspicious because the data P&S used (495 CVC sequences) are too small to warrant reliability. Our Otomi dataset is larger (2,242 sequences) but may be non-representative or selective. It is taken from Lexibank (List et al. 2022), which is itself derived from a word list included in the World Loanword Database (Haspelmath and Tadmor 2009; note, however, that only 11% of these words are loanwords).

In the IE bundle, most languages avoid three or four places, and only seven avoid two out of the four places. Thus, at most we could claim that there is no conclusive evidence for avoidance or non-avoidance in some IE languages. On the other hand, the possibility of there being a non-IE language without SPA remains open. Mayer et al. (2010) mention that such languages do exist, especially around the equator, without providing any examples. However, the size of the lexical samples upon which they based their observations were too small to be reliable (no more than 246 CVC sequences for any of 3,200 languages). Judging from the data kindly shared by Thomas Mayer (personal communication, December 2024), some examples are Pirahã, Amuesha, Yavitero, Kalina, and Sambe. We have evaluated one larger lexical sample for

Kalina (615 CVC sequences), and Yule's Q has indicated that this language has an SPA effect for all places but palatals.

## 5. Coronals

We have demonstrated that SPA is observable across various language groups, but two crucial reservations apply. The avoidance effect is smaller for dentals than for labials and dorsals. Second, palatals are avoided only in 53.9% of languages (including P&S's 28 ones, or 56% in our 100-language sample), primarily due to the Slavic languages lacking this avoidance. Moreover, homorganic sequences of palatals happen to be *favored* in 10 languages (7 in our sample). These unusual properties invite us to review how the coronal place classes are delimited and how SPA is formulated.[19] As explained in §2.3, we use the term coronal as a convenient cover label for dentals and palatals but assume that SPA applies to each of them. However, some studies hold that the SPA principle concerns the whole class of coronals (to be discussed in §5.1). Moreover, since the whole class of coronals happens to manifest a weaker avoidance effect, some other studies argue that coronal avoidance is sensitive to manner features and that the class should be divided into obstruents and sonorants for SPA to obtain (§5.2). Finally, the separation of dentals and palatals may be essentially correct, but the membership of some consonants in these classes should be reconsidered (§5.3). Our discussion will conclude with remarks on the so-called labial-coronal effect and the CVC sequences not directly affected by the SPA principle.

### 5.1. Coronal Avoidance

In line with P&S and others, we assume that SPA applies to four basic place classes: labials, dentals, palatals, and dorsals. However, some researchers assume fewer classes, namely those that correspond to the three major articulators: labials, coronals, and dorsals. The most straightforward reduction is the one that merges dentals and palatals into coronals.[20] Such a union

[19] It is also possible that the difference between dentals and palatals on the one hand and labials and dorsals on the other is a consequence of the statistical measure used. However, we have specifically chosen Yule's Q statistic because Grotberg (2022) has convincingly demonstrated that other statistical tools previously used by researchers, like O/E or $\chi^2$, are biased. They consistently underestimate the avoidance of coronals (dentals) because they are sensitive to the high frequency these consonants have in most languages.

[20] Remember that small capitals indicate that coronals is a label for the union of the classes of dentals and palatals. Other usages of the term *coronal* are discussed in the following subsections. They are written in lowercase letters. The union of dentals and

**Table 11**. Average Yule's Q values for homorganic CORONAL CVC sequences and the numbers of languages avoiding these sequences

| | Slavic (n = 18) | Non-Slavic IE (n = 44) | Non-IE ours (n = 38) | Non-IE P&S (n = 28) |
|---|---|---|---|---|
| **Cor_Cor** | −0.29 (18) | −0.39 (43) | −0.35 (36) | −0.46 (27) |
| **T_T** | −0.30 (18) | −0.36 (43) | −0.26 (38) | −0.30 (25) |
| **Č_Č** | −0.27 (3) | −0.37 (22) [8] | −0.30 (21) | −0.28 (13) |

(See Table 8 on the round and square brackets.)

was explicitly argued for by Mayer et al. (2010) for various languages on the grounds that these consonants behave similarly in the SPA statistic (see also Teil-Dautrey 2008 for Proto-Bantu or Grotberg 2022 for Latin and Medieval Castilian and a re-analysis of P&S's data using the single class of CORONALS).

Applying this merger to our sample, we discover that all but two languages avoid homorganic sequences of CORONALS (see Table 11 above). The exceptions are Pashto and Chinese; another is Kamilaroi in P&S's data. The avoidance of CORONALS shows a stronger avoidance effect than the avoidance of DENTALS in both non-IE language bundles; the effect also slightly weakens in the non-Slavic IE languages. In the Slavic languages, the avoidance effect is virtually the same, though it drops a little (average Yule's Q = −0.29 for CORONALS and −0.30 for DENTALS).

Although the general increase in the avoidance strength speaks for the union of DENTALS and PALATALS, there is one serious problem with this union: If homorganic sequences of CORONALS are subject to SPA, we expect that *any* combination of two CORONALS is also avoided, that is, SPA should be observable for repetitions of DENTALS and PALATALS as well as combinations of DENTALS with PALATALS in both orders (T_Č and Č_T). This question is part of a broader problem concerning the distribution of CORONALS and NON-CORONALS, namely the effect called "extended SPA".

P&S noticed that their sample languages showed a tendency to avoid not only CVC combinations of the four primary place classes but also of two superclasses, that is, the combinations T_Č and Č_T involving different CORONALS and P_K and K_P involving different NON-CORONALS. To the best of our knowledge, this extended SPA effect has not been reported before P&S or afterward. However, the underrepresentation of said combinations is noticeable in some other studies (e.g., Mayer et al. 2010), though not explicitly named.

palatals is not the only possible one; palatals could also be interpreted as dorsals. Due to the lack of space, we will not consider this option, but see §5.4 and fn 2.

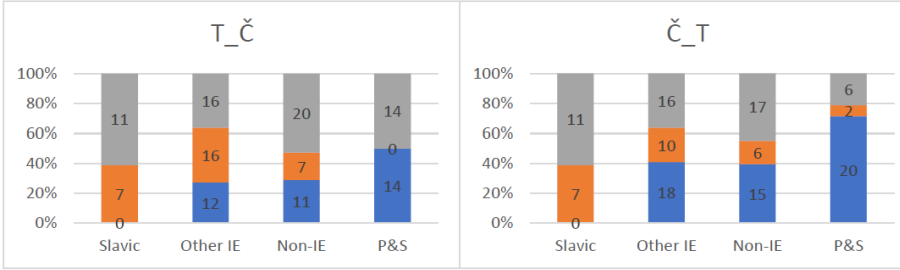


**Figure 1**. Graphs showing the proportion of the languages avoiding and not avoiding combinations T_Č and Č_T in our and P&S's language sample (Blue = avoidant languages, orange = favoring languages, gray = no significant inclination in either direction or the absence of such combinations. The statistics include –1 languages.)

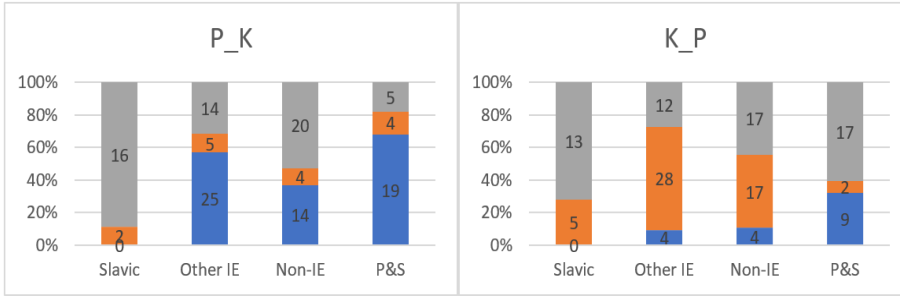


**Figure 2**. The same graphs for P_K and K_P (see Figure 1)

The attitude of our own languages to these combinations is visualized in Figure 1 (above), also including a re-analysis of the 28 non-IE languages from P&S. The graphs offer a comparison of the proportion of the varieties that significantly disfavor (blue) or significantly favor (orange) the combinations T_Č and Č_T, or in which there is no significant disproportion in either direction (gray; including two cases of the languages without PALATALS, Ancient and Modern Greek). Figure 2 gives the same for P_K and K_P. For simplicity, we do not report average Yule's Q values, though it is worth noting that all the combinations tend to be avoided with slightly more strength than they are preferred—with the exception of Slavic (e.g., T_Č has the average value of −0.31 for the avoidant non-IE languages in our sample and 0.17 for the favoring languages of the same group; but it is −0.15 against 0.22 for P_K in the same group).

It is immediately clear that no Slavic language avoids T_Č and Č_T; instead, a large portion of these languages *favors* them. In the case of T_Č, this includes all East Slavic languages together with Bulgarian, Serbo-Croatian, and Lower Sorbian; in the case of Č_T, the favoring languages are all South Slavic varieties together with Old Russian, Modern Czech, Slovak, and Proto-Slavic. The preference is apparently connected with their shared historical development. And it also seems to be an areal or genetic trait. Considering the situation in the other groups, we witness a noticeable inclination to avoid rather than prefer combinations of two different CORONALS. The proportion of the languages disfavoring T_Č and Č_T is largest in P&S's sample, of which only two languages happen to favor these sequences (Joola Kwaatay and Manjaku). Most of the avoidant languages are from Africa. Note also that the sequence Č_T tends to be avoided by more languages than its reverse, which has already been pointed out by P&S.

These observations go against the predictions of CORONAL SPA that unites DENTALS and PALATALS into one class. The proportion of the languages that avoid combinations of two different CORONALS is too small to warrant the conclusion that *any* combination of coronals is statistically underrepresented. The fact that CORONAL avoidance has been so widely reported must be due to the properties of DENTALS rather than CORONALS as a whole class. CVC sequences of DENTALS are almost uniformly avoided, but combinations of DENTALS with PALATALS happen to be avoided in a considerably smaller portion of languages. Many languages disprefer combinations of two PALATALS, but far from all. Moreover, some languages, such as the Slavic ones, tend to favor the mixed combinations of CORONALS.

Before moving to other topics, let us briefly look at Figure 2, which compares the proportions of the avoidant and favoring languages for the other combination types falling under the extended SPA principle, those involving two different NON-CORONALS. The Slavic languages fail to provide evidence for this type of avoidance. Once again, no Slavic variety disfavors any of these combinations. Instead, P_K is favored by Slovak and Old Czech but not by Modern Czech (which is curious). K_P is favored by all South Slavic languages but Bulgarian, and by Upper Sorbian. And once again it is P&S's non-IE languages in which these two sequences are most widely avoided, especially P_K. However, comparing the proportion of avoidant and favoring languages, we can hardly claim that the non-Slavic IE languages and the non-IE languages in our own sample tend to underrepresent K_P. There are more languages favoring these sequences than languages avoiding them. Therefore, only P_K sequences are demonstrably avoided across language families, except Slavic. We do not have any explanation for this discrepancy.

In short, our data do not support the existence of the extended SPA principle in Slavic, and in general, there are more languages without this effect

than those that impose restrictions on the combinability of CORONALS with NON-CORONALS.

## 5.2. Place-*cum*-Sonorancy Avoidance

Although many linguists have demonstrated the existence of coronal avoidance in several languages, we have repeatedly pointed out that the reader should be cautious about the meaning of the term *coronal*. It is often used as an equivalent to DENTALS either because (alveolo-)palatals or post-alveolars are not classified as DENTALS (see the following subsection) or because languages have very few (alveolo-)palatals and post-alveolars (e.g., Arabic, McCarthy 1994) or none (e.g., Maori, Rácz et al. 2016). Still, even these languages are characterized by a weaker coronal SPA effect, which is the same as saying that these languages have a weaker DENTAL SPA effect. And as demonstrated in the previous subsection, the effect is weaker even if the class of coronals includes alveolo-palatals, palatals, and post-alveolars. In short, no matter how coronals are defined, their avoidance tends not to be as strong as in other places.

It has been argued that the weakness or even the lack of a coronal SPA effect is a consequence of the large number of coronals present in phoneme inventories compared to the other place classes (Pierrehumbert 1993; Frisch et al. 2004). More distinctive features are therefore needed to capture differences between coronals. If similarity is measured by distinctive features, coronals are individually more dissimilar. And since SPA is essentially the avoidance of *similar* segments, additional, non-place features are expected to influence the coronal SPA effect.

Researchers have indeed reported that non-manner features influence SPA not only in the case of coronals but also other places. Padgett (1992) called them *subsidiary features*, claiming that they are, in a feature-geometric sense, dependent on and attached to the place node, which are the features [sonorant] and [continuant] in his conception. Yet phonological properties other than place have been shown to play a role in SPA (see, e.g., Berkley 1994, Dmitrieva 2008, and van Goch 2010, Rose and King 2007 on voicing, vowel type, stress, and distance). So, perhaps subsidiary features should simply be understood as SPA-influencing and -increasing properties (Wilson and Obdeyn 2009). The weight of such properties in SPA restrictions may be language-specific, though (see Coetzee and Pater 2008).

Padgett (1992) only formalized what had been known about the coronal SPA effect. Initially, the division of coronals by the feature [sonorant] was put forth for the first language in which SPA was described, Classical Arabic (McCarthy 1986, 1994).[21] This language restricts the occurrence of coronals

[21] As already mentioned, McCarthy (1994) and others (e.g., Padgett 1992 for Russian; see also the other references in this subsection) further demonstrate how the feature

within consonantal roots (already observed by Cantineau 1946 and Greenberg 1950). In the first two consonant positions, coronal obstruents freely combine with coronal sonorants, whereas repetitions of coronal obstruents and coronal sonorants are prohibited. The same division of coronals was confirmed in SPA studies of other Semitic languages (Akkadian by Reiner 1966; Tigrinya by Buckley 1997; Hebrew by Bachra 2001; Chaha and Amharic by Rose and King 2007) and other languages like Qafar (Hayward and Hayward 1989), Imdlawn Tashlhiyt Berber (Elmedlaoui 1995), Yamato Japanese (Kawahara et al. 2006), Shona and Wargamay (both in Wilson and Obdeyn 2009), Ngbaka (Danis 2017), Russian (Padgett 1992, 1995), English (Berkley 1994, 2000), and Dutch (van Goch 2010).

Our 100-language sample supports the observation that the avoidance of CORONALS depends on sonorancy, which we view as a difference between obstruents and sonorants. Eighty-six languages avoid repetitions of CORONAL obstruents (including all Slavic languages), and 92 avoid repetitions of CORONAL sonorants (including all Slavic languages except for Upper Sorbian). In contrast, combinations of CORONAL obstruents with CORONAL sonorants are avoided only in 29 languages (of which none is Slavic), and combinations of CORONAL sonorants with CORONAL obstruents are avoided in 15 languages (none is Indo-European!). These heterorganic CVC sequences are instead favored by 36 languages (obstruent + sonorant) and 66 languages (sonorant + obstruent). In both cases, this is true for virtually all Slavic languages (one positive Q value is not statistically significant).[22]

However, once again, we should be mindful of what these observations tell us. In the previous paragraph, we deliberately wrote "CORONALS" rather than "coronals" to indicate that we refer to a consonant class that may or may not include PALATALS (depending on whether a language has them), not just DENTALS. Thus, if combinations of CORONAL obstruents are indeed avoided, we expect that the avoidance concerns not just repetitions of DENTAL and PALATAL obstruents, and of DENTAL and PALATAL sonorants (on which see Table 13 on p. 213) but also combinations of the obstruents and sonorants without agreement in place. Our data provide weak support for these predictions.

Table 12 on the following page quantifies the extent and degree to which our languages avoid combinations of the CORONALS not agreeing in place

[continuant] affects SPA. We will not consider this problem here for lack of space. Suffice it to say that the influence this feature might have among coronals largely coincides with what we will call sibilancy avoidance in the following subsection.

[22] The sonorancy SPA effect for the 28 non-IE languages of P&S cannot be ascertained because we do not have information on the distribution of obstruents and sonorants. The authors only mention that "[their] measurements show that the dominant effect concerns PLACE OF ARTICULATION [emphasis original], not manner, nasality, or state of the glottis" (p. 342).

**Table 12.** Average Yule's Q values for CVC combinations of dentals and palatals with or without sonorancy agreement for a subsample of 86 languages

| | | Dental | | Palatal | |
|---|---|---|---|---|---|
| | | **Obstruents** | **Sonorants** | **Obstruents** | **Sonorants** |
| **Dental** | **Obstruents** | | | −0.35 (43) [2] | −0.29 (17) [1] |
| | **Sonorants** | | | − 0.45(3)[4] | −0.39 (16) [1] |
| **Palatal** | **Obstruents** | −0.31 (41) [4] | −0.15 (12) [3] | | |
| | **Sonorants** | −0.34 (18) [1] | −0.37 (31) [0] | | |

(Shading marks sonorancy agreement. See Table 8, p. 196, on the round and square brackets and Table 13, p. 213, for the empty cells.)

and/or sonorancy. The calculations are limited to a subsample of 86 languages that possess at least one palatal obstruent and one palatal sonorant (each language has at least one dental obstruent and sonorant). All Slavic varieties are included. The average Yule's Q is again calculated only for the languages that significantly avoid a given combination.

The table reveals several things. To begin with, the sonorant-first combinations are avoided more strongly than the combinations beginning with an obstruent. The sequences agreeing in manner/sonorancy are more widely disfavored than those without the agreement (see the shaded cells). The obstruent combinations are avoided in approximately 50% of the languages, whereas the sonorant combinations are in less than 35%. Finally, the sequences without agreement in place and manner are avoided in the smallest number of languages, never making a portion larger than 20% of the sample. In short, Table 12 supports our conclusion arrived at in the previous subsection: CVC combinations of dentals with palatals tend not to be avoided.

Even though there is little evidence for sonorancy having a substantial role in coronal avoidance, we must still account for the widely reported fact that nearly all the languages analyzed do show a place-*cum*-sonorancy avoidance effect for coronals (mind the typographical difference!). We offer an answer to this question in the next subsection, but first we would like to address another problem: Will SPA account better for the distribution of non-coronal consonants when they are divided into subclasses by sonorancy?

Berkley (2000: 29) paid attention to this question in English, showing that all mutual combinations of labial obstruents and sonorants, as well as all combinations of velar obstruents and sonorants, are almost equally

underrepresented irrespective of sonorancy agreement. Van Goch (2010: 79–82) reached the same conclusion for Dutch, and Padgett (1992, 1995) reported the same for labials in Russian. In contrast, Frisch et al. (2004: 192–93) state that there is some evidence for the subclassification of labials by manner in Classical Arabic, but the effect is not as strong as for coronals. Similarly, Coetzee and Pater (2008) indicate that sonorancy agreement influences, to some degree, the strength of SPA in Muna for all places.

Our material does not prove that SPA would be considerably influenced by sonorancy agreement. Table 13 on the following page compares the strength and extent of SPA as a function of sonorancy agreement for the four basic places. The values for LABIALS and DENTALS are derived from the whole 100-language sample. Those for PALATALS and DORSALS rely on fewer languages (86 and 33, respectively) because not all languages have obstruents and, especially, sonorants for these places.[23]

There is a striking difference between the DENTALS and the remaining places. While almost the same number of varieties avoid all combination types with very similar strength in the case of LABIALS and PALATALS, the avoidance of DENTALS is more common and stronger when there is sonorancy agreement (contrast the shaded and unshaded cells). This fact immediately explains why many previous studies reported sonorancy-dependent avoidance of coronals: As pointed out above, coronals are usually equivalent to DENTALS either because the class of coronals excludes (alveolo-)palatals or post-alveolars or because the languages analyzed have a very small inventory of (alveolo-) palatals and post-alveolars.[24]

The way sonorancy influences the distribution of DORSALS is less clear because our observations can be based on just 33 languages. The other languages either lack such segments or disallow CVC combinations of a DORSAL sonorant with any other consonant. Yet the statistics of Table 13 do not indicate that DORSAL SPA would be influenced by sonorancy agreement. It is only the presence of a DORSAL sonorant (especially as the $C_1$ consonant) that lowers the number of languages in which sequences of two DORSALS are avoided. On the other hand, sequences of two DORSAL obstruents are avoided in the majority of these 33 languages. The combinations with at least one DORSAL sonorant

[23] Since some of our source databases use phonetic transcription, some segments, like the velar nasal [ŋ], may be allophonic variants rather than separate phonemes. Note also that the DORSAL subsample includes only those languages that have DORSAL sonorants that combine with other consonants in the first word syllable. This is to say that our 100-language sample contains other languages with DORSAL obstruents and sonorants, but in many the sonorants have very restricted combinability.

[24] PALATALS once again stand out because their mutual combinations are absent in many languages (see the numbers in square brackets), even though the languages have both PALATAL obstruents and sonorants.

**Table 13**. Average Yule's Q values for homorganic CVC sequences with or without sonorancy agreement

| | | Labial | | | | Dental | |
|---|---|---|---|---|---|---|---|
| | | **Obst.** | **Son.** | | | **Obst.** | **Son.** |
| **Labial** | **Obst.** | −0.47 (76) [1] | −0.51 (80) [2] | **Dental** | **Obst.** | −0.27 (79) | −0.15 (28) |
| | **Son.** | −0.64 (75) [9] | −0.55 (76) [3] | | **Son.** | −0.22 (11) | −0.51 (93) |
| | | **Palatal** | | | | **Dorsal** | |
| | | **Obst.** | **Son.** | | | **Obst.** | **Son.** |
| **Palatal** | **Obst.** | −0.41 (34) [7] | −0.41 (25) [13] | **Dorsal** | **Obst.** | −0.37 (30) | −0.35 (20) |
| | **Son.** | −0.49 (16) [23] | −0.55 (23) [19] | | **Son.** | −0.43 (1) [7] | −0.66 (7) [14] |

(The values for labials and dentals are based on the total sample of 100 languages; the values for palatals on a subsample of 86 languages (the same as in Table 12); the values for dorsals on a subsample of 33 languages. Shading marks sonorancy agreement. See Table 8 on the round and square brackets.)

are furthermore significantly preferred in 12 languages, of which six possess a velar (or uvular) rhotic. This includes Upper Sorbian, the only Slavic variety in this subsample. None of the languages that limit dorsal sonorants to nasals prefer combinations containing a dorsal sonorant. Hence, the preference is at least partially explainable by the presence of dorsal trills.

### 5.3. Sibilancy Avoidance

Another explanation for the unusual distribution of homorganic sequences of dentals and palatals could reside in the contents of these identity classes. Recall that we define dentals as a class consisting of (inter)dental, (pre-)alveolar, and retroflex consonants, whereas the class of palatals comprises post-alveolars, alveolo-palatals, and true palatals. Although we call them collectively coronals, not all consonant classes mentioned have been characterized as coronals (e.g., by sharing the articulation of the tongue blade or by the feature [+coronal]). Since alveolo-palatals of the kind found in Polish (e.g., /ɕ/) are poorly represented in our language sample, we will not consider how their reclassification impacts the SPA effect. Likewise, we will not go into the problem

of palatal fricatives such as /ç/ and /ʝ/ in German and elsewhere, which Hall (1997: 15ff.) interpreted as non-coronals. They, too, are poorly represented in our data. Instead, we can look into another question suggested by one of the reviewers, namely the status of post-alveolars (palato-alveolars), which always happen to be sibilants.

First of all, it turns out that removing post-alveolars from the class of PALATALS does not improve the SPA statistic for this identity class. There are 14 languages in our sample for which this can be reliably tested. They limit their PALATAL obstruent inventory to true palatal plosives and post-alveolar sibilants.[25] One half of these languages avoid the CVC sequences that contain either true palatal plosives or post-alveolar sibilants; no language favors these sequences. Sequences just containing true palatal plosives are avoided in five languages. To be more precise, five languages lack such combinations, which may be a lacuna in our lexical data rather than total avoidance. Three languages significantly favor repetitions of palatal plosives (Modern Czech, Latvian, and Hungarian). In short, if the class of PALATALS includes post-alveolar sibilants, more languages avoid PALATAL CVC sequences than when the sibilants are not part of the class.

We are better positioned to assess how the avoidance effect is changed when post-alveolar sibilants are classified as DENTALS. Our sample contains 74 languages with post-alveolar sibilants. When these segments are treated as DENTALS, the repetitions of DENTAL obstruents happen to be underrepresented in more languages than when post-alveolars are members of the class (63 instead of 60). Furthermore, the avoidance strength measured by averaged Yule's Q increases from −0.26 to −0.28. Though slight, the increase can be sufficient proof that post-alveolar sibilants pattern with dento-alveolars rather than palatals. However, the increase is actually a manifestation of another type of avoidance not mentioned in the literature: the avoidance of repetitions of sibilants in CVC sequences.

To demonstrate this restriction, let us consider a smaller selection of the 47 languages that have both (pre-)alveolar and post-alveolar sibilants and no other coronal affricates and fricatives to affect the statistics; all these languages have dental/alveolar plosives too. Thus, the 47-language subsample excludes English, which contains the non-sibilant fricatives /θ/ and /ð/, or Polish, in which we find alveolo-palatal sibilants. Table 14 on the following page re-

[25] Four other languages have true palatal plosives and post-alveolar sibilants, but they also have true palatal fricatives. Since some linguists consider these fricatives to be non-coronals (Hall 1997), we exclude these four languages from the statistics. Also, we do not consider how palatal obstruents, however they are delimited, combine with palatal sonorants, as sonorancy agreement may play a role. To put it simply, the 14 languages we describe here are those belonging in types P4, P8, and P14 (see Table 9, p. 199).

**Table 14**. Average Yule's Q for combinations of dento-alveolar plosives with dento-alveolar sibilants and post-alveolar sibilants

| | **Dento-alv. plosives** | **Dento-alv. sibilants** | **Post-alv. sibilants** |
|---|---|---|---|
| **Dento-alv. plosives** | −0.28 (32) | −0.30 (22) [2] | −0.24 (13) [2] |
| **Dento-alv. sibilants** | −0.28 (21) [1] | −0.41 (36) [3] | −0.47 (20) [5] |
| **Post-alv. sibilants** | −0.27 (16) [1] | −0.49 (21) [6] | −0.42 (19) [5] |

(The values are calculated for 47 languages. Shading marks sibilancy avoidance. See Table 8 on p. 196 on the round and square brackets.)

produces average Yule's Q values for all mutual CVC combinations of dento-alveolar plosives, dento-alveolar sibilants, and post-alveolar sibilants.

The table demonstrates two things. First, the combinations involving no more than one sibilant (the unshaded cells) are underrepresented in fewer languages and less strongly than the combinations involving two sibilants (the shaded cells). Two-sibilant combinations are even absent in many languages (the number of languages is given in the square brackets). Second, the combinations of plosives with sibilants (affricates and/or fricatives) are disfavored in more languages and in a stronger manner when they agree in the place of articulation than when there is no such agreement, that is, when both are dento-alveolars rather than a combination of a dento-alveolar and a post-alveolar. To put it otherwise, if we set aside combinations of sibilants, the avoidance of homorganic sequences of DENTAL obstruents is indeed less strong than the avoidance of LABIALS and DORSALS (refer to Table 10, p. 203). The inclusion of post-alveolars in the class of DENTALS will increase the DENTAL avoidance strength, but only because many languages disfavor repetitions of sibilants, obviously due to their acoustic and perceptual similarity (cf. Kokkelmans 2021). We find it more appropriate to assume the existence of sibilancy avoidance as a special kind of similarity avoidance rather than to unite post-alveolar sibilants with dento-alveolars instead of palatals.

There are two other, interrelated pieces of evidence for treating post-alveolars as belonging to the distributional class of PALATALS, at least in Slavic languages. First, almost all Slavic languages significantly and strongly avoid CV sequences of PALATALS and back vowels; the only exception is Polabian, in which Yule's Q for such sequences is not significant. This avoidance is observable both for combinations with true palatals (= plosives and nasals in

**Table 15**. Yule's Q for CV combinations of palatals with back vowels in Slavic languages

| | Bel | Bul | LSo | Mac | MCz | MRu | OCS | OCz | ORu |
|---|---|---|---|---|---|---|---|---|---|
| **Plo** | × | × | × | absent | −0.87 | × | absent | 1.00 | × |
| **Aff** | −0.58 | −0.51 | −0.73 | −0.59 | −0.70 | −0.74 | −0.82 | absent | absent |
| **Fri** | (−0.17) | −0.53 | −0.49 | −0.54 | −0.57 | −0.58 | −0.63 | −0.72 | −0.46 |
| **Nas** | × | × | × | absent | −0.77 | × | absent | 1.00 | × |
| **Son** | −0.52 | (0.02) | −0.42 | −0.49 | −0.93 | −0.53 | −0.35 | −0.67 | −0.60 |
| | **Plb** | **Pol** | **Pom** | **PSl** | **SCr** | **Slk** | **Sln** | **Ukr** | **USo** |
| **Plo** | × | × | × | × | × | absent | × | × | × |
| **Aff** | × | absent | (−0.14) | −0.88 | −0.49 | −0.47 | −0.82 | −0.57 | −0.80 |
| **Fri** | (0.23) | −0.57 | −0.61 | −0.75 | −0.59 | −0.55 | −0.51 | −0.31 | −0.64 |
| **Nas** | × | absent | −0.86 | absent | (−0.60) | absent | × | × | × |
| **Son** | (0.24) | −0.79 | −0.61 | 0.71 | −0.29 | (−0.15) | (0.06) | −0.49 | −0.64 |

(Parenthesized values are not significant. Based on the same data as Table 4. "Absent" means a language lacks a certain combination of palatals even though it has these consonants.)

**Table 16**. Average Yule's Q values for CVC sequences unaffected by pure and extended SPA

| | Slavic (n = 18) | Non-Slavic IE (n = 44) | Non-IE ours (n = 38) | Non-IE P&S P&S (n = 28) |
|---|---|---|---|---|
| **P_T** | 0.29 (17) | 0.35 (44) | 0.24 (32) | 0.27 (21) |
| **T_P** | 0.35 (18) | 0.32 (43) | 0.25 (31) | 0.26 (25) |
| **T_K** | 0.19 (12) | 0.29 (41) | 0.22 (34) | 0.29 (24) |
| **K_T** | 0.19 (6) | 0.20 (27) | 0.23 (28) | 0.30 (22) |
| **P_Č** | × (0) | 0.17 (18) | 0.23 (17) | 0.28 (19) |
| **Č_P** | × (0) | 0.22 (17) | 0.22 (15) | 0.23 (18) |
| **Č_K** | × (0) | 0.32 (17) | 0.28 (15) | 0.26 (16) |
| **K_Č** | × (0) | 0.16 (5) | 0.24 (9) | 0.17 (11) |

(The parenthesized numbers indicate how many languages with significant *positive* Q values have been used for the calculation.)

Table 15 on the previous page) and post-alveolar sibilants (= affricates and fricatives in Table 15).[26]

The second piece of evidence comes from Modern Czech. This language (and probably other Slavic languages, too) not only disfavors combinations of palatals with back vowels but severely limits their distribution, which concerns both the true palatals /c/, /ɟ/, /ɲ/, and /j/ as well as the post-alveolars /t͡ʃ/, /d͡ʒ/, /ʃ/, /ʒ/, and /r̝/ (ř). Both types of combinations are mostly found across morpheme boundaries, in loanwords, or expressive and onomatopoeic vocabulary (Mathesius 1931). In other words, these CV combinations are virtually absent within native-origin morphemes; the few exceptions there are could be etymologically explained as being old and obscured instances of the three contexts mentioned (Bičan 2018 and Bičan, forthcoming).

## 5.4. Labial-Coronal Effect

The four basic place classes of labials, dentals, palatals, and dorsals produce 16 possible CVC combinations. The SPA principle, as formulated by P&S, divides them into three groups. One consists of homorganic combinations directly affected by pure SPA, the main research focus of this paper. The second group consists of coronal combinations of dentals with palatals and of non-coronal combinations of labials with dorsals. We saw in §5.1 that some languages disfavored these sequences, which phenomenon P&S had called extended SPA. The final group includes all the remaining combinations. It follows from the nature of SPA that if some sequences are disfavored/underrepresented, others must be favored/overrepresented. It is the third group of combinations that strongly inclines to this preference.

Table 16, shown on the previous page, provides average *positive* Yule's Q values for the eight kinds of combinations of non-coronals and coronals, which belong to the third group and which P&S found to be preferred cross-linguistically. This preference is confirmed by our sample languages and the recalculation of P&S's data, though again less obviously in Slavic. The eight sequence types break down into three bundles, which division is most obvious among the Slavic languages.

The first consists of combinations of labials with dentals (P_T and T_P). The preference for these sequences is cross-linguistically the strongest and most diffused one. Of the IE languages, it is only Pomeranian and Pashto for which this avoidance is not statistically significant, which may be due to the small size of our lexical data. The preference was already well spotted by P&S, but they did not explain it. It is a manifestation of a so-called labial-coronal

[26] The statistics are a little distorted for Polish, Pomeranian, and Serbo-Croatian because these languages also have alveolo-palatal affricates and fricatives, which have been included in the classes Aff and Fri in the table.

or fronting effect. Numerous cross-linguistic studies have demonstrated that sequences of a (bi)labial and a coronal are preferred both in infant and adult languages (see overviews in MacNeilage and Davis 2000 or Carrissimo-Bertola 2010). Since Table 16 on p. 216 indicates that the effect pertains little to PALATALS, the tendency should be dubbed a LABIAL-DENTAL effect instead. In fact, PALATALS were not included in the coronal class in the studies mentioned.

On the other hand, we do not notice any straightforward tendency to favor or disfavor the remaining combinations. The second bundle of sequences includes combinations of DENTALS with DORSALS (T_K, K_T). They are favored by fewer languages and less strongly, especially among the IE languages. In the case of K_T, the drop in favor may be connected to there being relatively many languages that significantly disfavor these sequences. Since none of these 12 languages are from Slavic, the explanation actually fails. Neither can we explain the attitude toward the T_K sequences (they are disfavored by three languages, again none being Slavic). The last bundle of sequences consists of those involving a PALATAL combined with a NON-CORONAL. None of these combinations are favored in Slavic languages, and the non-Slavic languages also tend to favor them to a lesser degree. Instead, these sequences happen to be disfavored in not a small number of languages. It may partly be diachronic hold-over, especially in the case of Č_K and K_Č in Slavic. It is well known that Slavic palatalizations produced postalveolar sibilants from velars; a similar mutation of velars into palatals is attested elsewhere (Bateman 2007). In other words, many palatal segments were originally velars, so the avoidance of Č_K and K_Č might have originally been the avoidance of K_K.[27]

## 6. Conclusion

Similar Place Avoidance (SPA) or its close variant, Obligatory Contour Principle (OCP or OCP-Place), is, besides the Sonority Sequencing Principle, perhaps the best-known and studied phonotactic constraint shared by many natural languages. Yet the studies do not agree on the formulation of the SPA principle, which slightly diminishes its power and casts doubt on its universality. We calculated the SPA statistics for the lexicons of 100 language varieties, of which 18 are Slavic, 48 non-Slavic Indo-European, and 38 non-IE languages. These languages avoid CVC combinations of consonants at the same place of articulation, but the situation is more complex than the existing SPA models suggest.

SPA models set up several consonant classes (identity classes), assuming that the CVC combinations of same-class consonants are significantly

[27] It is then not surprising that Padgett (1992) and Pozdniakov (2007, 2010) found that combinations of DORSALS with /t͡ʃ/, /ʃ/, and /ʒ/ are underrepresented in Russian. Our data are less suggestive of this kind of avoidance.

underrepresented, whereas different-class combinations are not. The difference between the models concerns whether coronal consonants are divided or grouped by place features only or whether non-place features such as sonorancy interfere. Repetitions of labials and dorsals are, with very few exceptions, avoided in all languages. The subdivision of labials by sonorancy does not improve the model because all four logically possible combinations of labial obstruents and sonorants are avoided in approximately the same number of languages. The same could be said about dorsals, though the rarity of dorsal sonorants distorts the picture. However, subdivision by sonorancy reveals that repetitions of sonorants agreeing in place are widely and strongly avoided cross-linguistically, which is true not only for dorsals but also for dentals and palatals, and as already mentioned, for labials too.

Table 17 on the following page visualizes the three most common models used in typological studies of SPA or recurring in the literature on SPA or OCP (model A in P&S; B in Mayer et al. 2010; C in several studies on Semitic languages). They disagree on the number and definition of the consonant classes for which consonant co-occurrence restrictions hold. Analyses of singular languages may work with more classes (e.g., velars and gutturals separated in Arabic) or fewer classes (e.g., Greek lacks palatals). The first, non-parenthesized values are averaged Yule's Q coefficients, measuring the association strength between two consonants in a CVC sequence or the degree to which the sequences are avoided. The second, parenthesized values are percentages of the languages from three groups (Slavic, non-Slavic IE, and non-IE) that significantly avoid a given sequence.

None of the three models adequately describes the distribution of coronal CVC sequences in Slavic and other languages (unless model C is equivalent to model D due to the lack of palatals; see below). Model B unites dentals and palatals into one class of coronals but leads to wrong predictions about the distribution of the united class. It is apparent from model A that palatals are markedly less affected by similarity avoidance than dentals. Homorganic sequences of palatals even tend to be favored in some languages (including Polish and Lower Sorbian), which is not evidenced for dentals. Furthermore, CVC combinations of coronals in which one consonant is dental and the second palatal (in either order) are not avoided in any Slavic language, and the same is true for around 60% of the other languages. Model B is thus not superior to model A.

The third model, C, the most common alternative to A and the original form of the OCP-Place principle, divides coronals into obstruents and sonorants. Repetitions of coronal obstruents and coronal sonorants are underrepresented in all Slavic languages and most other languages (several non-Slavic IE and non-IE languages do not avoid sequences of coronal obstruents). At the same time, combinations of coronal obstruents with coronal sonorants tend not to be avoided—in fact, they are favored across all language groups.

Still, this classification must be rejected for two reasons. The first follows from the same but elaborated argument against model B: Combinations of CORONAL obstruents agreeing in sonorancy but not in place are avoided in no more than 50% of Slavic languages; even elsewhere the avoidance is not much diffused. Still, we should note that most languages disfavor combinations of DENTAL obstruents with PALATAL obstruents (in both orders). It seems to be a consequence of sibilancy avoidance, that is, the dispreference of CVC sequences in which the consonants are dental, (pre-)alveolar, post-alveolar, and possibly also alveolo-palatal sibilants (but on Polish, see Table 6 on p. 194). Though not reported previously, a strong sibilancy avoidance effect is observable in many languages (but less obviously in the Slavic group).

**Table 17**. Three rejected SPA models

| A | LABIALS | DENTALS | PALATALS | DORSALS |
|---|---|---|---|---|
| | S: −0.51 (100%) | S: −0.30 (100%) | S: −0.27 (16.67%) | S: −0.43 (77.78%) |
| | I: −0.58 (100%) | I: −0.36 (97.73%) | I: −0.37 (50.00%*) | I: −0.43 (95.45%) |
| | N: −0.49 (92.10%) | N: −0.26 (100%) | N: −0.30 (55.26%) | N: −0.38 (92.11%) |

| B | LABIALS | CORONALS | DORSALS |
|---|---|---|---|
| | S: −0.51 (100%) | S: −0.29 (100%) | S: −0.43 (77.78%) |
| | I: −0.58 (100%) | I: −0.39 (97.73%) | I: −0.43 (95.45%) |
| | N: −0.49 (92.10%) | N: −0.35 (97.37%) | N: −0.38 (92.11%) |

| C | LABIALS | CORONAL OBSTRUENTS | DORSALS |
|---|---|---|---|
| | S: −0.51 (100%) | S: −0.31 (100%) | S: −0.43 (77.78%) |
| | I: −0.58 (100%) | I: −0.29 (90.91%) | I: −0.43 (95.45%) |
| | N: −0.49 (92.10%) | N: −0.23 (73.68%) | N: −0.38 (92.11%) |
| | | CORONAL SONORANTS | |
| | | S: −0.56 (94.44%) | |
| | | I: −0.51 (100%) | |
| | | N: −0.35 (81.58%) | |

(The first numerical values are the average Yule's Q values for 18 Slavic ("S"), 44 non-Slavic IE ("I"), and 38 non-IE ("N") languages. Parenthesized values are percentages of the languages significantly avoiding a given homorganic sequence. The asterisked percentage includes those languages that lack PALATAL sequences altogether.)

**Table 18**. The proposed SPA model

| D | labials | dental obst. | palatals | dorsals |
|---|---|---|---|---|
| | S: −0.51 (100%) | S: −0.32 (94.44%) | S: −0.27 (16.67%) | S: −0.43 (77.78%) |
| | I: −0.58 (100%) | I: −0.26 (90.91%) | I: −0.37 (50.00%*) | I: −0.43 (95.45%) |
| | N: −0.49 (92.10%) | N: −0.26 (57.89%) | N: −0.30 (55.26%) | N: −0.38 (92.11%) |
| | | dental son. | | |
| | | S: −0.63 (94.44%) | | |
| | | I: −0.56 (100%) | | |
| | | N: −0.44 (84.21%) | | |

(The numerical values are average Yule's Q values for 18 Slavic ("S"), 44 non-Slavic IE ("I"), and 38 non-IE languages ("N"). Parenthesized values are percentages of the languages significantly avoiding a given homorganic sequence. The asterisked percentage includes those languages that lack palatal sequences altogether.)

The second reason for rejecting model C concerns the predominance of dentals within the class of coronals. The observation that the sonorancy agreement determines the strength and diffusion of SPA is valid, but the agreement affects only dentals, not palatals. Irrespective of sonorancy agreement, combinations of palatals are avoided by almost the same number of languages (again, less obviously in Slavic, but these languages do not restrict homorganic CVC sequences of palatals in general). However, since most coronals are, in fact, dentals (including alveolars), the sonorancy agreement effect manifests itself even if palatals are included in statistical calculations.

The objections against models B and C and the combinability of dentals and palatals lead us to the fourth model of SPA reproduced in Table 18, shown above. Here, it is only dentals that are divided by sonorancy. The Slavic languages fit this model perfectly as they avoid both dental subclasses, but even the non-Slavic IE languages show a high degree of dispreference, particularly for repetitions of dental sonorants.

Model D is a combination of model A, originally introduced by P&S for African languages, and model C, derived from the phonotactics of Semitic languages. However, its validity for non-IE languages in general is only tentative. If there is a clear correlation between avoidance and preference within the IE languages, the contrast is less sharp in the non-IE language group. In the IE languages the sonorancy-agreeing combinations tend to be avoided

and not favored, whereas the sonorancy-disagreeing combinations tend not to be avoided; instead, they tend to be favored. In contrast, there are many non-IE languages that avoid the sonoracy-disagreeing combinations and few languages that prefer them. For example, combinations of DENTAL obstruents with DENTAL sonorants are underrepresented in 15 non-IE languages (39% of the 38 languages!), but they are overrepresented only in five languages. Thus, other factors may influence the distribution of DENTAL CVC sequences, and it is quite likely that the form and strength of the avoidance affect is to a large extent language-specific.

Be that as it may, it remains an open question why DENTAL SPA is so prominently affected by sonorancy agreement, at least in some languages. It has been suggested that this has to do with the size of coronal (i.e., DENTAL) inventories. More segments within this class require more distinctive features to distinguish them from each other and other segments. Consequently, "similarity for coronal pairs like /s, n/ is less than the similarity of /f, m/ due to the larger space of contrasts in the coronals[; as] a result, the co-occurrence restriction for /f, m/ is stronger than the co-occurrence restriction for /s, n/" (Frisch et al. 2004: 199; see also Pierrehumbert 1993: 375). However, while we agree that answers to the mystery of DENTALS should be sought in the richness and commonality of these segments, we reject the explanation by distinctive features. It can only be valid if we agree that it is distinctive features (in general or in certain conceptions) and/or phonological contrasts that truly express similarities between segments rather than their articulatory and acoustic properties (cf. Ohala 1990). Moreover, if SPA in general is to be explained by language perception and processing (as it usually is, see Introduction), then we should be able to demonstrate that, say, /f/ and /m/ are indeed perceived as more similar than /s/ and /n/.

Other mysteries lie behind the behavior of PALATALS. First, homorganic CVC sequences of these consonants are avoided only in three out of 18 Slavic varieties. Although these sequences are not universally disfavored even in other language groups, the portion of the avoidant languages is still larger than in Slavic—58% of the other languages (including those in P&S's sample) show a PALATAL avoidance effect or lack PALATAL sequences altogether. The contrast may partly be due to there being no Slavic language without post-alveolar sibilants (a possible exception may be Polabian once we leave out loanwords). The languages lacking these sibilants tend to avoid PALATAL sequences (22 non-Slavic languages are like this, 18 of which have an avoidance effect). However, there are striking differences among the languages that limit their PALATALS to post-alveolar sibilants and /j/, and the languages that have post-alveolar sibilants, /j/, other sonorants such as /ɲ/, and true palatal plosives such as /c/. Both types are among the most frequent PALATAL systems cross-linguistically; in Slavic they are the two most frequent systems. The vast majority of the non-Slavic IE and non-IE languages of both kinds have an avoid-

ance effect for the palatal sequences, whereas the effect has been found only for two Slavic languages out of fourteen.

The second mystery concerns the very lack of palatal avoidance. Why do many languages fail to impose distributional restrictions on the repetitions of palatals when most of these languages avoid homorganic sequences of labials, dorsals, and dentals (dental obstruents and dental sonorants)?[28] It is possible that the lack and/or weakness of palatal SPA relates to the origin of palatal consonants. They have often arisen from the palatalization of dorsals or dentals/alveolars before front vowels (Bateman 2007). Also, the approximant /j/ is a favored epenthetic and prothetic segment (Blevins 2008), or it might result from the dissolution of palatalized consonants (cf. Proto-Slavic **běžati* 'to run' with non-homorganic /běž/ > Modern Czech /bjɛʒɛt/ where /bj/ is from palatalized /bʲ/). Hence, homorganic palatal sequences were originally non-homorganic, and the lack of SPA may reflect this original state. It is then no surprise that there are other restrictions on the distribution of palatals that do not concern their co-occurrence with other consonants, but rather with vowels. One such restriction is the avoidance of immediate combinations of palatals with back vowels (see §5.3).

There is one final issue we would like to address briefly here. Although languages rather strongly and widely avoid CVC sequences with consonants from the same place of articulation, this avoidance is not a strict prohibition. Homorganic CVC sequences exist in all languages so far considered. The obvious question is what SPA-violating words are like. The most often cited examples are CVC sequences with identical consonants. In some languages, these are claimed to be the only violations possible (e.g., in Chol, as shown by Gallagher and Coon 2009; see also Rose 2000), for which reason some researchers ascribed a special status to repetitions of identical consonants within the SPA restrictions (Frisch 2004; Frisch et al. 2004). Yet, the morphological, lexicological, and etymological nature of SPA violations remain little explored. Possible SPA-violating lexical items and morphosyntactic contexts are summarized in P&S; they mention ideophones and other sound-symbolic vocabulary, child-language words and expressive lexical items in general, loanwords, morphological reduplications, and other morpheme-boundary sequences. Pozdniakov (2007) subsequently provided a detailed but partial lexicological analysis of SPA-violating CVC sequences in Modern Russian.

Our analysis of Slavic sheds some light on the nature of SPA-violating items and partly confirms the previous observations. To begin with, the question of loanwords cannot be adequately resolved within IE languages because these languages have mostly borrowed words from each other, that is, from languages restricted by SPA. Secondly, the problem of morpheme boundaries

[28] Another mystery is why the sequences of dental obstruents are not avoided in so many non-IE languages (see Table 18, p. 221).

is more complex than suggested. The situation in Modern Czech indicates that the same kind of SPA restrictions apply for roots and across suffixal boundaries, but no SPA effect is observable at prefixal boundaries. Interestingly, a LABIAL and DORSAL SPA effect is consistently present across word boundaries in the poetry of some writers, but neither effect has been observed at word boundaries in prose.

We have also looked at the origin, meaning, and function of the 1,256 native roots with a reliable etymology that contained a CVC sequence of LABIALS, DENTAL obstruents, DENTAL sonorants, PALATALS, and DORSALS. The roots had been excerpted from Modern Czech, Old Church Slavonic, and Slovenian (including many verbal roots from Czech dialects). As shown below in Table 19, around 60% of the LABIAL and DORSAL homorganic sequences occur in phonologically motivated (sound-symbolic) roots, which includes onomatopoeia, phonaesthematic or expressive formations, and child-language formations. The share of these roots is much smaller for DENTALS and PALATALS. The homorganic CVC sequences of these consonants are mostly secondary developments rather than the original building blocks of the roots. They arose from ancient morphological processes, of which reduplication and suffixation are most common (= "different domain" in Table 19) or are products of sound changes like assimilation or lenition (= "phonological changes"), or both. LABIAL and DORSAL sequences also arose from these processes, but to a much smaller degree. Only DENTAL sequences occur in roots inherited from Proto-Slavic or Proto-Indo-European in the unchanged phonological and morphological form (= "unmotivated original").

This etymological analysis of Slavic confirms that homorganic CVC sequences appear in phonologically motivated vocabulary or are products of sound changes and morphological processes but also reveals that there is a

**Table 19**. Origin of native Slavic roots with homorganic CVC sequences (n = 1,256)

| | N | Unmotivated non-original | | | Unmotivated orig. | Motivated |
|---|---|---|---|---|---|---|
| | | Dif. domain | Phon. changes | Both | | |
| LABIALS | 188 | 20.7% | 9.6% | 5.3% | 0.0% | 64.4% |
| DORSALS | 115 | 16.5% | 7.8% | 3.5% | 0.0% | 72.2% |
| DENT. OBS. | 244 | 34.0% | 25.4% | 5.3% | 4.5% | 30.7% |
| DENT. SON. | 290 | 40.3% | 5.5% | 24.8% | 1.0% | 28.3% |
| PALATALS | 419 | 8.1% | 42.7% | 17.7% | 0.0% | 31.5% |

sharp and statistically significant difference between homorganic sequences of non-coronals (labials and palatals) and coronals (dentals and palatals), which has been advocated throughout this paper. The former are much more inclined to be conveyors of sound symbolism than the latter. The former are also more strongly avoided cross-linguistically. The least associated with sound symbolism are homorganic sequences of palatals, which are also least often avoided cross-linguistically.

Sound symbolism could then be a place to look for an explanation of the differences in SPA strength. Non-coronals may be better predisposed for mimicking the extra-linguistic world than coronals; labials perhaps thanks to their visual saliency (cf. Kumagai 2020 and Kilpatrick et al. 2023), dorsals thanks to their acoustics (cf. Barrera-Pardo 2013 and Elsen 2017). It has been repeatedly pointed out that sound-symbolic vocabulary often features phonological abnormalities (e.g., Kořínek 1934 or Haiman 2018: 114–16). Sound symbolism and phonological peculiarity may go hand in hand, enhancing each other. If so, CVC sequences of labials and dorsals may be more strongly avoided than those of dentals so that their sound-symbolic capacity is more prominent and apparent among otherwise entirely arbitrary signs. However, it is also possible to offer an alternative explanation for the results in Table 19: non-coronals are less likely to be results of sound changes than coronals, and non-coronals are also less likely to figure in word-formative processes of Proto-Slavic and Proto-Indo-European roots. In short, we witness the effectiveness of language here in that some patterns are avoided only to be utilized more prominently elsewhere.

## References

Arvaniti, Amalia. (2007) "Greek phonetics: The state of the art". *Journal of Greek linguistics* 8: 97–208.

Bachra, Bernard N. (2001) *The phonological structure of the verbal roots in Arabic and Hebrew*. Leiden/Boston/Köln: Brill.

Barrera-Pardo, Darío. (2013) "The sound symbolism of velar sounds in English popular fiction". Eva Parra-Membrives and Albrecht Classen, eds. *Literatur am Rand: Perspectiven der Trivialliteratur vom Mittelalter bis 21. Jahrhundert*. Tübingen: Narr Verlag, 47–62.

Basbøll, Hans. (2005) *The phonology of Danish*. Oxford: Oxford University Press.

Bateman, Nicoleta. (2007) *A crosslinguistic investigation of palatalization*. PhD dissertation, University of California.

Berent, Iris, and Joseph Shimron. (1997) "The representation of Hebrew words: Evidence from the Obligatory Contour Principle". *Cognition* 64: 39–72.

Berkley, Deborah M. (1994) "Variability in Obligatory Contour Principle effects". Katherine Beals, Jeannette Denton, Robert Knippen, Lynette Melnar, Hisami Suzuki, and Erica Zeinfeld, eds. *The parasession on variation in linguistic theory*. Vol. 2 of *Papers from the 30th regional meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1–12.

————. (2000) *Gradient Obligatory Contour Principle*. PhD dissertation, Northwestern University.

Bičan, Aleš. (2018) "Etymologie a fonologie: případ Mathesiova fonotaktického pravidla" [Etymology and phonology: The case of Mathesius's phonotactic rule]. Petr Malčík, ed. *Vesper Slavicus: Sborník k nedožitým devadesátinám prof. Radoslava Večerky*. Praha: Nakladatelství Lidové noviny, 13–31.

————. (forthcoming) "Consonant-vowel combinations in Slavic and other languages".

Blevins, Juliette. (2008) "Consonant epenthesis: Natural and unnatural histories". Jeff Good, ed. *Linguistic universals and language change*. Oxford: Oxford University Press, 79–107.

Boll-Avetisyan, Natalie, and Réne Kager. (2014) "OCP-PLACE in speech segmentation". *Language and speech* 57(3): 294–421.

Booij, Geert. (2011) "Morpheme structure constraints". Mark Van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, eds. *Phonological interfaces*. Vol. 4 of *The Blackwell companion to phonology*. Oxford: Blackwell, 2049–69.

Buckley, Eugene. (1997) "Tigrinya root consonants and the OCP". *Penn working papers in linguistics* 4(3): 19–51.

Cantineau, Jean. (1946) "Esquisse d'une phonologie de l'Arabe Classique". *Bulletin de la Société Linguistique* 43: 93–140.

Carrissimo-Bertola, Manon. (2010) *Structures syllabiques des unités lexicales: "The fronting effect"*. MA thesis, Université Stendhal.

Cathcart, Chundra A. (2023) "The evolution of similarity avoidance: A phylogenetic approach to phonotactic change". Noah Elkins, Bruce Hayes, Jinyoung Jo, and Jian-Leat Siah, eds. *Proceedings of the 2022 Annual meeting on phonology*, 1–12. Available at: https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/5434/5206. Last accessed 5 February 2024.

Coetzee, Andries W. (2005) "The Obligatory Contour Principle in the perception of English". Sonia Frota, Marina Vigário, and Maria João Freitas, eds. *Prosodies*. Berlin/New York: De Gruyter Mouton, 223–45.

Coetzee, Andries W., and Joe Pater. (2008) "Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic". *Natural language and linguistic theory* 26: 289–337.

Cooper, Adam I. (2009) "Similarity avoidance in the Proto-Indo-European root". *Penn working papers in linguistics* 15(1): 55–64.

Danis, Nicholas S. (2017) *Complex place and place identity*. PhD dissertation, State University of New Jersey.

Davis, Stuart. (1991) "Coronals and the phonotactics of nonadjacent consonants in English". Carole Paradis and Jean-François Prunet, eds. *Phonetics and phonology 2: The special status of coronals*. New York: Academic Press, 49–60.

Derksen, Rick. (2008) *Etymological dictionary of the Slavic inherited lexicon*. Leiden/Boston: Brill.

Dmitrieva, Olga. (2008) "The gradient phonotactics of English CVC syllables". Unpublished manuscript. Available at: https://web.ics.purdue.edu/~odmitrie/papers/Dmitrieva2007_OnEnglishOCP.pdf. Last accessed 5 February 2024.

Doucette, Amanda, Timothy J. O'Donnell, Morgan Sonderegger, and Heather Goad. (2024) "Investigating the universality of consonant and vowel co-occurrence restrictions". *Glossa* 9(1): 1–39.

Elmedlaoui, Mohamed. (1995) "Géométrie des restrictions de cooccurrence de traits en sémitique et en berbère: synchronie et diachronie". *Canadian journal of linguistics/Revue canadienne de linguistique* 40(1): 39–76.

Elsen, Hilke. (2017) "The two meanings of sound symbolism". *Open linguistics* 3: 491–99.

Frisch, Stefan A. (1997) *Similarity and frequency in phonology*. PhD dissertation, Northwestern University.

————. (2004) "Language processing and segmental OCP effects". Bruce Hayes, Robert Kirchner, and Donca Steriade, eds. *Phonetically based phonology*. Cambridge, UK: Cambridge University Press, 346–71.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. (2004) "Similarity avoidance and the OCP". *Natural language and linguistic theory* 22: 179–228.

Frisch, Stefan A., and Bushra A. Zawaydeh. (2001) "The psychological reality of OCP-place in Arabic". *Language* 77(1): 91–106.

Gallagher, Gillian. (2015) "Natural classes in cooccurrence constraints". *Lingua* 166: 80–98.

Gallagher, Gillian, and Jessica Coon. (2009) "Distinguishing total and partial identity: Evidence from Chol". *Natural language and linguistic theory* 27: 545–82.

Graff, Peter N. H. M. (2012) *Communicative efficiency in the lexicon*. PhD dissertation, Massachusetts Institute of Technology.

Greenberg, Joseph A. (1950) "The patterning of root morphemes in Semitic". *Word* 6: 162–81.

Grotberg, April L. (2022) *Quantifying phonological feature co-occurrence*. PhD dissertation, Purdue University.

Haiman, John. (2018) *Ideophones and the evolution of language*. Cambridge, UK: Cambridge University Press.

Hall, T. Alan. (1997) *The phonology of coronals*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hamann, Silke. (2002) "Postalveolar fricatives in Slavic languages as retroflexes". Sergio Baauw, Mike Huiskes, and Maaike Schoorlemme, eds. *OTS yearbook 2002*. Utrecht: Utrecht Institute of Linguistics, 105–27.

Haspelmath, Martin, and Uri Tadmor, eds. (2009) *World loanword database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: http://wold.clld.org. Last accessed 1 November 2024.

Havlová, Eva, Adolf Erhart, and Ilona Janyšková, eds. (1989–2022) *Etymologický slovník jazyka staroslověnského* [Etymological dictionary of the Old Church Slavonic language], vols. 1–21. Prague: Karolinum; Brno: Tribun.

Hayward, Katrina M., and Richart J. Hayward. (1989) "'Guttural': Arguments for a new distinctive feature". *Transactions of the Philological Society* 87(2): 179–93.

Hollander, Myles, Douglas A. Wolfe, and Eric Chicken. (2014) *Nonparametric statistical method*. 3rd edition. New Jersey: Wiley.

Howson, Phil. (2017) "Upper Sorbian". *Journal of the International Phonetic Association* 47(3): 359–67.

Hualde, José Ignacio. (2005) "Quasi-phonemic contrasts in Spanish". Benjamin Schmeiser, Vineeta Chand, Ann Kelleher, and Angelo Rodriguez, eds. *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press, 374–98.

Hulden, Mans. (2017) "A phoneme clustering algorithm based on the Obligatory Contour Principle". Roger Levy and Lucia Specia, eds. *Proceedings of the 21st conference on computational natural language learning (CoNLL 2017)*. Vancouver: Association for Computational Linguistics, 290–300.

Janson, Tore. (1986) "Cross-linguistic trends in the frequency of CV sequences". *Phonology yearbook* 3: 179–95.

Kawahara, Shigeto, Hajime Ono, and Kiyoshi Sudo. (2006) "Consonant co-occurrence restrictions in Yamato Japanese". *Japanese/Korean linguistics* 14: 27–38.

Khan, Sameer ud Dowla. (2007) "Similarity avoidance in East Bengali fixed-segment reduplication". Erin Bainbridge and Brian Agbayani, eds. *Proceedings of the 34th Western Conference on Linguistics*. Fresno: Department of Linguistics at California State University, 257–71.

Kilpatrick, Alexander, Alexandra Ćwiek, Eleanor Lewis, and Shigeto Kawahara. (2023) "A cross-linguistic, sound-symbolic relationship between labial consonant, voiced plosive, and Pokémon friendship". *Frontiers in psychology* 14: 1–11.

Kinney, Ashley L. (2005) *Serial organization of speech sounds in Creole languages*. PhD dissertation, University of Texas.

Kokkelmans, Joachim. (2021) *The phonetics and phonology of sibilants*. PhD dissertation, University of Verona.

Kolgjini, Julie M. (2004) *Palatalization in Albanian: An acoustic investigation of stops and affricates*. PhD dissertation, University of Texas at Arlington.

Kořínek, Josef M. (1934) *Studie z oblasti onomatopoje* [Studies from the area of onomatopoeia]. Prague: Nákladem Filosofické fakulty University Karlovy.

Kumagai, Gakuji. (2020) "The pluripotentiality of bilabial consonants: The images of softness and cuteness in Japanese and English". *Open linguistics* 6: 693–707.

Ladefoged, Peter, and Ian Maddieson. (1996) *The sounds of the world's languages*. Oxford: Blackwell.

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch, and Russell D. Gray. (2022) "Lexibank, a public repository of standardized wordlists with computed phonological and lexical features". *Scientific data* 9(316): 1–16.

MacEastern, Margaret. (1999) *Laryngeal co-occurrence restrictions*. New York: Garland.

MacKay, Donald. (1970) "Phoneme repetition in the structure of languages". *Language and speech* 13(3): 199–213.

MacNeilage, Peter F., and Barbara L. Davis. (1999) "Consonant (vowel) consonant sequences in early words". *Proceedings of the 14th International Congress of Phonetic Sciences*. Vol. 3. San Francisco: University of California, Berkeley, 2489–92.

————. (2000) "Evolution of speech: The relation between ontogeny and phylogeny". Chris Knight, Michael Studdert-Kennedy, and James R. Hurford, eds. *The evolutionary emergence of language*. Cambridge, UK: Cambridge University Press, 146–60.

MacNeilage, Peter F., Barbara L. Davis, Ashley Kinney, and Christine Matyear. (2000) "The motor core of speech: A comparison of serial organization patterns in infants and languages". *Child development* 71: 153–63.

Maddieson, Ian. (1984) *Patterns of sounds*. Cambridge, UK: Cambridge University Press.

Mathesius, Vilém. (1931) "O výrazové platnosti některých českých skupin hláskových" [On the expressive validity of some Czech sound groups]. *Naše řeč* 15: 38–40.

Mayer, Thomas, Christian Rohrdantz, Frans Plank, Peter Bak, Miriam Butt, and Daniel A. Keim. (2010) "Consonant co-occurrence in stems across languages: Automatic analysis and visualization of a phonotactic constraint". Fei Xia and Lori Levin, eds. *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*. Uppsala: Association for Computational Linguistics, 70–78.

McCarthy, John. (1986) "OCP effect: Gemination and antigemination". *Linguistic inquiry* 17(2): 207–63.

————. (1994) "The phonetics and phonology of Semitic pharyngeals". Patricia A. Keating, ed. *Phonological structure and phonetic form*. Cambridge, UK: Cambridge University Press, 191–223. [Papers in Laboratory Phonology III.]

Nikitina, Tatiana. (2022) "At the intersection of synchrony and diachrony: A phonotactic analysis of the lexicon of Wan". *Languages in Africa* 3(2): 196–211.

Ohala, John J. (1990) "There is no interface between phonology and phonetics". *Journal of phonetics* 18: 153–71.

Ohala, Manjari. (1999) "Hindi". International Phonetic Association, ed. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge, UK: Cambridge University Press, 100–103.

Padgett, Jaye. (1992) "OCP subsidiary features". *NELS* 22(1): 335–46.

————. (1995) *Stricture in Feature Geometry*. Stanford: CSLI Publications.

Pierrehumbert, Janet. (1993) "Dissimilarity in the Arabic verbal roots". *NELS* 23: 367–81.

————. (1994) "Syllable structure and word structure: A study of triconsonantal clusters in English". Patricia A. Keating, ed. *Phonological structure and phonetic form*. Cambridge, UK: Cambridge University Press, 168–88. [Papers in Laboratory Phonology III.]

Plénat, Marc. (1996) "De l'interaction des contraints: Une étude de case". Jacques Durand and Bernard Laks, eds. *Current trends in phonology: Models and methods*. Salford: ESRI, 585–617.

Pozdniakov, Konstantin. (2007) "O *'ploxix' i 'xorošix' slovax* v russkom jazyke" [On 'bad' and 'good' words in the Russian language]. N. Vakhtin and G. Levinton, eds. *Hommage Albert Baiburin*. St. Petersburg: Université Européenne, 78–119. Available at: http://pozdnia-kov.free.fr/publications/2007_Baiburin.pdf. Last accessed 22 February 2024.

————. (2010) "Ograničenija distantnoj sočetaemosti soglasnyx v russkom jazyke" [Limitations of the distant combinability of the consonants in CVC sequences in Russian]. V. Z. Dem'jankov and V. J. Porhomovskij, eds. *V prostranstve jazyka i kul'tury: zvuk, znak, smysl. Sbornik statej v čest' 70-letija V. A. Vinogradova*. Moscow: Jazyki slavjanskix kul'tur, 113–26. Available at: http://pozdniakov.free.fr/publications/2010_Vinogradov.pdf. Last accessed 22 February 2024.

————. (2011) "Fonotaktika russkogo kornja i ee potencial'nye resursy" [Phonotactics of the Russian root and its potential resources]. *Hommage à Vadim Kasevich*. St. Petersburg University, 119–74. Available at: http://pozdniakov.free.fr/publications/2011_Phonotactics_Russian.pdf. Last accessed 22 February 2024.

Pozdniakov, Konstantin, and Guillaume Segerer. (2007) "Similar place avoidance: A statistical universal". *Linguistic typology* 11: 307–48.

Pystynen, Juho. (2014) "Similar place avoidance in language history". *Freelance reconstruction*. Blog post. Available at: https://protouralic.wordpress.com/2014/10/11/similar-place-avoidance-in-language-history/. Last accessed 5 February 2024.

Rácz, Péter, Jennifer Hay, Jeremy Needle, Jeanette King, and Janet Pierrehumbert. (2016) "Gradient Māori phonotactics". *Te Reo* 59: 3–21.

Reiner, Erica. (1966) *A linguistic analysis of Akkadian*. The Hague: Mouton and Co.

Rocasens, Daniel. (1990) "The articulatory characteristics of palatal consonants". *Journal of phonetics* 18: 267–80.

Rose, Sharon. (2000) "Rethinking geminates, long-distance geminates, and the OCP". *Linguistic inquiry* 31(1): 85–122.

Rose, Sharon, and Lisa King. (2007) "Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic languages". *Languages and speech* 50(4): 451–504.

Rosenthal, James A. (1996) "Qualitative descriptors of strength of association and effect size". *Journal of social service research* 21(4): 37–59.

Rousset, Isabelle. (2004) *Structures syllabiques et lexicales des langues du monde: Données, typologies, tendances universelles et contraintes substantielles*. PhD dissertation, Université Stendhal.

Sandell, Ryan. (2015) "Obligatory Contour Principle effects in Indo-European phonology: Statistical evidence and the morphology-phonology interface". Stephanie W. Jamison, H. Craig Melchert, Brent Harmon Vine, and Angelo Mercado, eds. *Proceedings of the 26th Annual UCLA Indo-European Conference*. Bremen: Hempen, 141–60.

Siptár, Péter, and Miklós Törkenczy. (2000) *The phonology of Hungarian*. Oxford: Oxford University Press.

Stanton, Juliet, and John F. Stanton. (2022) "In defense of O/E". Unpublished manuscript. Available at: https://ling.auf.net/lingbuzz/006391. Last accessed 5 February 2024.

Tang, Sze-Wing. (2000) "Identity avoidance and constraint interaction: The case of Cantonese". *Linguistics* 38(1): 33–61.

Teil-Dautrey, Gisèle. (2008) "Et si le proto-bantu était aussi une langue … avec ses contraintes et ses déséquilibres". *Diachronica* 25(1): 54–110.

Valleé, Nathalie, Solange Rossato, and Isabelle Rousset. (2009) "Favoured syllabic patterns in the world's languages and sensorimotor constraints". François Pellegrino, Egidio Marsico, Ioana Chitoran, and Christophe Coupé, eds. *Approaches to phonological complexity*. Berlin/New York: Mouton de Gruyter, 111–39.

van Goch, Wil. (2010) *The Obligatory Contour Principle: Consonant co-occurrence restrictions in Dutch*. PhD dissertation, Utrecht University.

Walter, Mary A. (2007) *Repetition avoidance in human language*. PhD dissertation, Massachusetts Institute of Technology.

————. (2010) "Harmony versus the OCP: Vowel and consonant cooccurrence in the lexicon". *Laboratory phonology* 1: 395–413.

Weiss, Michael. (2016) "The Proto-Indo-European laryngeals and the name of Cilicia in the Iron Age". Andrew Miles Byrd, Jessica DeLisi, and Mark Wenthe, eds. *Tavet Tat Satyam: Studies in honor of Jared S. Klein on the occasion of his seventieth birthday*. Ann Arbor, MI: Beech Stave, 331–40.

Wilson, Colin, and Marieke Obdeyn. (2009) "Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay". Unpublished manuscript. Available at: https://linguistics.ucla.edu/people/hayes/219/papers/WilsonObdeyn2009.pdf. Last accessed 5 February 2024.

Wright, Richard. (2004) "A review of perceptual cues and cue robustness". Bruce Hayes, Robert Kirchner, and Donca Steriade, eds. *Phonetically based phonology*. Cambridge, UK: Cambridge University Press, 34–56.

Yip, Moira. (1989) "Feature geometry and cooccurrence restrictions". *Phonology* 6: 349–74.

Zukoff, Sam. (2017) *Indo-European Reduplication: Synchrony, Diachrony, Theory*. PhD dissertation, Massachusetts Institute of Technology.

Zygis, Marzena. (2003) "Phonetic and phonological aspects of Slavic sibilant fricatives". *ZAS papers in linguistics* 3: 175–213.

Aleš Bičan
Department of Linguistics
and Baltic Studies
Faculty of Arts, Masaryk University
Brno, Czech Republic
and
Czech Language Institute
Czech Academy of Sciences
Brno, Czech Republic
ORCID: 0000-0003-2481-6023
bican@phil.muni.cz

# Looking Beyond Information Structure: Evidence from Magnitude Estimation Test Experiments for Weight Effects on Slovak Word Order

**Jakob Horsch and Martina Ivanová**

*Abstract*: The principle of end-weight (Quirk et al. 1985) posits that language users prefer to place short ("light") before long ("heavy") constituents because this is easier to cognitively process (Hawkins 1994). Weight effects on constituent order have been discussed for well over a century in languages like English and German (cf. Behaghel 1909, 1930). In Slavic languages, however, they have received little attention so far (cf. Kizach 2012: 251).

Rather, the focus has been on information structure (Short 2002: 494). However, Goldberg's Tenet #5 predicts that general cognitive restraints such as weight effects apply across languages (Goldberg 2003: 219). This article presents the results of a pilot study that investigates a phenomenon known as Heavy NP Shift, in which the $_{VP}$[NP PP] pattern changes to $_{VP}$[PP NP] when the NP is heavier than the PP.

Employing the Magnitude Estimation Test method (Hoffmann 2013), grammaticality acceptability ratings from 39 L1 Slovak speakers were elicited. The results show that Slovak is susceptible to weight effects, such that placing short before long constituents is always preferred. Moreover, the results provide evidence for the existence of a $_{VP}$[V NP PP] pattern in Slovak that has been identified as "basic" for English (Hawkins 1994: 20). This supports Hawkins's (2004) Performance-Grammar Correspondence Hypothesis, which posits that grammaticalized patterns in analytic languages are the result of performance preference and therefore preferred in synthetic languages.

## 1. Introduction*

The effect of constituent weight (i.e., constituent length and complexity) on word order has been observed and discussed for well over a century

* We would like to thank Monika Hornáček Banášová at the University of Ss. Cyril and Methodius in Trnava, Slovakia, and Peter Barrer at Comenius University in Bratislava, Slovakia, for their invaluable assistance in recruiting subjects and providing facilities for carrying out the Magnitude Estimation Test experiments. We express our gratitude to Thomas Hoffmann, who has provided valuable advice regarding the Magnitude Estimation Test method and statistical analysis of the data.

This paper was written in the context of the research project 09I03-03-V04-00713 "Weight Effects in English and Slovak" at the Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences. The project is generously funded by the Government of the Slovak Republic's Recovery and Resilience program (scholarship for excellent researchers R2-R4), funded in the context of the European Union's NextGenerationEU package.

(cf. Behaghel 1909, 1930). The "principle of end-weight" (Quirk et al. 1985) states that placing light before heavy constituents is preferred due to processing efficiency effects (Hawkins 1994), where "light" means short/less complex, and "heavy" means long/more complex.[1] This light-before-heavy principle can be explained as being rooted in locality-based approaches, which predict a processing advantage when the distance between major syntactic heads is minimized. For example, in the context of his theory of Early Immediate Constituents (EIC), Hawkins (1994) has argued that syntactic heads being placed closer to each other result in locality effects (cf. also Gibson's (1998, 2000) syntactic prediction locality theory). This has led to $_{VP}$[NP PP] becoming the "basic order of the English VP" (Hawkins 1994: 20) because PPs tend to be heavier (i.e., longer and more complex) than NPs. To illustrate this, consider examples (1) and (2) (from Hawkins 1994: 20):

(1) John $_{VP}$[gave $_{NP}$[a book] $_{PP}$[to Mary]]

(2) John $_{VP}$[gave $_{PP}$[to Mary] $_{NP}$[a book]]

According to EIC, English speakers will prefer (1), because here the heads of the three syntactic constituents (V, NP, PP) in the VP are adjacent, which is the optimal situation from a locality perspective for ease of processing. Note that this follows a cognitive approach that does not assume an "unmarked" word order due to semantic roles or case.

Now, when the NP is heavier than the PP, a phenomenon called Heavy NP Shift[2] (HNPS) occurs: The NP is "shifted" so that V and DO are separated by a prepositional phrase (Mains et al. 2015), thereby increasing processing efficiency as in (3) (example from Hawkins 1994: 57):

(3) I $_{VP}$[gave $_{PP}$[to Mary] $_{NP}$[the valuable book that was extremely difficult to find]]

It is noteworthy that HNPS does not occur categorically. Studies have shown that it is most likely when the weight difference between a light PP

[1] In the following text, we operationalize constituent weight as the number of orthographic words in a constituent, a simple and tried-and-tested method (cf., e.g., Szmrecsányi 2004: 1037) based on the observation that the "number of words, number of nodes, and number of phrasal nodes […] are so highly correlated that it is impossible to choose among them on empirical grounds" (Wasow and Arnold 2003: 121). Accordingly, in the following text, "light" is to be understood as fewer, and "heavy" as more words.

[2] The term "shift" originates in analyses based on derivational models of syntax (Mains et al. 2015).

and a heavy NP is at least four words (Hawkins 1994; cf. also Wasow 1997; Stallings and MacDonald 2011). In fact, even when there are considerable differences in weight, HNPS does not always occur (cf. Wasow 2002). In this context, previous studies (e.g., Konieczny 2000; Konieczny and Döring 2003) have offered an explanation in anti-locality effects, which are explained through expectation-based accounts. In contrast to locality theories which are based on memory limitations, expectation-based approaches posit different key cognitive constraints as determining syntactic complexity (cf. Gibson and Wu 2013). A prominent example of an expectation-based approach is the so-called surprisal theory, according to which the difficulty of processing an input in the context in which it appears decreases as the input's conditional probability increases (Hale 2001; Levy 2008; Smith and Levy 2008, 2013). To estimate surprisal, different psychometric variables have been used recently in the theoretical studies, including acceptability judgements (if surprisal correlates with perception, one should expect responses that are surprising in context to obtain lower scores, i.e., a negative relationship between surprisal and score) (e.g., Lau et al. 2017; Wallbridge et al. 2022). Thus, expectation-based theories can also provide explanations for longer-before-shorter orders (cf. Balling and Kizach 2017).

Furthermore, ceiling effects have been observed, to the effect that "increasing NP length appears to bring HNPS only to a par with the canonical [i.e., basic] constituent order, such that HNPS is never significantly preferred over the canonical [i.e., basic] word order" (Medeiros et al. 2021: 437). In other words, speakers are highly reluctant to deviate from $_{VP}$[NP PP], underscoring the fact that it is "basic" and has been "grammaticalized" in English (Hawkins 1994: 20).

What makes HNPS particularly interesting is that it allows investigating cognitive effects on word order with minimal "interference" from other factors because it is a "rare case of phrase ordering that does not affect grammatical role assignment" (Stallings et al. 1998: 395). That is, the direct object NP stays a direct object NP even when shifted;[3] this is not the case with other syntactic alternations such as the ditransitive/to-dative alternation. In fact, HNPS is a perfect candidate for investigating the effects of processing efficiency on constituent order because, unlike other syntactic alternations,[4] it is triggered exclusively by constituent length/complexity. Neither is it stylistically

[3] In generative approaches, HNPS is argued to belong to the category of wh-movement, i.e., it is assumed that A-movement is responsible for the shift (Haegeman and Guéron 1999: 221–24), because the moved NP is no longer in the canonical direct object position and has not been assigned a different case by its new position, which would be the case in A-movement (cf. also Haegeman 1994: 418–21).

[4] HNPS has also been discussed in the context of patterns other than $_{VP}$[V NP PP], e.g., the dative alternation. However, all of these syntactic alternations appear to

motivated nor have animacy effects been found (Stallings et al. 1998: 410). Moreover, Kizach (2014) has noted that speakers "care more about complexity than about givenness" because "the potential benefit of heeding complexity (having short-before-long wherever possible) is […] much larger". Similarly, Šimík (2021: 3) finds that "[i]nformation structure and syntax […] are related only very loosely" in Czech, explicitly refuting the notion that "syntax, which generates [word order] alternations is tightly related to information structure, which motivates them".

The claim that weight effects prevail over givenness when it comes to word ordering is particularly relevant for research on the Slavic languages. While there is a large body of research on the principle of end-weight and HNPS in languages such as English and German (e.g., Behaghel 1909: 139, 1930: 85; Quirk et al. 1972; Quirk et al. 1985; Hawkins 1994, 2004, 2014; Wasow 1997; Arnold et al. 2000; Stallings et al. 1998; Stallings and MacDonald 2011; Mains et al. 2015; Melnick 2017; Medeiros et al. 2021), in Slavic languages little attention has been paid to weight effects so far. For example, Kizach (2012: 251) has noted that "whereas newness has repeatedly been argued as being a relevant factor for Russian, little or no attention has been paid to weight". Rather, the focus has been on information structure, i.e., givenness.[5] For example, Short notes that it is consistently claimed in grammars of Slovak that "pragmatic constituent order takes precedence over syntax" (Short 2002: 494), and that "[m]odern Slovak sources decline to refer to any unmarked order of constituents in terms of basic word order" (Short 2002: 566).[6] In other

be semantically motivated. For the purposes of the present investigation, the term "heavy NP shift" entails none of these other syntactic alternations.

[5] We are grateful to an anonymous reviewer for pointing out that "[w]hile the term 'information structure' was introduced […] by Halliday (1967), by and large in the sense of givenness, its meaning nowadays is broader, encompassing also topic-comment and background-focus structuring (see, e.g., Lambrecht 1994 […] or Krifka 2008)".

[6] Note that Šimík et al. (2014) investigated the impact of givenness on the position of the direct object (DO) in Czech with respect to three clause-mate constituents—subject (S), verb (V), and VP-modifying PP (X). In two controlled acceptability judgment experiments, two main observations were established: (i) for objects in all-new contexts and (ii) for given objects. In contexts with given objects, the order S-V-X-DO (condition 1) was significantly less acceptable than all other orders; the orders S-V-DO-X, S-DO-V-X, and DO-S-V-X (conditions 2, 3, and 4, respectively) showed no significant difference (conditions 2 and 3 violated the GIVEN–NEW principle and condition 4 did not violate it). In this study, only two variables were used in the design of the experiment: referentiality of the object and position of the object. However, in one of the examples presented in Šimík et al.'s (2014) Experiment 1 (in all-new contexts), object NP and VP-modifying PP differed with respect to weight (light/heavy). The results of the study showed that there was no significant difference

words, neither has processing efficiency been discussed as a factor affecting word order in Slovak, nor has there been any mention of a resulting "basic" word order such as the one that Hawkins (1994) determined for English (see above).[7] This is evident from our review of numerous grammars of the Slovak language (Pauliny et al. 1963; Orlovský 1971; Mistrík 1983, 1988, 2003; Pauliny 1981, 1997; Pavlovič 2012; Ivanová 2016) as well as more than six decades' worth of issues of two of the most authoritative linguistic journals on the Slovak language, *Jazykovedný časopis* (https://www.juls.savba.sk/ediela/jc/) and *Slovenská reč* (https://www.juls.savba.sk/ediela/sr/).

However, it is unlikely that weight effects should not play a role in word ordering in Slovak. As a synthetic language, its word order is certainly more flexible than English, but it should nevertheless follow the principle of end-weight and exhibit Heavy NP shift. This is predicted by Goldberg's Tenet #5, which posits that "[c]ross linguistic generalizations are explained by appeal to general cognitive constraints" (Goldberg 2003: 219). Moreover, Hawkins's Performance-Grammar Correspondence Hypothesis (PGCH) states:

> When the grammar of one language is more restrictive and eliminates one or more structural options that are permitted by the grammar of another, the restriction will be in accordance with performance preferences. The preferred structure will be retained and "fixed" as a grammatical convention, the dispreferred structures will be removed. (Hawkins 2004: 5)

In other words, PGCH predicts that the fixed word orders of analytic languages are the result of performance preference due to processing efficiency, and that these fixed word orders will mirror word orders that are preferred in languages with more flexible word orders, i.e., synthetic languages. This

between the word orders *heavy non-referential object NP — light PP* and *light PP — heavy non-referential object NP* (on a Likert-type rating scale with nine points, the first word order achieved a mean rating of 7.79 and the second word order a mean rating of 7.6). Unfortunately, Šimík et al. (2014) did not investigate any weight distinction of the investigated object NPs and PPs in Experiment 2 for given objects. We therefore suggest that the results of their experiment (which is robustly corroborated by the data), i.e., that the scrambling of the given object is just as acceptable as keeping it *in situ* as long as it is not placed clause-finally, should be further tested for weight effects.

[7] As noted by an anonymous reviewer, "for the closely related Czech, it has been taken for granted that there is something like basic word order, see, e.g., Sgall et al. (1980) et seq., where basic order is called 'systémové' (systemic in English texts) and where it is one of the core principles determining the order within the new (not 'contextually bound') area of the utterance". Furthermore, the reviewer has pointed out that Siewierska and Uhlířová (1998) in their survey "also don't seem to mention explicitly a basic order (which they discuss primarily based on Czech and Polish)".

has been demonstrated in corpus-based studies on the Slovak comparative correlative (CC), a biclausal construction (e.g., [*Čím menej rečí tu bude,*]$_{C1}$ [*tým skôr zaspím*]$_{C2}$ 'The less talking there is here, the sooner I will fall asleep'), where Horsch (2020, 2021) showed that there is a statistically significant preference for constituent orders in both subclauses C1 and C2 that mirror the grammaticalized order in English.

These observations have led us to the following **research questions**:

i. As in English, is there a preference for placing light before heavy constituents in Slovak?

ii. Has this led to a basic, grammaticalized order such as the one that has been determined for English?

Our **hypotheses** are as follows:

i. Goldberg's Tenet #5 (Goldberg 2003: 219) maintains that general cognitive restraints such as processing efficiency will result in cross-linguistic generalizations. Based on research on English that has adduced ample evidence in support of the principle of end-weight, and thus, a preference for light constituents to precede heavy constituents, we predict that the same holds true for Slovak.

ii. Hawkins's (2004) PGCH predicts that the fixed word orders of analytic languages such as English will be mirrored in preferred word orders in more synthetic languages such as Slovak. We therefore predict that Slovak also has a constituent order that can be described as "basic". Given the findings for English, we predict that this is the $_{VP}$[NP PP] order.[8]

To answer the research questions outlined above, we decided to carry out experiments using the Magnitude Estimation Test method to elicit grammaticality acceptability judgments from L1 Slovak speakers. The method is discussed in Section 2; the results are presented in Section 3. Section 4 offers

[8] Šimík and Wierzba (2017), using gradient acceptability judgments and multiple regression, showed that, for Czech, Slovak, and Polish, postverbal objects were equally acceptable irrespective of whether the objects were preceded or followed by an indirect object or an adverbial, i.e., the orders $_{VP}$[object NP PP] and $_{VP}$[PP object NP] showed no significant difference in acceptability. Their results display no significant effects on the relative ordering of object NP and PP. However, the interaction between weight order and phrase order was not tested in this study.

a discussion of these results in light of current research, and in Section 5 we offer a conclusion, including suggestions for further research.

## 2. Methodology: Magnitude Estimation Test (MET)

The MET method (Bard et al. 1996; Cowart 1997; Hoffmann 2013) features grammaticality acceptability judgments,[9] i.e., it requires subjects to "judge stimuli relatively to a reference item" (Hoffmann 2013: 99) instead of restricting them to a scale with absolute values, e.g., the Likert scale. The method is based on the assumption that humans are better at making relative judgments than absolute judgments (Hoffmann 2013: 99).[10] As Cowart has pointed out, the ratings thus obtained translate to "proportional relation among the numbers assigned to different stimuli", which "should reflect the proportions of the stimuli themselves; thus a sentence that is judged twice as good as another should get a number twice as high as that assigned to the other sentence" (Cowart 1997: 73–74).

In practice, MET questionnaires present a set of sentences (test items) to subjects, who rate them with a number that they are basically free to choose. However, they are asked to provide this number "proportional to a constant

[9] Note that such experiments measure *degrees of acceptability* (Hoffmann 2013: 104; who also refers to Featherston 2007a, b; Fanselow 2007; and Newmeyer 2007) and not grammaticality in the Chomskian sense, i.e., whether a sentence belongs to the language system, and is therefore "grammatical", or does not, making it "ungrammatical" (as discussed in Hoffmann 2013: 104). That is to say, while the experiments that are part of the present investigation were carefully designed and implemented following best practices in an attempt to "get as close to grammaticality as possible" (Hoffmann 2013: 104), their results do not permit any definitive claims about the grammaticality (or ungrammaticality, for that matter) of the investigated structures in the Chomskian sense.

[10] As pointed out by an anonymous reviewer, MET as a method has recently also been critically scrutinized (e.g., Weskott and Fanselow 2011; Sprouse 2011; Fukuda et al. 2012; Verhagen et al. 2020). However, as noted by Hoffmann (p.c. in August 2024), MET seems to yield reliable results akin to other methods. Verhagen et al., for example, arrive at the conclusion that "ratings expressed on a Magnitude Estimation scale did not differ systematically from the ratings expressed on a Likert scale" (Verhagen et al. 2020: 61–62), and similarly, Fukuda et al. (2012: 335) found "overwhelming consistency in their results" from Likert scale and MET tests. On a final note, the main point of criticism seems to be that the method is "more difficult for experimenters to implement and requires more mathematical sophistication on the part of experiment participants" (Fukuda et al. 2012: 336).

However, in the authors' own experience with conducting MET experiments for the present study as well as a previous one (Horsch 2023a, b), subjects did not report any notable challenges in understanding and applying the method. Moreover, preparations for the experiments and their implementation did not cause any considerable difficulties.

reference sentence" (Hoffmann 2013: 99) that they have assigned a number of their choosing at the beginning of the experiments. Thus "subjects do not have to rate stimuli on a scale […] which might artificially limit their choices" but rather "decide on their own scale and make as many fine-grained choices as they deem necessary" (Hoffmann 2013: 103). Our questionnaires also featured grammatical and ungrammatical fillers, which later provided baselines against which test items can be compared. This makes it possible to assess test items not just in relation to each other, but also in relation to grammatically correct and incorrect items.

To answer the research questions outlined in §2, we decided to test two variables with two levels each, WEIGHT ORDER (levels: LIGHT-HEAVY, HEAVY-LIGHT) and PHRASE ORDER (levels: NP-PP, PP-NP). We wanted to test all possible factor combinations (conditions):

- LIGHT NP — HEAVY PP
- HEAVY NP — LIGHT PP
- LIGHT PP — HEAVY NP
- HEAVY PP — LIGHT NP

Following Cowart (1997: 49–50), the aim was to present to every participant all four of these conditions but use different lexical material each time to ensure that "the informant is never able to confront a sentence in quite the same way twice" (Cowart 1997: 50). Because of the low number of conditions (four), we decided to create[11] eight lexicalizations to test each condition twice. This also allowed us to install a safeguard against lexical effects since two lexicalizations have been shown to be able to neutralize lexical effects of any items other than the investigated variables (cf. Schütze and Sprouse 2013). Each lexicalization featured a VP consisting of a V, an NP and a PP, the latter of which featured the prepositions *do* '(in)to', *na* 'on(to)', *pod* 'under', *z* 'out of', 'from', *medzi* 'between', and *od* 'from' (4a–h):[12]

[11] Note that due to time constraints, we created experimental sentences from scratch and did not use authentic sentences; it is acknowledged that some experimental studies featuring acceptability ratings use authentic or shortened authentic examples (e.g., Divjak 2017).

[12] In the valency patterns captured in valency dictionaries, object actants and obligatory adverbials are captured by the scheme S–V–O–PP with object actants preceding adverbials. However, this is merely a convention which does not exclude the possible scrambling of the object actant in the sentence structure. The possible influence of the valency status on their ordering within sentence structure has been investigated in several studies on word order for Czech. For example, word order involving object actants and adverbial adjuncts was investigated by Rysová (2011), who demonstrated that valency status of adverbials (obligatory versus non-obligatory) may have an in-

(4) **Lexicalizations**[13]

a. Eva dala (ten drahý) kabát (, čo jej manžel kúpil vo Viedni,) do (profesionálnej) čistiarne (, ktorú jej odporučila suseda,)

'Eve brought the (that expensive) coat (that her husband had bought her in Vienna) to the (professional) dry cleaner (that her neighbor had recommended)'

b. Sestrička odoslala (rôzne) vzorky (, ktoré deň predtým odobrala pani Novákovej,) na (rozsiahlu) kontrolu (, ktorú si vyžiadal jej neurológ,)

'The nurse sent (various) samples (that she had taken from Ms. Nováková the day before) to an (extensive) examination (that her neurologist had asked for)'

c. Hostia položili (svoj veľký) kufor (, ktorý s námahou vynášali po schodoch,) pod (biely) stôl (, čo stojí pri okne v kuchyni,)

'The guests put the (their big) suitcase (that they had laboriously carried up the stairs) under the (that white) table (that is next to the window in the kitchen)'

fluence on word order (with non-obligatory adverbials preceding patients and obligatory adverbials following patients). However, the investigation was focused mostly on manner adverbials and lightness or heaviness of valency actants, and adjuncts were not tested as possible variables affecting word order. In another study, Rysová and Mírovský (2014) focused on testing valency as a word order factor in Czech by studying the results of a series of corpus queries and the decision trees algorithm. The main aim of the study was to test the hypothesis that obligatory adverbials (in terms of valency) follow the non-obligatory (i.e., optional) ones in the surface word order. For the 10-fold cross-validation in the decision trees experiments, various sets of features were tested, including form and length of one of the governed nodes and its subtree (e.g., length in words, length in characters, verbal modality, dependent clause). This study showed that valency (obligatoriness of sentence members) is not a strong factor influencing surface word order in Czech. However, it was also shown that there are several features in their feature selection that influence word order of contextually non-bound free modifiers of a verb in Czech, one of them being the length of verb modifiers (Rysová and Mírovský 2014: 979). In the experiment presented in our study, both obligatory and non-obligatory adverbial PPs were used in the lexicalizations. Hence, further investigations on the possible influence of valency status on the preferential word order is necessary to satisfactorily settle the issue.

[13] Note that the order of NP and PP of these lexicalizations (4a–h) was varied in the actual questionnaire, such that *all possible combinations* (i.e., NP-PP *and* PP-NP) were tested. They are shown here with NP preceding PP simply for ease of exposition.

(4) d. Marek vybral (ten vlnený) sveter (, čo má červené a modré pruhy,) z (mokrej) tašky (, ktorú si celú oblial kávou,)

'Mark took the (that woolen) sweater (that had red and blue stripes) from the (large) bag (all over which he had poured coffee)'

e. Učiteľka vpustila (malých) školákov (, ktorí sa práve vrátili z prestávky,) do (školskej) jedálne (, ktorá príjemne voňala od palaciniek,)

'The teacher let the (little) schoolkids (who just came back from break) into the (school) canteen (that smelled pleasantly like pancakes)'

f. Snúbenci pozvali (tú milú) tetu (, ktorá už dlhé roky žije v zahraničí,) na (letnú) svadbu (, ktorá sa mala konať v júli,)

'The bridal pair invited the (that nice) aunt (who had lived many years abroad) to [their] (summer) wedding (that was going to take place in July)'

g. Janka vložila (svoju novú) knihu (, ktorú jej daroval strýko z Rakúska,) medzi (tie) ostatné (, ktoré už mala odložené na poličke,)

'Joanne put the (her new) book (that her uncle from Austria had given her) between (those) others (that she had already put on the shelf)'

h. Katka dostala (ten veľký) darček (, ktorý bol zabalený v zlatom papieri,) od (najlepších) kamarátok (, ktoré pozvala na svoje narodeniny,)

'Kate got a (that big) present (that was wrapped in golden paper) from her (best) friends (that she had invited to her birthday)'

These eight lexicalizations were then used to create all four factor combinations twice, resulting in 8 × 4 = 32 tokens, which were then divided into four material sets of eight tokens each using the Latin squares method (Hoffmann 2011: 29, cf. also Keller 2000: 60; Keller and Alexopoulou 2005: 1121). The aim was to present every factor combination to every participant twice, but without using the same lexicalization more than once in every material set. Thus, material set 1 appeared as follows (5a–h):

(5) **Test items, material set 1**[14]

a. Eva dala kabát do profesionálnej čistiarne, ktorú jej odporučila suseda.
(LIGHT NP — HEAVY PP)

'Eve brought the coat to the professional dry cleaner that her neighbor had recommended.'

b. Sestrička odoslala rôzne vzorky, ktoré deň predtým odobrala pani Novákovej, na kontrolu.
(HEAVY NP — LIGHT PP)

'The nurse sent various samples, which she had taken from Ms. Nováková the day before, to an examination.'

c. Hostia položili pod stôl svoj veľký kufor, ktorý s námahou vynášali po schodoch.
(LIGHT PP — HEAVY NP)

'The guests put under the table their large suitcase, which they struggled to carry up the stairs.'

d. Marek vybral z mokrej tašky, ktorú si celú oblial kávou, sveter.
(HEAVY PP — LIGHT NP)

'Mark took from the large bag, over which he had poured coffee, the sweater.'

e. Učiteľka vpustila školákov do školskej jedálne, ktorá príjemne voňala od palaciniek.
(LIGHT NP — HEAVY PP)

'The teacher let the schoolkids into the school canteen, which pleasantly smelled of pancakes.'

f. Snúbenci pozvali tú milú tetu, ktorá už dlhé roky žije v zahraničí, na svadbu.
(HEAVY NP — LIGHT PP)

'The bridal pair invited that nice aunt, who had lived many years abroad, to the wedding.'

[14] Translations, which were not part of the actual experiment, have been added. The translations may partially deviate from canonical English word order to better match the original and partly to serve as word-for-word glosses as well.

(5) g. Janka vložila medzi ostatné svoju novú knihu, ktorú jej daroval strýko z Rakúska.
(LIGHT PP — HEAVY NP)

'Jana put between those others her new book, which her uncle from Austria had given her.'

h. Katka dostala od najlepších kamarátok, ktoré pozvala na svoje narodeniny, darček.
(HEAVY PP — LIGHT NP)

'Kate got from her best friends, whom she had invited to her birthday party, a present.'

Following Hoffmann (2013: 108), as many material sets as conditions (i.e., four) were created using the Latin squares method. Sixteen filler items were then created to achieve a filler:stimuli ratio of 2:1, following Cowart's suggestion that "[i]n general, there should be at least twice as many filler sentences as there are [test items]" (Cowart 1997: 92). Half of the filler items (i.e., eight) were grammatical and half were ungrammatical. The latter featured two types of violations,[15] subject-verb agreement (e.g., 6a) and word order (6b):

(6) **Examples of ungrammatical filler items**

a. **Rodičia** sa ma **spýtalo**, či chcem ísť s nimi na dovolenku alebo zostať doma.

'(My) parents asked me if I wanted to join them for a holiday trip or rather stay home.'

b. Adam **ponáhľal sa**, lebo musel stihnúť vlak do Viedne na pracovný pohovor.

'Adam was in a hurry because he had to catch a train to Vienna for a job interview.'

The 16 fillers were added to the eight test items, resulting in a set of 24 tokens. Four material sets, each consisting of 24 tokens, were created such that each material set featured the four possible factor combinations with different lexicalizations and in different places in the questionnaire. The tokens were

[15] Subject-verb agreement violations featured subjects with plural markers and past-tense verbs with singular neuter markers, e.g., in (6a) the subject is *rodičia* 'parents', i.e., plural and masculine animate, whereas the past-tense verb (*sa*) *spýtalo* has a singular neuter marker (*-o*) instead of the plural masculine animate marker (*-i*).

Word order violations featured clitics in third position when they should have been in second, e.g., in (6b) the reflexive pronoun *sa* is in third position when it should be in second position (i.e., following the subject *Adam*).

randomized[16] using blocking, "a controlled randomization procedure" where "each experimental sentence [is assigned] to a block so that only one sentence of each type appears in each block" (Cowart 1997: 94).

The questionnaires, which were printed on ten paper pages, also featured two training sessions to familiarize subjects with the MET method. The questionnaires were filled out under "carefully constructed experimental settings" (Hoffmann 2013: 99) to make sure that the data obtained was valid, objective, and reliable. To this end, they were personally administered by one of the authors at Comenius University in Bratislava and the University of Ss. Cyril and Methodius in Trnava. All subjects were recruited with the help of colleagues, who also organized suitable rooms for carrying out the METs so a comparable setting could be ensured for all subjects. In line with general ethics guidelines, all questionnaires were fully anonymous, only asking subjects to provide metadata such as their age, gender, country, and mother tongue(s), but never their name. Before filling out the questionnaires, subjects were informed about the time frame (about 45 minutes) and the three phases of the experiment (a non-linguistic training phase, a linguistic training phase, and the final experiment). They were also informed that there were no risks for them and that they would remain completely anonymous, and they were informed about their rights, including their right to withdrawal from participation without providing a reason at any time and their right to omit or refuse to answer or respond to any question asked of them. Finally, they were provided with one of the authors' email addresses in case of questions about the study. All of this information was printed on an information sheet that subjects were allowed to take home.

## 3. Results

A total of 39 L1 Slovak speakers (30 f, 7 m, 2 non-binary) were recruited to fill out the MET questionnaires: 29 students from three different groups from Comenius University in Bratislava (COM) and 10 from the University of Ss. Cyril and Methodius (UCM) in Trnava. The 39 subjects were all university students with an age range of 20–26.

Table 1 on the following page provides an overview of how many subjects from each group filled out which material sets (MS). It shows that at least nine subjects filled out each material set.

[16] The following restrictions were implemented when randomizing the items: The first and last three items were fillers, the first item was a grammatical filler, and test items were never adjacent to each other, i.e., always separated by a filler item.

**Table 1.** Overview of subject groups and material sets

| | MS 1 | MS 2 | MS 3 | MS 4 | Totals |
|---|---|---|---|---|---|
| **COM 1** | 3 | 3 | 3 | 2 | 11 |
| **COM 2** | 2 | 2 | 1 | 1 | 6 |
| **COM 3** | 2 | 3 | 3 | 4 | 12 |
| **UCM** | 3 | 3 | 2 | 2 | 10 |
| **Totals** | 10 | 11 | 9 | 9 | 39 |

All questionnaires were scanned and the ratings entered into a spreadsheet,[17] where *z*-scores[18] were calculated for further analysis. The *z*-score mean of one particular LEXICALIZATION (0.419) turned out to deviate considerably from the other seven (range: −0.235 to 0.213), so it was excluded from further analysis. This was deemed unproblematic, since every condition was tested with two LEXICALIZATIONS per material set.

Mixed-effects modeling was employed to test for significant random and fixed effects. To this end, the `lmer()` function from the *lme4* package (Bates et al. 2015, 2020) for *R* for Windows was employed. Specifically, the `step()` function from the *lmerTest* package for *R* (Kuznetsova et al. 2015, 2020) was used. This is a function that "automates the step-down approach" used in stepwise regression (Kuznetsova et al. 2015: 8) using the `lmer()` function from *lme4*. Table 2 on the following page provides an overview of the tested fixed and random effects and their levels.

[17] All results, as well as scans of the questionnaires, have been made available on an OSF repository that is accessible via the following link: https://osf.io/kvrwh/.

[18] It is acknowledged that an anonymous reviewer has noted that "Z-score transformation relies on data being normally distributed, which they rarely are in judgment data". However, as noted by Cooksey, "transforming raw scores into *z*-scores does not change the shape of a distribution or rank ordering of individuals within that distribution" (Cooksey 2020: 126). In fact, all it does is make the data comparable so that it can be fed into the mixed-effects models. Furthermore, as an inspection of the distribution of our data revealed, the reviewer's criticism is only partially applicable, anyway: As the histogram in the Appendix shows, while the distribution cannot be described as prototypically Gaussian, neither is it completely random (as noted by Hoffmann, p.c. in August 2024).

**Table 2.** Overview of fixed and random effects, with levels

| Fixed effects | Random effects |
|---|---|
| WEIGHT ORDER | SUBJECT (001, 002, 003, … 039) |
| (LIGHT_HEAVY, HEAVY_LIGHT) | GENDER (MALE, FEMALE, NON-BINARY) |
| PHRASE ORDER (NP_PP, PP_NP) | GROUP (COM1, COM2, COM3, UCM) |
| | LEXICALIZATION[19] (A, B, C, D, F, G, H) |
| | MATERIAL SET (MS1, MS2, MS3, MS4) |

The full model subjected to stepwise regression using *lmerTest* was thus as follows:

Z_SCORE ~ WEIGHT_ORDER * PHRASE_ORDER + (1|MS)
+ (1|SUBJECT) + (1|GENDER) + (1|LEXICALIZATION) + (1|GROUP)

Stepwise regression showed that none of the random effects were significant. This suggests that the experiment design and implementation were successful in minimizing the influence of extra-linguistic factors and testing the desired linguistic variables only. Table 3 shows the results and the order of elimination of random effects from the model.

**Table 3.** Likelihood ratio tests for random effects and their order of elimination

| | $\chi 2$ | Elimination no. | *p*-value |
|---|---|---|---|
| GENDER | 0.00 | 1 | 1.000[20] n.s. |
| SUBJECT | 0.10 | 2 | 0.751 n.s. |
| MS | 0.23 | 3 | 0.633 n.s. |
| GROUP | 0.37 | 4 | 0.541 n.s. |
| LEXICALIZATION | 0.63 | 5 | 0.427 n.s. |

[19] Lexicalization E was excluded from analysis because it deviated considerably from the others (see above).

[20] The model estimated a standard deviation of 0 for the random effect GENDER, which resulted in *lme4* returning the message `boundary (singular fit)` and a *p*-value of

**Table 4.** F-tests for the fixed effects, and their order of elimination

| | *F* | Elimination no. | *p*-value |
|---|---|---|---|
| WEIGHT ORDER: PHRASE ORDER | 0.142 | 1 | 0.706 n.s. |
| WEIGHT ORDER | 35.70 | kept | < 0.001 *** |
| PHRASE ORDER | 5.90 | kept | 0.016 * |

As Table 4 shows, both fixed effects turned out to be significant. WEIGHT ORDER was very significant ($p < 0.001$***); PHRASE ORDER was significant ($p < 0.05$*). The interaction WEIGHT ORDER:PHRASE ORDER was not significant ($p = 0.70645$ n.s.). Thus, the final model determined by stepwise regression using *lmerTest* was as follows:

```
z_score ~ weight_order + phrase_order
```

In other words, both linguistic variables significantly influenced subjects' ratings, with WEIGHT ORDER having a more profound impact than PHRASE ORDER. The implications are discussed in more detail in §4 below.

Figure 1 on the following page plots the *z*-score means of each of the four test items for a visual inspection of the results. Table 5 on the following page provides the *z*-score means as well as standard deviation (SD) and standard error (SE) values. The grammatical filler *z*-score mean was 0.60 (SD 0.72; SE 0.04); for the ungrammatical fillers, this value was −0.62 (SD 0.95; SE 0.05). These values are plotted as "baselines" against which test items can be judged.

As expected (cf. hypotheses (i) and (ii)), the best-rated condition was LIGHT NP — HEAVY PP (*z*-score mean 0.428), i.e., those test items where a light constituent preceded a heavy constituent and where an NP preceded a PP. Conversely, HEAVY PP — LIGHT NP was rated worst (*z*-score mean −0.333). Taking a closer look at WEIGHT ORDER, the plot shows that LIGHT before HEAVY was always preferred by the subjects, regardless of PHRASE ORDER. Recall that this was a very significant effect in the mixed-effects model ($p < 0.001$), indicating that subjects were particularly sensitive to WEIGHT ORDER. Moreover, as Table 5

1.000. This is an indication that the slope "could not be estimated" by the model (Winter 2020: 259)—that is, "the level of 'between-group' variabilityis not sufficient to warrant incorporating random effects in the model" (Bates 2010: 10). In other words, GENDER had a negligible influence on the outcome of the experiment and was therefore removed from the model before proceeding with stepwise regression (as suggested by Bates 2010: 10; cf. also Pasch et al. 2013 and Barr et al. 2013).

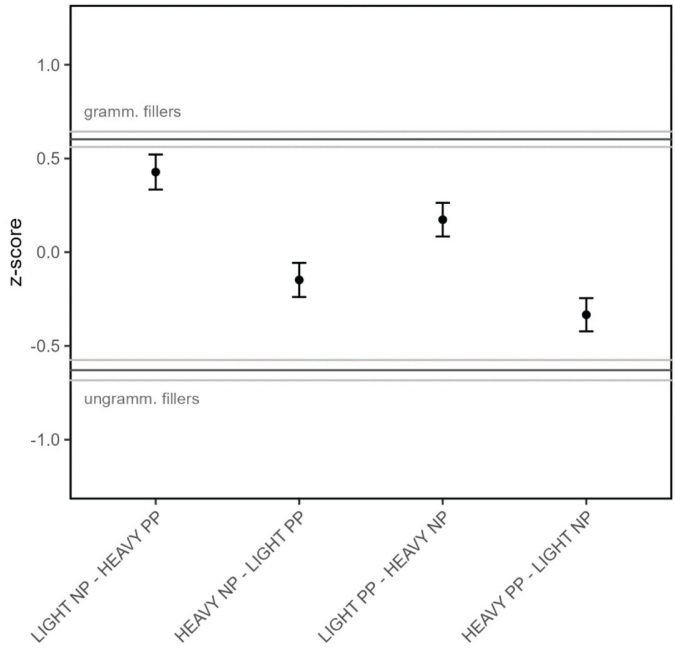


**Figure 1.** *Z*-score means of WEIGHT ORDER : PHRASE ORDER (error bars indicate standard errors)

**Table 5.** *Z*-score means of WEIGHT ORDER : PHRASE ORDER

| WEIGHT ORDER | PHRASE ORDER | *z*-score | SD | SE |
|---|---|---|---|---|
| LIGHT_HEAVY | NP_PP | 0.428 | 0.770 | 0.093 |
| HEAVY_LIGHT | NP_PP | − 0.148 | 0.756 | 0.091 |
| LIGHT_HEAVY | PP_NP | 0.173 | 0.748 | 0.090 |
| HEAVY_LIGHT | PP_NP | − 0.333 | 0.724 | 0.088 |

shows, only LIGHT before HEAVY had positive *z*-scores, whereas HEAVY before LIGHT received negative *z*-scores. This confirms hypothesis (i) that weight effects affect constituent order in Slovak. Concerning PHRASE ORDER, the plot shows that NP before PP was always preferred, confirming hypothesis (ii), i.e.,

the existence of a "basic" $_{VP}$[NP PP] pattern that has been identified in English. However, note that this effect was less significant ($p < 0.05$*) than WEIGHT ORDER, indicating that subjects were less sensitive to it.

In summary, mixed-effects modeling and the plots have confirmed both hypotheses outlined in §1, that is, LIGHT before HEAVY constituents are always preferred. The same applies to NP before PP. Also, the results show that WEIGHT ORDER plays a considerably more important role than PHRASE ORDER in Slovak.

## 4. Discussion

The data presented here are the result of the first-ever investigation of weight effects in Slovak using an experimental approach, and the first using the MET method to investigate weight effects in any language. We have been able to answer both research questions, demonstrating that weight effects do affect Slovak constituent order and that there is evidence for the existence of a "basic" $_{VP}$[NP PP] pattern. These findings are in line with claims made in the literature upon which our hypotheses were based. Firstly, regarding hypothesis (i), Goldberg's Tenet #5 posits that "[c]ross-linguistic generalizations are explained by appeal to general cognitive constraints" (Goldberg 2003: 219). Weight effects are the result of processing efficiency (Hawkins 1994), which is a general cognitive constraint in this sense. Therefore, it is unsurprising that they affect word order in Slovak in the same way as in English, that is, producing a preference for placing light before heavy constituents. This effect was shown to be very significant ($p < 0.001$***) in the data that we collected from L1 Slovak speakers.

Secondly, regarding hypothesis (ii), Hawkins's (2004) PGCH predicts that grammaticalized patterns in analytic languages are the result of performance preference and will mirror those that are preferred in synthetic languages. Our results confirm this prediction, as evidenced by the preference for the NP-PP phrase order:[21] All other things being equal (i.e., weight order), subjects always preferred NPs preceding PPs. This effect was significant ($p < 0.05$*), indicating that phrase order plays a less important role than weight order. An

[21] We are grateful to an anonymous reviewer who pointed out the experiment in Šimík and Wierzba (2017), which "involved the same type of comparison; the authors found a numerical, but not statistically significant, preference for the NP-PP order in Slovak, Czech, and Polish". As the anonymous reviewer observed, it should be noted "that Šimík [and] Wierzba's experiment had more power (many more lexicalizations and also more participants)" and that, "[g]iven that the phrase order effect is not as large (as compared to the weight order effect), it's good to be slightly more cautious about the result" of the present study. Furthermore, "it cannot be ruled out that the NP-PP preference reduces to [sic] the weight effect. After all, PPs have at least two words (and nodes, in most syntactic frameworks), while NPs can easily have just one".

interaction of weight order and phrase order could not be detected, suggesting that they are independent of each other.

## 5. Conclusion

The results presented here are part of a pilot study. As such, they are to be considered a first, rather rudimentary, and therefore anything but conclusive attempt at determining whether weight effects play a role at all in Slovak word order. However, our first results do indicate that this is the case. Moreover, we have found evidence for the existence of a "basic" NP-PP phrase order in the Slovak VP. Both of these findings address a significant gap in research on the effects on word order beyond information structure. This applies to Slovak in particular and the Slavic languages in general, where the principle of end-weight and processing efficiency have received little to no attention so far; cf., e.g., Kizach's remarks on Russian (Kizach 2012: 251). Given our findings, however, we believe that weight effects and basic phrasal patterns cannot be ignored in research on word order in Slavic languages. The fact that both weight order and phrase order turned out to be statistically significant effects in our study has serious implications for approaches that disregard weight effects or consider them secondary to information structure. Certainly, future editions of Slovak grammars, which have so far all but ignored weight effects, should at the very least briefly mention the principle of end-weight. They should also discuss the idea of basic or grammaticalized phrasal patterns that arise as a result of performance preference (cf. Hawkins's 2004 PGCH), addressing a serious gap in research that Fried (2017: 241) has described as follows: "syntactic patterning […] has been generally left just about untouched in Slavic linguistics".

As a pilot study, our data is anything but comprehensive. For example, we tested only four out of eight possible conditions (and not, e.g., LIGHT NP — LIGHT PP or HEAVY PP — LIGHT PP). From a methodological point of view, we acknowledge that the experiment design leaves room for improvement. For example, authentic sentences from, e.g., corpora can be used to create test items (see fn 7). We also did not take into consideration so-called "wrap-up" effects[22] that may affect acceptability ratings of sentences. Furthermore, the phrases we used as test items varied in length (measured as number of words), if only slightly; verbs with different valency were used, and some of the relative clauses also differed with respect to perspective or subject maintenance

[22] That is, participants' decisions about sentence acceptability/grammaticality may be different for words toward the end of sentences compared to the preceding words because their decisions become more fixed at later points. For this reason, it has been suggested that test items be constructed in such a way that they are followed by a coordinated main clause to reduce such wrap-up effects (e.g., Balling and Kizach 2017).

and restrictiveness. All of these factors might have skewed the participants' acceptability ratings, such that in future, more extensive investigations will be necessary to examine the relevance of these factors for acceptability ratings. Due to time constraints, however, none of these issues could be addressed in this pilot study.

Nevertheless our—admittedly very preliminary—results do demonstrate that there is a need for more investigations on factors influencing word order in Slovak (and Slavic languages, for that matter) beyond information structure. Such studies should be based on more data and more types of data, and they should take into account more linguistic variables. For example, corpus data can be a valuable addition to the experimental data presented. Syntactic patterns consisting of a verb followed by an NP and a PP in either order can be extracted from corpora that have been annotated using Universal Dependencies, and frequencies of each order can be determined. Following Kizach's (2012: 270) findings that there was a "correlation between the most efficient orders and the most frequent orders" in his corpus study using data from the Russian National Corpus, it is plausible that Slovak corpus data will show similar effects. That is, a higher frequency of corpus tokens with light before heavy constituents, as well as a higher frequency of the $_{VP}$[NP PP] pattern over $_{VP}$[PP NP]. Corpus data can also help shed more light on the effects of weight difference between NP and PP on the likelihood of HNPS occurring (see discussion in §1), and ceiling effects in this regard (cf. Medeiros et al. 2021).

Moreover, previous research on English has shown that the argument structure properties of particular verbs may play a role. Stallings et al. (1998) claim in the context of their "verb disposition hypothesis" that verbs that can take sentential complements in addition to NPs, such as *find* and *reveal*, are more likely to be involved in HNPS than those that can only take NPs, such as *transfer* and *delay* (Stallings et al. 1998: 397). Such effects can also be investigated using corpus data, by looking at the likelihood of HNPS occurring with particular verbs. Also, as Kizach (2012: 270) has noted, "an interesting topic for future research would be to investigate how givenness (theme-rheme) and weight interact". For example, future investigations could incorporate the variable INFORMATION STRUCTURE, with the levels TOPICAL and FOCUSED, depending on the left context of a corpus token or the contents of a short text that subjects must read before filling out an MET questionnaire. This way, it can be determined which factor is more important when it comes to word order, weight effects or information structure.

Finally, extending research on word order in general and HNPS in particular to other languages appears warranted in light of our findings, as our results have confirmed the cross-linguistic predictions made by Goldberg's Tenet # 5 (Goldberg 2003: 219) and Hawkins's (2004) PGCH.

## References

Arnold, Jennifer, Anthony Losongeo, and Ryan Ginstrom. (2000) "Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering". *Language* 76(1): 28–55.

Balling, Laura W., and Johannes Kizach. (2017) "Effects of surprisal and locality on Danish sentence processing: An eye-tracking investigation". *Journal of psycholinguistic research* 46(5): 1119–36.

Bard, Ellen G., Dan Robertson, and Antonella Sorace. (1996) "Magnitude estimation of linguistic acceptability". *Language* 72(1): 32–68.

Barr, Dale J., Roger Levy, Christopher Scheepers, and Harry J. Tily. (2013) "Random effects structure for confirmatory hypothesis testing: Keep it maximal". *Journal of memory and language* 68(3): 255–78.

Bates, Douglas. (2010) *lme4: Mixed-effects modeling with R*. New York: Springer.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. (2015) "Fitting linear mixed-effects models using lme4". *Journal of statistical software* 67(1): 1–48.

————. (2020) Reference manual for Package "lme4". Available at: http://cran.r-project.org/web/packages/lme4/lme4.pdf.

Behaghel, Otto. (1909) "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern". *Indogermanische Forschungen* 25: 110–42.

————. (1930) "Von deutscher Wortstellung". *Zeitschrift für Deutschkunde*. Vol. 44 of *Zeitschrift für den deutschen Unterricht*. Leipzig: B. G. Teubner, 82–89.

Cooksey, Ray W. (2020) "Descriptive statistics for summarising data". *Illustrating statistical procedures: Finding meaning in quantitative data*. Singapore: Springer, 61–139.

Cowart, Wayne. (1997) *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, CA: Sage.

Divjak, Dagmar. (2017) "The role of lexical frequency in the acceptability of syntactic variants: Evidence from *that*-clauses in Polish". *Cognitive science* 41(2): 354–82.

Fanselow, Gisbert. (2007) "Carrots—perfect as vegetables, but please not as a main dish". *Theoretical linguistics* 33(3): 353–67.

Featherston, Sam. (2007a) "Data in generative grammar: The stick and the carrot". *Theoretical linguistics* 33(3): 269–318.

————. (2007b) "Reply". *Theoretical linguistics* 33(3): 319–33.

Fried, Mirjam. (2017) "Construction grammar in the service of Slavic linguistics, and vice versa". *Journal of Slavic linguistics* 25(2): 241–76.

Fukuda, Shin, Grant Goodall, Dan Michel, and Henry Beecher. (2012) "Is magnitude estimation worth the trouble?" Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman, eds. *Proceedings of the 29th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 328–36.

Gibson, Edward. (1998) "Linguistic complexity: Locality of syntactic dependencies". *Cognition* 68(1): 1–76.

————. (2000) "The dependency locality theory: A distance-based theory of linguistic complexity". Alec P. Marantz, Yasushi Miyashita, and Wayne O'Neil, eds. *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press, 94–126.

Gibson, Edward, and Iris H. H. Wu. (2013) "Processing Chinese relative clauses in context". *Language and cognitive processes* 28(1–2): 125–55.

Goldberg, Adele E. (2003) "Constructions: A new theoretical approach to language". *TRENDS in cognitive sciences* 7(5): 219–24.

Haegeman, Liliane. (1994) *Introduction to government & binding theory*. 2nd ed. Oxford: Blackwell.

Haegeman, Liliane, and Jacqueline Guéron. (1999) *English grammar—a generative perspective*. Oxford: Blackwell.

Hale, John T. (2001) "A probabilistic Earley parser as a psycholinguistic model". *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, vol. 2. Pittsburgh, PA: Association for Computational Linguistics, 159–66.

Hawkins, John A. (1994) *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press. [Cambridge Studies in Linguistics 73.]

————. (2004) *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

————. (2014) *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Hoffmann, Thomas. (2011) *Preposition placement in English: A usage-based approach*. Cambridge, UK: Cambridge University Press. [Studies in English Language.]

————. (2013) "Obtaining introspective acceptability judgements". Manfred Krug and Julia Schlüter, eds. *Research methods in language variation and change*. Cambridge, UK: Cambridge University Press, 99–118.

Horsch, Jakob. (2020) "Vidíš tým lepšie, čím bližšie sa pozeráš: Slovak comparative correlative CC' constructions from a construction grammar perspective". *Jazykovedný Časopis* 71(1): 25–40.

————. (2021) "Slovak comparative correlatives: A usage-based construction grammar account". *Constructions and frames* 13(2): 193–229.

Horsch, Jakob. (2023a) *The comparative correlative construction in World Englishes: A usage-based construction grammar approach*. PhD thesis, University of Eichstätt-Ingolstadt, Eichstätt, Germany. Available at: https://opus4.kobv.de/opus4-ku-eichstaett/frontdoor/index/index/docId/833.

————. (2023b) "Corroborating corpus data with elicited introspection data: A case study". *Jazykovedný Časopis* 74(2): 60–69.

Ivanová, Martina. (2016) *Syntax slovenského jazyka*. 2nd ed. Prešov, Slovakia: University of Prešov.

Keller, Frank. (2000) *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. PhD thesis, University of Edinburgh.

Keller, Frank, and Theodora Alexopoulou. (2005) "A crosslinguistic, experimental study of resumptive pronouns and that-trace effects". Bruno G. Bara, Lawrence Barsalou, and Monica Bucciarelli, eds. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. London: Psychology Press, 1120–25. Available at : https://escholarship.org/uc/item/6xg3x3hx.

Kizach, Johannes. (2012) "Evidence for weight effects in Russian". *Russian linguistics* 36(3): 251–70.

————. (2014) "Complexity, givenness and g-maze". Abstract of talk at the LinG/RTG2636 Colloquium, University of Göttingen. Available at: https://www.uni-goettingen.de/en/487714.html. Last accessed 25 October 2023.

Konieczny, Lars. (2000) "Locality and parsing complexity". *Journal of psycholinguistic research* 29(6): 627–45.

Konieczny, Lars, and Philipp Döring. (2003) "Anticipation of clause-final heads: Evidence from eye-tracking and SRNs". Peter P. Slezak, ed. *Proceedings of the 4th International Conference on Cognitive Science*, 13–17 July 2003. Sydney: University of New South Wales, 330–35.

Krifka, Manfred. (2008) "Basic notions of information structure". *Acta linguistica Hungarica* 55(3–4): 243–76.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. (2020) lmerTest: Tests for random and fixed effects for linear mixed effects models (lmer objects of lme4 package). R package version 3.1-3. Available at: https://CRAN.R-project.org/package=lmerTest.

Kuznetsova, Alexandra, Rune H. B. Christensen, Cecile Bavay, and Per B. Brockhoff. (2015) "Automated mixed ANOVA modeling of sensory and consumer data". *Food quality and preference* 40: 31–38.

Lambrecht, Knud. (1994) *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge, UK: Cambridge University Press. [Cambridge Studies in Linguistics 71.]

Lau, Jey Han, Alexander Clark, and Shalom Lappin. (2017) "Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge". *Cognitive science*, 41(5): 1202–41.

Levy, Roger. (2008) "Expectation-based syntactic comprehension". *Cognition* 106: 1126–77.

Mains, Paul, Kevin McGowan, and David J. Medeiros. (2015) "Gradient acceptability by length in heavy NP shift". Presented at the 89th Annual Meeting of the Linguistic Society of America, Portland, OR.

Medeiros, David, Paul Mains, and Kevin B. McGowan. (2021) "Ceiling effects on weight in heavy NP shift". *Linguistic inquiry* 52(2): 426–40.

Melnick, Robin Jeffrey. (2017) *Consistency in variation: On the provenance of end-weight*. PhD thesis, Stanford University.

Mistrík, Jozef. (1983) "Moderná slovenčina. 2. vyd". Bratislava: Slovenské pedagogické nakladateľstvo.

————. (1988) *A grammar of contemporary Slovak*. Bratislava: Slovenské pedagogické nakladateľstvo.

————. (2003) *Gramatika slovenčiny*. Bratislava: Slovenské pedagogické nakladateľstvo.

Newmeyer, Frederick J. (2007) "Commentary on Sam Featherston, 'Data in generative grammar: The stick and the carrot'". *Theoretical linguistics* 33(3): 395–99.

Orlovský, Jozef. (1971) *Slovenská syntax*. Bratislava: Obzor.

Pasch, Bret, Benjamin M. Bolker, and Steven M. Phelps. (2013) "Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice". *The American naturalist* 182(5): E161–E173.

Pauliny, Eugen. (1997) *Krátka gramatika slovenská*. Bratislava: Národné literárne centrum—Dom slovenskej literatúry.

————. (1981) *Slovenská gramatika (Opis jazykového systému)*. Bratislava: Slovenské pedagogické nakladateľstvo.

Pauliny, Eugen, Jozef Ružička, and Jozef Štolc. (1963) *Slovenská gramatika*. Bratislava: Slovenské pedagogické nakladateľstvo.

Pavlovič, Jozef. (2012) *Syntax slovenského jazyka I*. Available at: https://pdfweb.truni.sk/download?e-skripta/syntax-1.pdf.

Quirk, Randolph, Sidney Greenbaum, and Geoffrey Leech. (1972) *A grammar of contemporary English*. London: Longman.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. (1985) *A comprehensive grammar of the English language*. New York: Longman.

Rysová, Kateřina. (2011) "Ke slovosledu v konstrukcích s obligatorními participanty (se zaměřením na slovosled patientu a způsobového slovesného doplnění)". Vladimír Petkevič and Alexandr Rosen, eds. *Korpusová lingvistika Praha 2011—Gramatika a značkování korpusů*. Prague: Nakladatelství Lidové noviny, 62–69.

Rysová, Kateřina, and Jiří Mírovský. (2014) "Valency and word order in Czech—A corpus probe". Nicoletta Calzolari, Khaled Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, eds. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík, Iceland: European Language Resources Association, 975–80.

Schütze, Carson T., and Jon Sprouse. (2013) "Judgment data". Robert J. Podesva and Devyani Sharma, eds. *Research methods in linguistics*. Cambridge, UK: Cambridge University Press, 27–50.

Sgall, Peter, Eva Hajičová, and Eva Buráňová. (1980) *Aktuální členění věty v češtině*. Prague: Academia.

Short, David. (2002) "Slovak". Bernard Comrie and Greville G. Corbett, eds. *The Slavonic languages*. London: Routledge, 533–92. [Routledge Language Family Descriptions.]

Siewierska, Anna, and Ludmila Uhlířová. (1998) "An overview of word order in Slavic languages". Anna Siewierska, ed. *Constituent order in the languages of Europe*, vol. 1. Berlin, New York: De Gruyter Mouton, 105–50. [Empirical Approaches to Language Typology.]

Smith, Nathaniel J., and Roger Levy. (2008) "Optimal processing times in reading: A formal model and empirical investigation". *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* 30(30): 595–600.

————. (2013) "The effect of word predictability on reading time is logarithmic". *Cognition* 128(3): 302–19.

Sprouse, John. (2011) "A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments". *Language* 87(2): 274–88.

Stallings, Lynne M., and Maryellen C. MacDonald. (2011) "It's not just the 'heavy NP': Relative phrase length modulates the production of heavy-NP shift". *Journal of psycholinguistic research* 40(3): 177–87.

Stallings, Lynne M., Maryellen C. MacDonald, and Padraig G. O'Seaghdha. (1998) "Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift". *Journal of memory and language* 39(3): 392–417.

Szmrecsányi, Benedikt. (2004) "On operationalizing syntactic complexity". Gérard Purnelle, Cédrick Fairon, and Anne Dister, eds. *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis*. Louvain-la-Neuve, Belgium: Université Catholique de Belgique (UCLouvain), 1032–39.

Šimík, Radek. (2021) *Information structure and its relation to syntax, semantics, and reference*. Unpublished manuscript. Available at: https://lingbuzz.net/005822.

Šimík, Radek, Marta Wierzba, and Beste Kamali. (2014) "Givenness and the position of the direct object in the Czech clause". Cassandra Chapman, Olena Kit, and Ivona Kucerová, eds. *Proceedings of Formal Approaches to Slavic Linguistics (FASL) 22: The McMaster Meeting 2013*. Ann Arbor, MI: Michigan Slavic Publications, 302–18.

Šimík, Radek, and Marta Wierzba. (2017) "Expression of information structure in West Slavic: Modeling the impact of prosodic and word-order factors". *Language* 93(3): 671–709.

Verhagen, Véronique, Maria Mos, Joost Schilperoord, and Ad Backus. (2020) "Variation is information: Analyses of variation across items, participants, time, and methods in metalinguistic judgment data". *Linguistics* 58(1): 37–81. DOI 10.1515/ling-2018-0036.

Wallbridge, Sarenne, Peter Bell, and Catherine Lai. (2022) "Investigating perception of spoken dialogue acceptability through surprisal". *Interspeech 2022: The 23rd Annual Conference of the International Speech Communication Association*. Red Hook, NY: Curran Associates, 4506–10.

Wasow, Thomas. (1997) "Remarks on grammatical weight". *Language variation and change* 9(1): 81–105.

————. (2002) *Postverbal behavior*. Stanford, CA: Center for the Study of Language and Information.

Wasow, Thomas, and Jennifer Arnold. (2003) "Post-verbal constituent ordering in English". *Determinants of grammatical variation in English*. Berlin: De Gruyter, 119–54.

Weskott, Thomas, and Gisbert Fanselow. (2011) "On the informativity of different measures of linguistic acceptability". *Language* 87(2): 249–73.

Winter, Bodo. (2020) *Statistics for linguists: An introduction using R*. New York/London: Routledge.

## Appendix

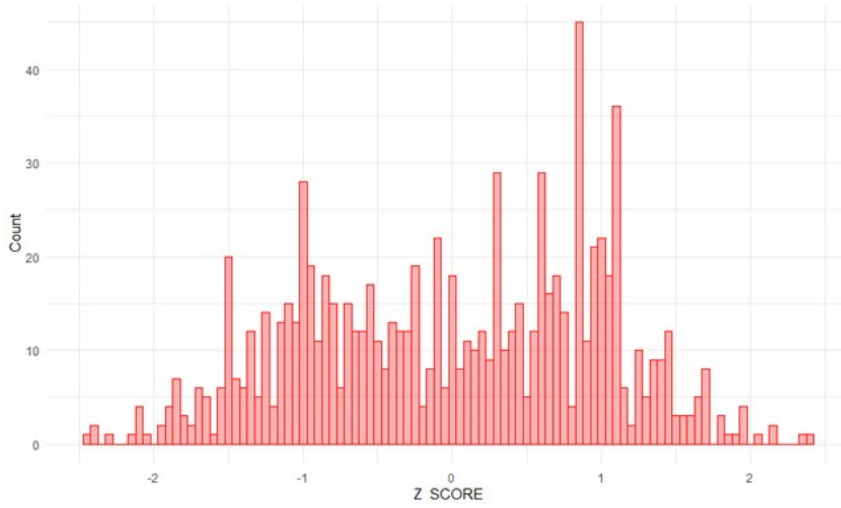


**Figure 2.** Distribution of *z*-scores in the MET data set

Jakob Horsch
Slovak National Corpus Department
Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences
Bratislava, Slovakia
jakob.horsch@korpus.sk

Martina Ivanová
Institute of Slovak and Media Studies
Faculty of Arts
University of Prešov
Prešov, Slovakia
and
Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences
Bratislava, Slovakia
martina.ivanova@unipo.sk

# Local and Non-Local Binding of Reflexives and Pronominals in Russian Object Control Infinitives

**Tatiana Perevozchikova**

*Abstract*: It has been suggested that reflexives and pronominals in Russian object control infinitives can refer either to the matrix subject (non-local binding) or to the infinitive subject (local binding), depending on the matrix verb. Specifically, it has been argued that the non-local binding of a reflexive and the local binding of a pronominal are most likely to occur when the matrix verb induces a strong cohesion between the matrix and the infinitive clause. The present study investigates Russian speakers' preferences in the interpretation of possessives in object control infinitives with the aim of testing the cohesion hypothesis by means of a referent selection task. The results show that the matrix verb does influence the interpretation of possessive reflexives and pronominals in object control infinitives, but not as predicted by the cohesion hypothesis. For a possessive reflexive, local binding is generally preferred, but non-local binding is also possible and most likely with a matrix verb that induces weak cohesion between the matrix and the infinitive clause. For a possessive pronominal, non-local binding is preferred both with a matrix verb that induces a strong cohesion and with a matrix verb that induces a weak cohesion, while local binding dominates with a matrix verb from the middle cohesion range. The study concludes that the cohesion between the matrix and the infinitive clause is not a relevant factor underlying the effect of the matrix verb in the interpretation of Russian possessives in object control infinitives. An alternative explanation in terms of implicit causality is proposed, which argues that the interpretation of reflexives is subject to a strong syntactic locality constraint, which can be weakened when pragmatic inferences from the basic cognitive representation of the event conditioned by a matrix verb make a non-local referent salient. The interpretation of pronominals is subject to a weak anti-locality constraint, which can be overridden when pragmatic inferences suggest a local referent as the more salient interpretation.

*Keywords:* pronominal, reflexive, possessive, reference, non-local binding, Russian, object control infinitive

## 1. Introduction

The binding domain of a reflexive pronoun, i.e., the domain in which the reflexive must find its antecedent, is usually the clause. However, reflexives can sometimes have an antecedent outside the clause in which they are contained,

resulting in non-local binding dependencies. In Russian, non-local binding is restricted to cases of medium-distance binding (Reuland and Koster 1991: 24) in infinitive complements introduced by an object control verb (Rappaport 1986), as illustrated by (1) and (2).

(1) Komendant$_1$ velel dvorniku$_2$ otnesti
commandant.NOM.SG.M order.PST.SG.M janitor.DAT.SG.M take.INF
veščі žilca k sebe$_{1,2}$.
thing.ACC.PL tenant.GEN.SG.M to REFL.DAT
'The commandant ordered the janitor to take the tenant's things to his place.' (Peškovskij 1956: 163)

(2) Professor$_1$ poprosil assistanta$_2$ pročitat'
professor.NOM.SG.M ask.PST.SG.M assistant.ACC.SG.M read.INF
svoj$_{1,2}$ referat.
POSS.REFL.ACC.SG.M talk.ACC.SG.M
'The professor asked the assistant to give his talk.' (Klenin 1974: 30)

The reference of the personal reflexive in (1) and the possessive reflexive in (2) has been considered ambiguous between the local antecedent, which is the subject of the infinitive, and the non-local antecedent, which is the subject of the matrix verb. In addition, it has been observed that pronominals used in an object control infinitive clause can also have both the local and the non-local antecedent (Bílý 1981; Kazenin 2000; Klenin 1974; Růžička 1973; Timberlake 1979; Yokoyama 1975).

Assuming this referential ambiguity, existing studies have focused either on describing the syntactic operations that enable the non-local binding of a reflexive (Dotlačil 2005; Kazenin 2000; Zubkov 2018) or on identifying the semantic-pragmatic factors that can influence the referent choice (Bílý 1981; Klenin 1974; Růžička 1973; Timberlake 1979, 2004: 249–52; Yokoyama 1975). Research from both perspectives suggests that the availability of the non-local binding for a reflexive and the local binding for a pronominal in object control infinitives varies according to the matrix verb and for individual speakers.

Two specific suggestions have been made. First, it has been argued that the likelihood of the non-local binding of a reflexive and the local binding of a pronominal in an infinitive clause is highest when the matrix verb induces a high degree of cohesion between the matrix clause and the infinitive clause (Timberlake 1979, 2004: 249–52). Second, it has been suggested that Russian speakers adhere to one of three "pronoun dialects" in the interpretation of pronominals and reflexives in object control infinitives: The first requires complementary distribution of reflexives and pronominals, the second allows free variation for both, and the third allows free variation only for reflexives

but not for pronominals (Klenin 1974: 40–41, 57). These claims have so far been based on the interpretation of a few examples by a few speakers and have not been subjected to systematic statistical analysis investigation.

The aim of this paper is to provide statistical generalizations about Russian speakers' preferences in the interpretation of a reflexive possessive (hereafter RP) and a non-reflexive possessive (hereafter NRP) in object control infinitives. Two main questions are addressed:

i. Is the interpretation of an RP and an NRP in object control infinitives constrained by the matrix verb in terms of cohesion between the matrix and the infinitive clause?
ii. Is there interspeaker variation in interpretation preferences as captured by the three pronoun dialects described by Klenin (1974)?

The study focused primarily on possessives but included a partial comparison between argumental and possessive pronouns as a secondary issue. In order to keep the number of test items manageable for the participants, the interpretation of argumental pronouns and reflexives was investigated in a smaller set of items without varying the matrix verb.

The study is based on data from a referent selection task conducted with 120 Russian speakers and subsequently analyzed using a generalized linear mixed-effects regression. As we will see, the results challenge Timberlake's hypothesis, showing that the matrix verb does influence the interpretation of an NRP and (to a lesser extent) an RP, but this influence is not conditioned by the cohesion between the matrix and the infinitive clause. Furthermore, it is shown that, contrary to Klenin's (1974) claim, there is no "pronoun dialect" in which the interpretation of a pronominal would be more restricted than the interpretation of a reflexive.

Before presenting the details of the study, I will give a brief overview of the Russian system of reflexives with a focus on non-local binding, then describe existing structural accounts of non-local binding in Russian and review semantic and pragmatic accounts centered on the role of the matrix verb. The theoretical section concludes with the formulation of hypotheses to be tested. In the next section, I will describe the methodology. The fourth section presents the data analysis, followed by a discussion of the results in section 5. The article concludes with an overall summary and directions for future work.

## 2. Theoretical Preliminaries

### 2.1. Basic Facts about Reflexives and Non-Local Binding in Russian

In Russian, there is an argumental reflexive *sebja* and a possessive reflexive *svoj*, as illustrated in (3).

(3) Oleg$_1$ vidit sebja$_1$ / svoë$_1$
Oleg.NOM.SG.M see.PRS.3SG REFL.ACC POSS.REFL.ACC.SG.N
otraženie v zerkale.
reflection.ACC.SG.N in mirror.LOC.SG.N
'Oleg sees himself / his reflection in the mirror.'

Both reflexives are morphologically simple and unspecified for person, number, and gender of their antecedent. The argumental reflexive has different case forms except for the nominative; the accusative *sebja* as in (3) is its citation form. The possessive reflexive *svoj* has an adjectival morphology, agreeing in case, number, and gender with the head of the NP it modifies.

The reference of both reflexives depends on another NP (their antecedent) in the sentence, which constitutes the case of binding. The binding behavior of Russian reflexives and pronominals is considered to be well captured by principles A and B of the Binding Theory as formulated by Chomsky (1981: 188). A reflexive takes its antecedent within a minimal clause, whereas a pronominal does not (Rappaport 1986; Kazenin 2000).

However, complications in the application of the canonical Binding Theory for Russian arise in infinitive complements introduced by an object control verb, as in (4) and (5).

(4) Olja$_1$ poprosila Tanju$_2$ nalit'
Olja.NOM.SG.F ask.PST.SG.F Tanja.ACC.SG.F pour.INF
sebe$_{1,2}$ / ej$_{1,2}$ kofe.
REFL.DAT / 3SG.F.DAT coffee
'Olja asked Tanja to pour herself / her coffee.'

(5) Olja$_1$ poprosila Tanju$_2$ nalit'
Olja.NOM.SG.F ask.PST.SG.F Tanja.ACC.SG.F pour.INF
svoej$_{1,2}$ / eë$_{1,2}$ gost'e kofe.
POSS.REFL.DAT.SG.F POSS.3SG.F guest.DAT.SG.F coffee
'Olja asked Tanja to pour her guest coffee.'

Peškovskij (1956) was the first to point out that reflexivization in Russian object control infinitives is not clause-bound and that the reference of a reflexive pronoun in the infinitive clause is ambiguous between the matrix and the infinitive subject. Subsequent studies have observed that the interpretation of a pronominal in such sentences is also ambiguous between the local and the non-local binding (Bílý 1981; Klenin 1974; Růžička 1973; Timberlake 1979; Yokoyama 1975). Although referential ambiguity has been described for possessives and argumental pronouns, some scholars mentioned that possessives are likely to allow more freedom in their interpretation than argumental pronouns (Padučeva 1983: 17; Timberlake 2004: 241; Kazenin 2000). For example, Kazenin (2000) assumes that only possessive pronominals can refer to both the matrix and the infinitive subject, thus excluding the infinitive subject as a possible referent for argumental pronominals. Differences in the availability of non-local interpretation between possessive and argumental reflexives have been documented for Scandinavian languages, where possessives have been shown to allow non-local binding more than argumental reflexives (Lundquist 2013, 2014). Whether possessives in Russian allow more flexibility in their interpretation than argumental pronouns remains to be shown, and the present study will be a step in this direction.

Recent research has made it increasingly clear that the long-distance interpretation of reflexives varies in subtle ways from dialect to dialect within a language, and from speaker to speaker within a single dialect (Charnavel et al. 2017: 8–9; Lundquist 2013). In Russian, too, the availability of non-local referents for reflexives and local referents for pronominals in object control infinitives has been considered subject to interspeaker variation (Bílý 1981; Klenin 1974; Růžička 1973). Klenin (1974: 40–41) described three "pronoun dialects", representing the preferred interpretive strategies of different speakers. In dialect 1, a reflexive refers only to the infinitive subject and cannot be replaced by a pronominal; a pronominal refers only to the matrix subject. In dialect 2, both a reflexive and a pronominal can refer to either the matrix subject or the infinitive subject. In dialect 3, a reflexive can refer to both the infinitive subject and the matrix subject, but a pronominal can only refer to the matrix subject. So far, claims about interspeaker variation in Russian have been based on the opinions of a few speakers about a few examples. Large-scale empirical studies are needed to determine whether the pronoun dialects described really exist and which of them is the most common. In this article, I hope to provide sufficient evidence to answer this question and thus contribute to understanding the extent of interspeaker variation in the non-local interpretation of reflexives, which has not been adequately recognized in the literature (Charnavel et al. 2017: 9).

## 2.2. Theoretical Accounts of Non-Local Binding in Russian

To accommodate the binding properties of Russian reflexives and pronominals in object control infinitives such as in (4) and (5), several adjustments to the canonical Binding Theory have been proposed by structural accounts.

Rappaport (1986) specified the binding domain for Russian reflexives as the minimal finite clause containing them. This definition, however, contradicts the possibility of the pronominal *ej* in example (4) co-referring with the matrix subject according to Principle B. To resolve this contradiction, Kazenin (2000) proposes different binding domains for Russian reflexives and pronominals, i.e., for reflexives the binding domain is the minimal finite clause, whereas for pronominals it is any minimal cause containing them. This definition implies that a pronominal used in an infinitive clause can be bound within the clause, i.e., it can have the infinitive subject as its antecedent. Kazenin (2000) points out that the problem with his account is that it cannot explain why the infinitive subject can bind a possessive but not an argumental pronominal, which he thinks can only refer to the matrix subject.

Two recent accounts, Dotlačil (2005) and Zubkov (2018), explain the availability of non-local binding for Russian reflexives by considering the structural properties of infinitive clauses. Dotlačil (2005) argues that non-local binding in Czech and Russian arises from a restructuring in the architecture of infinitive clauses. Specifically, in Czech and Russian infinitive clauses, PRO, the element that closes the binding category, can be omitted, leading to non-local binding. Once PRO is present in an infinitive clause, a reflexive must be locally bound (Dotlačil 2005). The problem with this account is that it does not go beyond a few examples to propose a unified procedure for determining when PRO is present and when it is not. Nor does it deal with the interpretation of pronominals.

The multiple Agree-based dependency account (Zubkov 2018; Reuland and Zubkov 2022) claims that non-local binding in Russian object control infinitives can be explained by multiple Agree-based dependencies, established separately for person and number features. Non-local binding arises due to the optional restructuring of infinitive clauses, which results in the absence of a high person probe (the feature instance that initiates the Agree operation) in the C-domain. The person probe in the matrix clause evaluated by the nominative can cross intervening probes, which explains the fact that reflexives in infinitive clauses can be bound by an animate subject of the matrix verb. This approach also addresses the non-complementarity between reflexives and pronouns in infinitive clauses. Since a pronoun has a fully specified φ-feature bundle, it can only be bound by an antecedent if the latter evaluates all probes that can attempt to evaluate the former. If any of the probes diverge, as is the case in infinitive clauses, the pronoun is not excluded and complementarity breaks down (Zubkov 2018: 120). Zubkov's account thus assumes that the

non-local interpretation of reflexives results from an optional restructuring of infinitives. Whether or not an infinitive clause undergoes restructuring is considered to depend on the control verbs, which may differ in the richness of the CP they choose, and on individual speakers, who may differ in whether or not they allow CPs without a high probe. The theory thus recognizes the role of the matrix verb in the non-local interpretation of the Russian reflexive but does not elaborate on it.

What structural accounts have in common is that they all explain why a non-local interpretation of a reflexive is possible in object control infinitives, but they do not predict when the long-distance interpretation actually occurs. The latter problem has been addressed by some non-structural accounts, which are discussed below.

### 2.3. The Role of the Verb in the Interpretation of Pronouns and Reflexives

Recent psycholinguistic findings challenge the view that the interpretation of reflexives involves only syntactic information, in contrast to pronominals, whose interpretation relies primarily on non-structural factors. Instead, it has been shown that the processing of both reflexives and pronominals is constrained by the interaction of syntactic and semantic-pragmatic factors, and that syntactic constraints dominate for reflexives, while semantic-pragmatic constraints dominate for pronominals (Brown-Schmidt, Byron, and Tanenhaus 2005; Kaiser 2003; Kaiser and Do 2012; Kaiser and Trueswell 2008; Kaiser et al. 2009). A similar idea underlies Huang's (2000, 2006) revised neo-Gricean pragmatic account of anaphora, which presumes that pragmatics provides a set of complementary regularities that constrain the interpretation and production of anaphora on top of structural regularities.

In most accounts of long-distance anaphora, semantic-pragmatic factors have been described in terms of logophoricity (Charnavel et al. 2017: 10–26), often operationalized with reference to the ability of the predicate to encode the discourse roles associated with point of view, such as *source*, *self*, and *pivot* in Sells (1987). Experiments on English have shown that the non-local interpretation of a reflexive is more likely with speech verbs (*say*, *claim*) that assign their subjects the role of the source of information than with perception predicates (*hear*, *see*) that assign their subjects the role of the perceiver of information (Kaiser and Do 2012; Kaiser et al. 2009; Sloggett 2017).

In Russian, it was also observed that the reference of possessives in object control infinitives could be influenced by the matrix verb (Bílý 1981: 134–37; Klenin 1974: 142; Růžička 1973: 465–71; Yokoyama 1975: 91). Růžička (1973: 738) described 19 matrix verbs that can introduce an object control infinitive containing a reflexive, pointing out that all matrix verbs used in this construction have the general semantic feature "behavioral control". Among them, there

are verbs that describe verbal communication (e.g., *prikazat'* 'order', *predložit'* 'offer', *prosit'* 'request'), and those that do not (e.g., *zastavit'* 'force', *prinudit'* 'compel', *pomoč'* 'help'). Růžička states that the two groups trigger different preferences in the interpretation of reflexives in their infinitive complements. With the verbs *zastavit'*, *dat'*, *prinudit'*, and *pomoč'*, the reflexive tends to, or even has to be, interpreted as referring to the matrix subject (Růžička 1973: 465–71).

Yokoyama (1975) similarly differentiates between two groups of matrix verbs, i.e., verbs introducing direct discourse including *skazat'* 'tell', *poprosit'* 'ask', *velet'* 'order', *prikazat'* 'order', *ugovorit'* 'talk into', *umolit'* 'plead into', *priglasit'* 'invite', and *razrešit'* 'permit', and verbs not introducing direct discourse such as *zastavit'* 'force', *dat'* 'let', *mešat'* 'hinder', and *pomoč'* 'help'. She hypothesizes that with verbs introducing direct discourse, a reflexive in the embedded clause is likely to take the local referent, and a pronominal tends to take the non-local referent (Yokoyama 1975: 91). The opposite is true for matrix verbs that do not introduce direct discourse. Thus, in (6), the RP *svoj* is most likely to refer to the infinitive subject *reportëra* 'reporter' and the NRP *ego* to the matrix subject *student* 'student'. In (8), on the other hand, the RP *svoj* refers to the subject *otec* 'father' and the NRP *ego* to the object *posetitelju* 'visitor' (Yokoyama 1975: 91). Yokoyama's argumentation is based on the analysis of the transformation of direct discourse such as (7) into indirect discourse with an object control infinitive, as in (6), inspired by Kuno (1972). She claims that with verbs introducing direct discourse, the reflexive of the direct speech, i.e., *svoj* in (7), remains as it is in the infinitive clause, while the first- and second-person pronominals of the direct discourse, i.e., *moj* and *Vaš* in (7), change into the third-person pronominal *ego* (Yokoyama 1975: 92).

(6) Student$_1$ poprosil reportëra$_2$ vzjat'
student.NOM.SG.M ask.PST.SG.M reporter.ACC.SG.M take.INF
svoj$_2$ / ego$_1$ stakan.
POSS.REFL.ACC.SG.M POSS.3SG.M glass.ACC.SG.M
'The student asked the reporter to take his glass.'
(Yokoyama 1975: 91)

(7) Student poprosil reportëra:
student.NOM.SG.M ask.PST.SG.M reporter.ACC.SG.M
"Voz'mite svoj / Vaš
take.IMP.PL POSS.REFL.ACC.SG.M POSS.2PL.M.ACC
/ moj stakan".
POSS.1SG.M.ACC glass.ACC.SG.M
'The student asked the reporter: "Take your / my glass."'
(Yokoyama 1975: 92)

(8) Moj otec$_1$ ne dal
POSS.1SG.NOM.SG.M father.NOM.SG.M NEG give.PST.SG.M
posetitelju$_2$ podnjat' svoj$_1$ / ego$_2$
visitor.DAT.SG.M raise.INF POSS.REFL.ACC.SG.M POSS.3SG.M
stul.
chair.ACC.SG.M
'My father did not let the visitor raise his chair.'
(Yokoyama 1975: 91)

Yokoyama's account has been criticized for its incorrect classification of matrix verbs (Bílý 1981: 136). For example, it is not clear why *razrešit'* 'permit' is considered a verb of direct discourse, while its synonym *pozvolit'* 'allow' is not. Nor is it clear why the verbs *offer, advise*, and *recommend* are classified as verbs that do not introduce direct discourse.

Timberlake (1979: 112; 2004: 249–52) proposed a different classification of matrix verbs. The early version of his hypothesis (Timberlake 1979: 112) states that matrix verbs in object control infinitives form a hierarchy along the parameter of increasing independence, or isolation, of the infinitive clause with respect to rules operating from the matrix into the infinitive clause, including reflexivization, genitive of negation, or restriction on aspect. The main distinction in this hierarchy is between two types of matrix verbs: the verbs *dat'* 'give, let', *pomoč'* 'help', and *zastavit'* 'make, force' on one hand and the verbs *pozvolit'* 'allow', *prikazat'* 'order', *predložit'* 'offer', *zapretit'* 'forbid', *poprosit'* 'ask, request', etc., on the other. Timberlake predicts that a matrix verb of the former type allows or requires reflexivization more than a verb of the latter type. This means that a reflexive in an infinitive clause is more likely to refer to the matrix subject with matrix verbs of the *force* or *let* type than with matrix verbs such as *offer* or *ask*. For instance, Timberlake (1979: 128) claims that in example (9) with a matrix verb of the *force* type, the RP refers to the matrix subject rather than to the infinitive subject.

(9) Roditeli$_1$ zastavljali Serëžu$_2$ ne slušat'
parent.NOM.PL force.PST.PL Serëža.ACC.SG.F NEG listen.INF
ego$_2$ / svoju$_{1,2}$ rakovinu
POSS.3SG.M POSS.REFL.ACC.SG.F shell.ACC.SG.F
vo vremja edy.
during meal.GEN.SG.F
'Parents make Serëža not listen to ( = play with) his/their shell at mealtime.'

**Table 1.** Likelihood of pronoun antecedents in infinitive complements according to the matrix verb (based on Timberlake 2004: 252)

| | Matrix predicate | Frequency of the construction in the RNC | Antecedent = matrix subject | Antecedent = infinitive subject |
|---|---|---|---|---|
| 1 | ***zastavit'* 'force'** | 1332 | sebja \| ?eë<br>svoju \| ?eë | sebja \| *eë<br>svoju \| eë |
| | *dat'* 'give, let' | 800 | | |
| | *pomoč'* 'help' | 782 | | |
| 2 | ***predložit'* 'propose'** | **590** | sebja \| eë<br>svoju \| eë | sebja \| *eë<br>svoju \| ?eë |
| | *poručit'* 'delegate' | 235 | | |
| | *prikazat'* 'order' | 292 | | |
| | *pozvolit'* 'allow' | 453 | | |
| 3 | ***poprosit'* 'request'** | **287** | ?sebja \| eë<br>*svoju \| eë | sebja \| *eë<br>svoju \| ?eë |
| | *umolit'* 'beseech' | 3 | | |
| | *ugovorit'* 'persuade' | 59 | | |
| | *ubedit'* 'convince' | 42 | | |
| | *priglasit'* 'invite' | 44 | | |

In the later version of the theory, Timberlake (2004: 249–50) classifies matrix verbs into three types and formulates the difference between them in terms of the cohesion between the matrix and infinitive clauses, as follows:

> The two clauses are very cohesive if the subject of the matrix predicate controls the outcome of the event, as with *dat'* 'give, let', *pomoč'* 'help', and *zastavit'* 'force'. The two predicates are not cohesive if the matrix subject transfers responsibility for the event to the matrix object (implicit subject), as with *umolit'* 'beseech', *ugovorit'* 'persuade', *ubedit'* 'convince', *priglasit'* 'invite', and *poprosit'* 'ask'. Intermediate are *prikazat'* 'order', *pozvolit'* 'allow', *predložit'* 'propose', and *poručit'* 'delegate'. (Timberlake 2004: 249–50)

As cohesion decreases, the possibility of using a reflexive to refer to the matrix subject decreases, and the pressure to use a reflexive to refer to the infinitive subject increases (Timberlake 2004: 252). This hypothesis has been formalized in Table 1 above, which was adapted from Timberlake (2004: 252)[1] and extended to include the frequency of the object control infinitives with

[1] The notations ? and * are from the original table in Timberlake (2004: 252).

each of the matrix verbs in the Russian National Corpus (RNC), which will be discussed in the methodological section. The table represents a hierarchy of matrix predicates according to the cohesion between the matrix and the infinitive clauses. The first group of verbs (the *force* type) represents the highest degree of cohesion and the third group (the *request* type) the lowest degree of cohesion. As seen from the last column, an RP can always refer to the infinitive subject. However, the likelihood of an RP referring to the matrix subject is highest with matrix verbs of the first group (the *force* type) and lowest with matrix verbs of the third group (the *request* type). Correspondingly, an NRP tends to refer to the matrix subject, but the likelihood that it refers to the infinitive subject is highest with verbs of the first group and lowest with verbs of the third group.

Although Timberlake (2004) did not justify why exactly each of these verbs was assigned to one of the three groups, his definition of cohesion is reminiscent of the event integration and clause integration introduced in the influential typological study of complementation by Givon (1980). The basic idea is that the semantics of the matrix predicate determines the degree of event integration, i.e., the integration of the event described by the matrix clause and the event described by a complement clause into a coherent overall event, which in its turn determines the degree of syntactic integration between the matrix and complement clauses. Matrix predicates are grouped into six semantic types and ordered along a scale based on the influence of the matrix agent over the embedded agent (Givon 1980: 369). The higher the predicate is on the scale, the greater the influence of the matrix agent over the embedded agent, and the less choice, control, and independence are ceded to the embedded agent. Givon (1980) describes two types of "other manipulation" predicates relevant to this study. The highest rank on the scale is taken by success (implicative) predicates, i.e., predicates that encode intended manipulations as necessarily successful (*make*, *cause*, *force*, *prevent*). Lower on the scale are strong attempt predicates, i.e., predicates that do not imply that the intended manipulation is necessarily successful (*order*, *allow*, *permit*, *ask*, *tell*, *let*). As the verbs *make* and *force* are higher on the scale, they are supposed to induce a stronger semantic and syntactic cohesion between the matrix and the complement clause than the verbs *ask* and *tell*, which is in line with Timberlake's (1979: 112) assumption. At the same time, Givon (1980) points out that the degree of semantic and syntactic integration may vary within predicates of the same type due to the semantic nuances of each verb and the pragmatic inferences associated with them.

To summarize this section, it has been observed that the interpretation of reflexives and pronominals in Russian object control infinitives depends on the matrix verb, but it is not clear what features of the matrix verb are responsible for this effect. One account suggests that the relevant distinction is between matrix verbs that introduce direct discourse and those that do not (Yokoyama

1975). Another account argues that the relevant feature is the degree of cohesion between the matrix and the infinitive clause, which is determined by the semantics of the matrix verb (Timberlake 1979, 2004). Both accounts predict the highest likelihood of the non-local interpretation of reflexives and the local interpretation of pronominals with matrix verbs of the *force* type. In addition, Timberlake (2004) proposes a higher probability of the non-local interpretation of reflexives and the local interpretation of pronominals with *propose*-type matrix verbs than with *request*-type matrix verbs. Neither of these proposals has been tested empirically. This study will address both, but will focus on testing Timberlake's hypothesis, which provides a more fine-grained classification of matrix verbs.

## 2.4. Research Questions and Hypotheses

The main goal of this study is to find out which interpretation (local or non-local) is preferred by Russian speakers for the reflexive possessive (RP) and the non-reflexive possessive (NRP) in object control infinitives, and to what extent the interpretation of both possessives is constrained by the matrix verb. Specifically, the study aims to test Timberlake's (2004) cohesion hypothesis formulated as H1:

> H1: The interpretation of an RP and an NRP depends on the matrix verb such that the non-local interpretation of an RP and the local interpretation of an NRP is more likely with a matrix verb inducing a strong cohesion between the matrix and the infinitive clause (the *force*-type verb) than with a matrix verb inducing a weaker cohesion between the clauses (*propose*- and *request*-type verbs).

The second aim is to assess the degree of interspeaker variation in the interpretation preferences of reflexives and pronominals in Russian object control infinitives. If Klenin's (1974) suggestion is correct, Russian speakers should follow one of the three strategies (or pronoun dialects), formulated below as H2–H4:

> H2: There are speakers who adhere to strict complementarity between an RP and an NRP, allowing only the local interpretation for an RP, and only the non-local interpretation for an NRP (dialect 1).
>
> H3: There are speakers who assume free variation in the interpretation of an RP and an NRP, i.e., they accept both local and non-local interpretations of an RP and an NRP (dialect 2).

H4: There are speakers who accept free variation between the local and the non-local interpretation for an RP but accept only the non-local interpretation for an NRP (dialect 3).

Although the focus of the study is on possessives, the interpretation preferences of argumental pronouns are also tested, but on a smaller set of items and without manipulating the matrix verb. Thus, the study aims to provide a general idea of the interpretation preferences of argumental pronouns in addition to possessives, but it is not intended to be a detailed comparison between the two types of pronouns.

## 3. Methodology

The hypotheses were tested by means of a forced choice referent selection task, a format that has been used in previous studies for investigating pronoun reference in adult native speakers (Mertins 2021), second language learners (Pitz et al. 2017), and in children (Baauw et al. 2011). In this study, the task was designed in a written format and implemented over the Internet using the freely available software Google Forms. The task included biclausal sentences with an object control verb and an infinitive complement, which contained either an argumental pronoun in the direct object position or a possessive pronoun as a modifier of the direct object. Each sentence was followed by a question asking for the pronoun referent. Participants were required to answer the question by choosing one of the two suggested alternatives, the matrix subject (non-local interpretation) or the matrix object (local interpretation).

### 3.1. Test Items and Design

Three types of items were constructed: test items with possessive pronouns, test items with argumental pronouns, and filler items.

The test items with possessives were designed to measure the influence of two independent variables, the type of possessive and the type of matrix verb, on the dependent variable, i.e., the referent choice of the possessive. The possessive type had two levels, reflexive and non-reflexive, and the matrix verb type was coded as a three-level variable: a *force*-type verb, a *propose*-type verb, and a *request*-type verb. For each type, one verb was selected from the list of possible matrix verbs proposed by Timberlake (2004: 252) and reproduced in Table 1 on p. 270. The selection criterion was the frequency with which the verb was found as a matrix verb introducing an infinitive complement containing a possessive in the main subcorpus of the Russian National Corpus (RNC).[2]

[2] As there is no category of possessive pronouns in the lexico-grammatical search in the Russian National Corpus, a more general category of adjectival pronouns was

As seen from the third column of Table 1 on p. 270, this construction is most frequent with verbs of the *force* type and least frequent with verbs of the *request* type. Within the *force* type, the verb *zastavit'*, illustrated by (10), is more frequent in this construction than *dat'* or *pomoč'*. Within the *propose* type, infinitive complements with a possessive most frequently occur with the verb *predložit'*, illustrated by (11). Within the *request* type, this construction is most common with the verb *poprosit'*, as in (12). Correspondingly, we chose the verb *zastavit'* to represent the *force* type, the verb *predložit* to represent the *propose* type, and *poprosit'* to represent the *request* type.

(10) A kak že inače: tol'ko čto on, možno skazat', vernul k žizni devočku, vsego liš' **zastaviv eë vspomnit' svoë imja.**

[RNC 2014]

'How could it be otherwise: he had just, one might say, brought the girl back to life, just by **making her remember SELF's name**.'

(11) Počti s poroga xozjain **predložit vam otvedat' ego nastojki** na list'jax èdèl'vejsa.

[RNC 2008–2010]

'Almost from the threshold, the owner **will invite you to try his tinctures** on edelweiss leaves.'

(12) On provël interesnyj èksperiment, v kotorom **poprosil dvux komandirov ocenit' svoix soldat** s točki zrenija fizičeskix kačestv, intellekta, liderstva […].

[RNC 2017–2020]

'He conducted an interesting experiment in which he **asked two commanders to rate SELF's soldiers** in terms of physical attributes, intelligence, leadership […]'

The combination of three values of the verb type and two values of the pronoun type resulted in six test conditions, illustrated with examples in Table 2 on p. 276. Each test item consisted of a sentence describing the same basic situation: *Person 1 influenced Person 2 to act on his/her thing* followed by a question asking whether Person 1 or Person 2 was the most likely possessor of the thing. All test sentences were designed using the following structure: a common male name in the nominative + a matrix verb in the past tense perfective aspect + a second common male name in the dative or accusative case + an infinitive + an accusative noun modified by a possessive pronoun. The infinitive complements were designed such that both the subject and the object

used to search for the relevant constructions.

of the matrix verb could be considered plausible referents given encyclopedic world knowledge, as illustrated by the example in Table 2.

The questions following the test sentences were constructed in two synonymous ways. In half of the items, participants had to answer the question *Komu prinadležit X?* 'Whom does X belong to?', where X was the direct object of the infinitive. The answer options in this case were represented by the names of Person 1 and Person 2 in the dative case. In the other half of the items, the question was *O č'ej vešči idët reč'?* 'Whose thing is this about?'. For this question, answer options were given as names of Person 1 and Person 2 in the genitive case.

A total of 20 different biclausal sentences with possessive pronouns were constructed, each of which was used in the six test conditions described above. This resulted in a total of 120 test items with possessive pronouns, which were then divided into six lists of 20 test items each: 10 sentences with a matrix verb of one type, half with an RP and half with an NRP, and 10 sentences with a matrix verb of another type, again half with an RP and half with an NRP. Each list was presented to 20 participants.

In addition to the test items with possessives, there were 10 test items with argumental pronouns. They had the same syntactic structure as the items with possessives, but they used a different matrix verb to introduce the infinitive complement, namely *skazat'* 'tell'.

The verb *skazat'* was chosen to represent the most neutral verb used in the construction relevant to this study because it has the properties [+ behavioral control, + verbal communication] without any other semantic properties specified (Růžička 1973: 738).

In addition to the test items with possessive and argumental pronouns, 20 filler items with a similar syntactic structure but unrelated to the topic of the study were included in the task. Thus, each participant was presented with a total of 50 items (20 test items with possessives, 10 test items with personal pronouns, and 20 filler items).

### 3.2. Participants

A total of 120 Russian native speakers were recruited for the study through personal networks. The resulting sample of participants was not balanced. It was skewed toward women, with 87 female participants (73%) compared to 33 male participants (28%). In terms of education, there were 99 participants (83%) with higher education compared to 21 participants (17%) with vocational training. In terms of age, 47% of the participants were in the 25–39 age group, 24% were in the 40–60 age group, and 20% were in the 18–24 age group. There were only two participants in the 12–17 age group (2%) and eight participants over 60 (7%), so younger and older speakers were underrepresented in the sample.

**Table 2.** An example of a test item with possessives in six test conditions and of a test item with argumental pronouns in two conditions

| | **Matrix verb** | **Pronoun** | |
|---|---|---|---|
| 1 | *force* | RP | Ivan **zastavil** Olega unesti **svoj** čemodan v druguju komnatu. Komu prinadležit čemodan?<br>1) Ivanu 2) Olegu<br>'Ivan **forced** Oleg to bring **SELF'S** suitcase to another room. Whom does the suitcase belong to?'<br>1) Ivan.DAT 2) Oleg.DAT' |
| 2 | | NRP | Ivan **zastavil** Olega unesti **ego** čemodan v druguju komnatu. O č'ëm čemodane idet reč'? O čemodane<br>1) Olega 2) Ivana<br>'Ivan **forced** Oleg to bring **his** suitcase to another room. Whose suitcase is this about?'<br>1) Oleg.GEN 2) Ivan.GEN' |
| 3 | *propose* | RP | Ivan **predložil** Olegu unesti **svoj** čemodan v druguju komnatu.<br>'Ivan **proposed** to Oleg to bring **SELF'S** suitcase to another room.' |
| 4 | | NRP | Ivan **predložil** Olegu unesti **ego** čemodan v druguju komnatu.<br>'Ivan **proposed** to Oleg to bring **his** suitcase to another room.' |

*—continued on next page—*

| | —continued— Matrix verb | Pronoun | |
|---|---|---|---|
| 5 | *request* | RP | Ivan **poprosil** Olega unesti **svoj** čemodan v druguju komnatu.<br>'Ivan **asked** Oleg to bring **SELF's** suitcase to another room.' |
| 6 | | NRP | Ivan **poprosil** Olega unesti **ego** čemodan v druguju komnatu.<br>'Ivan **asked** Oleg to bring **his** suitcase to another room.' |
| 1 | *tell* | R | Katja skazala Ksjuše kupit' **sebe** bilet.<br>'Katja told Ksjuša to buy **SELF** a ticket.' |
| 2 | | NR | Katja skazala Ksjuše kupit' **ej** bilet.<br>'Katja told Ksjuša to buy **her** a ticket.' |

### 3.3. Procedure and Data Analysis

The experiment was conducted over the Internet. Each participant received a link to the questionnaire. Each session started with written instructions and a collection of relevant participant data. Then, 50 items followed in a single block in an individually randomized order. Each item, including the object infinitive sentence and the question together with the two answer alternatives, was presented on a single screen. The alternatives were displayed below each other, with one of them at the bottom in one half of the items and the other alternative at the bottom in the other half. After marking their choice by clicking on one of the radio buttons, participants pressed a button to move to the next screen. Participants could not go back to already provided answers to correct them. There was no time limit for providing an answer.

Participants' choices were coded as reference to the matrix subject (non-local interpretation) or to the matrix object (local interpretation). The data were first submitted to graphical examination in the form of mosaic plots including a chi-squared test. Subsequently, the data were analyzed using a generalized linear mixed-effects regression in *R* by means of functions *lrm()* and *glrm()*. The modeling aimed to predict the choice of the referent of a possessive pronoun (matrix subject or matrix object) from two linguistic fixed effects, the

type of possessive (reflexive or non-reflexive) and the type of the matrix verb (the *force*, *propose*, or *request* type), and from two random effects, the effect of participant and test item. Additionally, the fixed effect of the following socio-demographic variables was added to the model: participants' age, gender, and education. The reference values for the variables were set as subject for the dependent variable, the reflexive for the type, and the *force* type as the matrix verb type.

## 4. Results

All participants who received the link to the study completed the task, so 120 questionnaires have been evaluated in total. Each participant indicated their interpretation preferences for 20 test items with possessives and 10 test items with argumental pronouns. This resulted in a total of 2,400 data points for possessives, with 400 items in each of the six conditions, and 1,200 data points for argumental pronouns, with 600 items in each of the two conditions.

### 4.1. The Influence of the Matrix Verb (H1)

To recall, the dependent variable in the task was the referent choice for a possessive pronoun, which could be either the matrix subject or the matrix object. Table 3 on the following page gives an overview of how often each of these two referents was chosen in the six conditions. The table reveals a preference for the object interpretation of an RP, i.e., for all three matrix verbs, participants interpreted an RP as indicating that the possessor is the matrix object in more than 66% of test items. However, the object interpretation was more frequent with a matrix verb of the *propose* type (78%) and *force* type (75%) than with a matrix verb of the *request* type (66%). For an NRP, the numbers in Table 3 suggest a strong preference for the subject interpretation with a matrix verb of the *request* type (79% of test items), a weaker preference for the subject interpretation with a matrix verb of the *force* type (63%), and a preference for the object interpretation with a *propose*-type verb (65%).

The bottom part of Table 3 shows the interpretation preferences for argumental pronouns with the matrix verb *skazat'* 'tell'. The argumental reflexive was interpreted as referring to the matrix object and the argumental pronominal as referring to the matrix subject in 92% of occurrences, which suggests their complimentary distribution.

The proportions of the subject and object referents for possessive pronouns in each of the six conditions are visualized in the form of a mosaic plot in Figure 1 on the following page. The bottom part of the figure represents the referent choices for an RP. The blue part in all three bars shows that the object referent was chosen for an RP with all three matrix verbs more often

**Table 3.** The number of subjects and object referents chosen for a reflexive possessive (RP), a non-reflexive possessive (NRP), argumental reflexive (R), and argumental pronominal (NR)

| Condition | Verb | Pronoun | Referent = subject | Referent = object | Total items |
|---|---|---|---|---|---|
| 1 | **force** | RP | 100 (25%) | 300 (75%) | 400 |
| 2 | | NRP | 251 (63%) | 149 (37%) | 400 |
| 3 | **propose** | RP | 90 (22%) | 310 (78%) | 400 |
| 4 | | NRP | 140 (35%) | 260 (65%) | 400 |
| 5 | **request** | RP | 137 (34%) | 263 (66%) | 400 |
| 6 | | NRP | 317 (79%) | 83 (21%) | 400 |
| *Total* | | | *1035 (43%)* | *1365 (57%)* | *2400* |
| 1 | **tell** | R | 47 (8%) | 553 (92%) | 600 |
| 2 | | NR | 555 (93%) | 45 (7%) | 600 |


**Figure 1.** Referent choice for possessives according to matrix verb type; R = reflexive, NR = non-reflexive, S = matrix subject, O = matrix object

than would be expected under the null hypothesis, which predicts equal chances of the subject and the object referents for both an RP and an NRP independent of the matrix verb.

The red color means that the matrix subject was chosen as a referent of an RP less often than would have been expected under the null hypothesis. Additionally, the color intensity signals the strength of this effect. A high intensity of the blue and red colors in the case of *force*- and *propose*-type verbs represents Pearson residuals greater than four, indicating that the departure from the null hypothesis is highly significant. For the verbs of the *request* type, the light blue and red colors indicate smaller residuals (between two and four), which nevertheless point to a significant effect.

Let us now focus on the distribution of subject and object referents of an RP with the three matrix verbs. The chi-squared test indicates that there is no difference in the distribution of RP referents between matrix verbs of the *force* and *propose* types, but there is a significant difference between matrix verbs of the *force* and *request* types. Subject interpretation is more frequent, and object interpretation is less frequent, with a verb of the *request* type compared to *force*- and *propose*-type verbs ($\chi^2$ (799) = 8.21, $p$ = .004).

Turning to the interpretation of an NRP visualized in the upper half of the mosaic plot, we see that subject referents dominate over object referents with matrix verbs of the *force* and *request* types, whereas the object interpretation overrides the subject interpretation with verbs of the *propose* type. The color-coding indicates that the distribution of the subject and object referents with all three matrix verbs significantly differs from that predicted by the null hypothesis. The blue color reveals that the subject referents of an NRP with a matrix verb of the *force* or *request* type were significantly more frequent than expected under the null hypothesis. With a *propose*-type verb, on the other hand, the light blue color indicates that the object interpretation was significantly more frequent than expected if the test condition had no influence. The red color indicates that object referents were chosen less frequently than expected with the matrix verbs of the *force* and *request* types, whereas with the *propose*-type verb, the subject interpretation of an NRP was less frequent than predicted by the null hypothesis. The chi-squared test reveals that with matrix verbs of the *force* type, subject referents of an NRP were chosen more often than with matrix verbs of the *propose* type ($\chi^2$ (799) = 65.04, $p$ < 00001), and less often than with matrix verbs of the *request* type ($\chi^2$ (799) = 26.44, $p$ < 00001).

To summarize, the mosaic plot indicates that the type of matrix verb and the type of possessive significantly influence the choice of a possessive referent. For an RP, the object interpretation dominates, but the subject interpretation is more likely with a *request*-type verb and less likely with the *force*- and *propose*-type verbs. For an NRP, the object referent is most likely with a matrix

verb of the *propose* type, less likely with a verb of the *force* type, and least likely with a verb of the *request* type.

For a more detailed analysis, the data were processed using a generalized linear mixed-effects regression. A binominal mixed model was fit, with the referent of the possessive as a dependent variable, the matrix verb, possessive type as well as participants' age, gender, and education as fixed effects, and participants and items as random effects. Additionally, interaction between matrix verb and possessive type was included in the model because the mosaic plot suggests the interaction might be important. The reference values were set to object for the referent, *force* type for the matrix verb, and NRP for the possessive type. This means that the model predicts the likelihood of the subject interpretation of an RP with matrix verbs of the *propose* and *request* types. The output of the model is given in Table 4, shown on the following page.

The last column of Table 4, indicating *p*-values, reveals that the main effects of linguistic variables and of one interaction are significant. There is a highly significant negative effect of the possessive type, meaning that the chances of a subject referent decrease when the possessive is reflexive compared to when it is non-reflexive.

The effect of the matrix verb is also significant. When the matrix verb is of the *request* type, the likelihood of the subject interpretation increases both for an RP and for an NRP. When the matrix verb is of the *propose* type, the likelihood of subject interpretation decreases. However, a significant interaction between the matrix verb and the possessive type, marked with * in Table 4, indicates that this influence does not affect an RP and an NRP to the same extent. Based on the graphical representation of the data in Figure 1 on p. 279, we can interpret this effect such that an NRP but not an RP is more likely to get a subject interpretation with a matrix verb of the *propose* type than with a matrix verb of the *force* type.

Among the sociodemographic factors, only participants' gender was marginally significant. Female participants were slightly more likely to prefer subject referents than male participants were.

Turning to the random effects summarized in the lower part of Table 4, we see a sizable variation in referent choice between test items and between participants. The referent choices in particular items vary from the average choices with an SD of more than 0.5 points, meaning that individual test items made a large difference in the referent choices. There is also a sizable random effect of participants because the choices of individual participants vary from the overall choices with an SD of 0.4 points, indicating that individual participants differ in their interpretation preferences.

The total explanatory power of our model is substantial (conditional $R^2 = 0.33$), and the part related to the fixed effects alone (marginal $R^2$) is of 0.23. This means that fixed factors and random effects together explain 33% of the variance in the referent choices, whereas the factors matrix verb, possessive

**Table 4.** Output of the model predicting the likelihood of the matrix subject referent of an RP from the matrix verb and possessive type

| **Fixed Effects** | **Estimate** | **Std. Error** | **z value** | ***p*** |
|---|---|---|---|---|
| Intercept | 0.57 | 0.17 | 3.18 | 0.0014 |
| Matrix verb (*propose*) | -1.26 | 0.16 | -7.75 | < 0.0001 |
| Matrix verb (*request*) | 0.91 | 0.18 | 5.18 | < 0.0001 |
| Possessive (REFL) | -1.83 | 0.17 | -10.97 | < 0.0001 |
| Matrix verb (*propose*) *possessive | 1.16 | 0.23 | 4.91 | < 0.0001 |
| Matrix verb (*request*) *possessive | -0.40 | 0.24 | -1.69 | 0.089 |
| Gender (feminine) | 0.31 | 0.14 | 2.23 | 0.027 |

| **Random effects** | **Variance** | **Std. Deviation** |
|---|---|---|
| Item | 0.35 | 0.59 |
| Participant | 0.24 | 0.39 |

type, and participants' gender alone explain 23% of the variance in the data. The low percentage of variance explained by the model suggests that matrix verb and possessive type are not sufficient to account for the referent choices of possessive pronouns and that there are other explanatory factors that have to be included in the model.

### 4.2. Interspeaker Variation (H2–H4)

In order to account for interspeaker variation in interpretation preferences, I will quantify which of the pronoun dialects proposed by Klenin (1974) occur in the data and how often. As a reminder, speakers of dialect 1 strictly follow the complementarity principle. In dialect 2, there is free variation for both reflexives and pronominals. Dialect 3 allows free variation only for reflexives and restricts pronominals to the non-local interpretation. Apart from these dialects, there was another common interpretation strategy in our data, namely the one where the reflexive mostly refers to the matrix object, while the pronominal can refer to either the subject or the object. I will call this interpretation, which assumes free variation between the local and the non-local interpretation for a pronominal, while restricting a reflexive to the object interpretation, dialect 4.

A participant was considered as adhering to dialect 1 in the interpretation of pronouns if they chose a subject referent for a reflexive and an object referent for a pronominal one time at most in the 20 test items. A participant was classified in dialect 2 if they chose a subject referent for a reflexive and an object referent for a pronominal at least two times in the 20 test items. Participants who chose the subject referent for a reflexive in at least two test items, and those who did not choose the object referent for a pronominal in more than one item, were classified in dialect 3. Finally, participants who chose the subject referent of a reflexive in one item at most and the object referent for a pronominal in at least two items were considered speakers of dialect 4.

Table 5 on the following page gives the number of participants in the sample who adhered to one of the four pronoun dialects according to their preferred strategy in the interpretation of possessive and argumental pronouns. As is evident from the table, most participants (67%) interpreted possessives according to the principle of free variation, allowing both an RP and an NRP to refer either to the matrix subject or to the matrix object. The other three pronoun dialects that impose more restrictions on the distribution of possessive referents were less frequent in the data. Thus, 20% of participants tended to prefer the object referent for an RP, while allowing both subject and object referents of an NRP (dialect 4). Another 14% followed the complimentary principle, restricting the referent of an RP to the object and the referent of an NRP to the subject (dialect 2). Only three participants allowed flexibility in

**Table 5.** Number of speakers adhering to a particular pronoun dialect

| | Dialect 1 | Dialect 2 | Dialect 3 | Dialect 4 |
|---|---|---|---|---|
| Possessive | 14 (12%) | **80 (67%)** | 3 (2%) | 23 (19%) |
| Argumental | **83 (69%)** | 5 (4%) | 11 (9%) | 21 (18%) |

**Table 6.** Number of participants adhering to a particular combination of dialects in the interpretation of possessive and argumental pronouns

| | Possessive Dialect 1 | Possessive Dialect 2 | Possessive Dialect 3 | Possessive Dialect 4 |
|---|---|---|---|---|
| Argumental Dialect 1 | 14 (12%) | 46 (38%) | 3 (2%) | 20 (17%) |
| Argumental Dialect 2 | - | 2 (1%) | - | - |
| Argumental Dialect 3 | - | 11 (9%) | - | - |
| Argumental Dialect 4 | - | 15 (13%) | - | 6 (5%) |

the interpretation of an RP, while restricting the interpretation of an NRP to subjects (dialect 3).

For the interpretation of argumental pronouns, the frequency distribution of the four dialects in our data was different from the one of possessive pronouns. Most speakers (69%) observed the complimentarity principle, restricting the reflexive to the object referent and the pronominal to the subject referent. The second most common strategy observed in 18% of participants was dialect 4, which restricts a reflexive to the object interpretation, while allowing both interpretations for a pronominal. Few participants (9%) allowed both referents for the reflexive, while restricting a pronominal to the subject interpretation (dialect 3). The least frequent dialect (4% of participants) was the one that allows free variation in the interpretation of the reflexive and pronominals.

Let us now consider what combinations of the interpretation strategies of the possessive and argumental pronouns exist in our data. Table 6 is a matrix

representing the number of speakers who followed a particular combination of strategies in the interpretation of possessives on one hand and argumental pronouns on the other. The most frequent combination observed in 46 participants (38%) is dialect 2 for possessive pronouns and dialect 1 for argumental pronouns. These speakers allowed free variation in the interpretation of possessives but restricted the interpretation of argumental pronouns to the complementarity principle. The second most frequent combination (in 17% of participants) was dialect 4 for possessives and dialect 1 for argumental pronouns. Like the previous group, these speakers restricted the interpretation of argumental pronouns to the complementarity principle, and additionally, they restricted the interpretation of a possessive reflexive to object referents. The combination of dialect 2 for possessives with dialect 4 for personal pronouns was observed in 13% of participants, closely followed by the combination of dialect 1 both for possessives and argumental pronouns (12%). What is common to all observed combinations is the fact that the interpretation of argumental pronouns, and especially of argumental reflexives, is more restricted than the interpretation of possessives. The empty cells in the matrix show that there was no combination in our data where the interpretation of possessives would be more restricted than the interpretation of argumental pronouns.

To summarize, the analysis of interspeaker variation presented in this section confirms H2 and H3 but rejects H4. It also reveals another common strategy, namely free variation in the interpretation of a pronominal but only local interpretation of a reflexive. Furthermore, it shows that the most frequent strategy in the interpretation of possessives is free variation between local and non-local reference for both an RP and an NRP. By contrast, in the interpretation of argumental pronouns, most speakers followed the complementarity principle, restricting a reflexive to the local referent and a pronominal to the non-local referent. The second most common strategy for both possessives and argumental pronouns is to allow free variation for a pronominal while restricting a reflexive to the local interpretation.

## 5. Discussion

This study investigated preferences of Russian native speakers in the interpretation of possessive and argumental pronouns in object control infinitives. The first question addressed is to what extent the matrix verb constrains the interpretation of possessive pronouns. Specifically, is there a higher probability for the non-local interpretation of a possessive reflexive and for the local interpretation of a possessive pronominal with a matrix verb that implies a stronger cohesion between the matrix and the infinitive clause compared to a matrix verb that implies a weaker cohesion between the two clauses?

The results suggest that the referent choice of a possessive pronominal and, to a lesser extent, of a possessive reflexive is influenced by the matrix

verb. For a possessive reflexive, there is a preference for the local interpretation with all three matrix verb types, i.e., the object referent was preferred in at least 65% of the uses. However, the strength of this preference was influenced by the matrix verb, although not in the direction predicted by Timberlake's hypothesis. Contrary to the expectation that the non-local binding of a possessive reflexive should be most likely with the verb of the *force* type, participants in the study were less likely to choose the non-local referent with the verb *zastavit'* 'force' than with the verb *poprosit'* 'ask, request'. This effect was significant, though small. On the other hand, there was no clear anti-locality bias in the interpretation of a possessive pronominal, i.e., the choice between the local and non-local interpretation depended on the matrix verb. The non-local interpretation was preferred with matrix verbs of the *force* and *request* types (the matrix subject was chosen as a referent in more than 60% of cases), but the local interpretation was preferred with matrix verbs of the *propose* type (the infinitive subject was chosen in 65% of cases). This pattern clearly contradicts the cohesion hypothesis, which predicts that the local interpretation of a pronominal is most likely with the *force*-type verb.

This suggests that the criterion of cohesion as operationalized by Timberlake is not a likely factor underlying the influence of the matrix verb on the interpretation of possessives in object control infinitives. Yokoyama's (1975) account cannot explain binding preferences in our study either. Two of the matrix verbs used in this study, *poprosit'* 'request' and *predložit'* 'propose', belong to the group of verbs that introduce direct discourse, and the verb *zastavit'* 'force' is a verb that does not introduce direct discourse in Yokoyama's (1975: 90) classification. If this distinction was the factor underlying the effect of the matrix verb, the binding preferences with *poprosit'* and *predložit'* should be similar to each other (a local interpretation should be preferred for the reflexive and a non-local interpretation for the pronominal), but different from those with the verb *zastavit'*, which should prompt the non-local interpretation of the reflexive and the local interpretation of the pronominal. However, the results of this study show that for a possessive pronominal, the non-local interpretation is strongly preferred with the verb *poprosit'* and the local interpretation with the verb *predložit'*. A similar effect is observed for a possessive reflexive. That is, although the local binding was preferred with both verbs, the non-local interpretation was more likely with *poprosit'* than with *predložit'*. Therefore, it can be concluded that the explanation proposed by Yokoyama (1975) does not hold.

Since neither Timberlake's nor Yokoyama's classification of matrix verbs can account for our data, I will look for an explanation in the rich literature describing the effects of the verb on the interpretation of pronouns. It has long been known that semantic information encoded in the verb in terms of implicit causality influences the interpretation of pronouns with interclausal

antecedents (e.g., Arnold 2010; Garvey and Caramazza 1974; McDonald and MacWhinney 1995; Stewart, Pickering, and Sanford 2000).

The term implicit causality refers to a type of systematic inference that people make from verbs, e.g., when asked to continue a sentence like *John questioned Mary because…*, people are likely to mention a cause associated with John, whereas in a sentence like *John praised Mary because…* they tend to produce a cause associated with Mary (Pickering and Majid 2007: 1). Verbs like *to question* are classified as NP1-biased verbs (i.e., biased towards the first noun phrase or subject) and verbs like *to praise* are classified as NP2-biased verbs (i.e., biased towards the second noun phrase or object). There is a consensus that the explanation of implicit causality effect lies in the interaction between semantic and pragmatic factors. Pickering and Majid (2007: 6) argue that implicit causality "provides an abstraction of the type of reason that is most likely to be provided for the event and indicates which entity the reason tends to be about. Hence, it is an inference from a description of an event". In the same vein, Bott and Solstad (2014, 2021) believe that the implicit causality bias is an epiphenomenon of the verb's use in a particular discourse context, rather than a feature inherent in the verb's lexical semantics. They argue that verbs showing the implicit causality effect are lexically determined to elicit expectations for particular types of explanation because they have semantically and pragmatically determined slots associated with the biased argument. When these slots are not filled with causally relevant information, the expectation of a particular explanation arises, which then determines the preference for which of the two arguments an anaphor refers to. A similar idea underlies the account of van den Hoven and Ferstl (2018), who suggest that verbs trigger certain schemas that contain basic cognitive representations of the event that are shared across participants (the Idealized Cognitive Models). Speakers rely on an idealized cognitive model to make inferences, unless there is evidence that it does not hold in a given discourse context (van den Hoven and Ferstl 2018). In experimental studies such as the present one, where the influence of discourse context is minimized, participants are likely to resolve the pronominal reference by relying on the basic cognitive representations of the events.

To uncover what the basic cognitive representations might be for the events described by the three matrix verbs in this study, we start with the dictionary explications. The new explanatory dictionary of synonyms (Apresjan 2004: 371) states that *zastavit'* is "a typical verb for the canonical act of coercion, intentional influence on the patient that overcomes his possible resistance and is the immediate reason for the patient performing the action desired by the agent". The dictionary adds that due to its broad meaning, the verb *zastavit'* has many peripheral and semantically weakened uses, where the components of its canonical meaning are bleached. For example, there are contexts where the agent wants to urge the patient to do something that, according to the

opinion of the agent, might be pleasant or useful for the patient. In these uses of *zastavit'*, the meaning component "violence against the will of the patient" gets lost.

In our study, *zastavit'* was used in the structure *Person 1 forced Person 2 to act upon* SELF*'s/his thing*, illustrated by example (13).[3]

(13) Ivan$_1$ zastavil Olega$_2$ unesti
Ivan.NOM.SG.M force.PST.SG.M Oleg.ACC.SG.M carry.INF
svoj$_{1?,2}$ / ego$_{1,2}$ čemodan v
POSS.REFL.ACC.SG.M POSS.3SG.M suitcase.ACC.SG.M to
druguju komnatu.
another.ACC.SG.F room.ACC.SG.F
'Ivan forced Oleg to carry his suitcase to another room.'

(14) Gena$_1$ zastavil Slavu$_2$ povtorit'
Gena.NOM.SG.F force.PST.SG.M Slava.ACC.SG.F repeat.INF
svoj$_{1?,2}$ / ego$_{1?,2}$ rasskaz s
POSS.REFL.ACC.SG.M POSS.3SG.M story.ACC.SG.M from
načala.
beginning.GEN.SG.N
'Gena forced Slava to repeat his story from the beginning.'

(15) Andrej$_1$ zastavil Vadima$_2$ soxranit'
Andrej.NOM.SG.M force.PST.SG.M Vadim.ACC.SG.M save.INF
svoj$_{1?,2}$ / ego$_{1,2?}$ fajl na diske.
POSS.REFL.ACC.SG.M POSS.3SG.M file.ACC.SG.M on CD.LOC.SG.M
'Andrej forced Vadim to store his file on the CD.'

Given the semantic meaning of *zastavit'*, we can infer that both Person 1 and Person 2 could be logically plausible possessors of the thing. It is possible that Person 1 forces Person 2 to act upon the thing belonging to Person 1. Alternatively, Person 2 can be forced to perform an action upon the thing that he himself owns. Correspondingly, example (13) can be interpreted as Ivan forcing Oleg to carry the suitcase that belongs either to Ivan or to Oleg. In line with these possibilities, we found no clear interpretation bias for a possessive pronominal in sentences with *zastavit'*. Instead, there was a significant variation between individual test items. There were items like (13), where the

[3] Here and in examples to follow, the indexing indicates interpretation preferences based on the analysis of individual test items of the study. The question mark indicates the option that was chosen less frequently.

participants chose the subject referent as often as the object referent. On the other hand, there were items like (14), where the object referent was preferred, or items like (15), where the subject referent dominated.

For a possessive reflexive in sentences with *zastavit'*, however, there was a strong interpretation bias towards the object (75%). I suggest this is because in the case of semantic-pragmatic indeterminacy described above, the influence of the default syntactic rule for the reflexive is strongest. That is, the syntactically closest referent is preferred if there are no important semantic or pragmatic reasons that would make the matrix subject a more likely referent.

Let us now examine the dictionary entry for the verb *poprosit'* 'ask, request'. In Apresjan (2004: 882), the following explication is given: "Person X wants an action P to be done: X thinks that person Y can do P without thinking that Y must do P; X tells Y that he wants Y to do P; X says it in such a way that Y understands that X does not think that Y has to do P". Importantly, the dictionary remarks that the action performed is typically in the interest of the speaker and the act of asking is often humiliating for him (Apresjan 2004: 882–83). In our study, the verb *poprosit'* was used in the structure *Person 1 asked Person 2 to act upon* SELF*'s/his thing*, exemplified by example (16).

(16) Ivan$_1$ poprosil Olega$_2$ unesti
Ivan.NOM.SG.M ask.PST.SG.M Oleg.ACC.SG.M carry.INF
svoj$_{1,2}$ / ego$_{1,2?}$ čemodan v
POSS.REFL.ACC.SG.M POSS.3SG.M suitcase.ACC.SG.M to
druguju komnatu.
another.ACC.SG.F room.ACC.SG.F
'Ivan asked Oleg to carry his suitcase to another room.'

Given the core meaning of the verb *poprosit'*, the most likely interpretation of this situation is the one where Person 1 asks Person 2 to act upon the thing belonging to Person 1. Correspondingly, the most frequent interpretation of (13) should be Ivan asking Oleg to carry the suitcase belonging to Ivan. In accordance with this expectation, there was a higher chance of subject referents in such items for both possessives. For a pronominal, the subject referent was strongly preferred over object referent (79% vs. 21%), whereas for a reflexive, the object referent still dominated but was less frequent than with two other matrix verbs (66% vs. 75% / 78%).

Finally, the verb *predložit'* 'suggest, propose' is described in a dictionary as a synonym of *rekomendovat'* 'recommend' and *sovetovat'* 'advise'. The general meaning of these verbs according to Apresjan (2004: 1075) is as follows: "Believing that Y is interested in the opinion of X of what Y should do in this situation, X says that it would be better for Y to do P". The dictionary remarks that this meaning presupposes that the speaker believes that the action performed

is beneficial for the addressee. Given the pragmatic inferences from the semantic meaning of *predložit'* in the situation *Person 1 proposes Person 2 to act upon SELF's/his thing*, the most logical interpretation is the one where the possessor of the thing is Person 2. Therefore, the preferred interpretation of items like (17) should be the one where Ivan proposes that Oleg carry the suitcase belonging to Oleg.

| | | | | | |
|---|---|---|---|---|---|
| (17) | Ivan$_1$ | predložil | Olegu$_2$ | unesti | |
| | Ivan.NOM.SG.M | propose.PST.SG.M | Oleg.DAT.SG.M | carry.INF | |
| | svoj$_{1?,2}$ | / ego$_{1?,2}$ | čemodan | v | |
| | POSS.REFL.ACC.SG.M | POSS.3SG.M | suitcase.ACC.SG.M | to | |
| | druguju | komnatu. | | | |
| | another.ACC.SG.F | room.ACC.SG.F | | | |

'Ivan asked Oleg to carry his suitcase to another room.'

In line with this, we found that object referents were preferred for both possessives. For the reflexive, this preference results in the highest percentage of object referents among the three matrix verbs (78%). For the pronominal, this interpretation even overrides the preference for the subject referent dictated by the structural constraint.

An interesting result of this study is the fact that the general direction of the effect of the matrix verb was similar for reflexives and pronominals. That is, compared to the matrix verb *zastavit'*, the verb *poprosit'* increased the chances for a matrix subject referent, and the verb *predložit'* increased the chances for an object referent, for reflexives and pronominals alike. This finding supports the claim of shared biases advanced by Kaiser and Do (2012). It assumes that although pronouns and reflexives are guided by opposite syntactic biases, they share a bias towards an antecedent that is highly activated / accessible / salient in participants' mental models of discourse. In this study, it could be that in the idealized cognitive model of the verb *poprosit'* described above, the matrix subject is profiled as a more salient referent than the infinitive subject is, whereas for the verb *predložit'* it is the other way round, i.e., the infinitive subject is profiled as a more salient referent.

The key difference between reflexives and pronominals is the strength of this pragmatic constraint compared to the syntactically conditioned interpretation preferences, as suggested by the form-specific multiple constraint models (Kaiser 2003; Kaiser and Do 2012; Kaiser and Trueswell 2008; Kaiser et al. 2009). This study shows that the infinitive subject (local interpretation) is generally a preferred referent for a possessive reflexive, but this preference is weakened when the semantic meaning of the matrix verb and the corresponding pragmatic inferences suggest that the matrix subject is a likely referent. The interpretation of a possessive pronominal, on the other hand, is subject to

a weak structural constraint of anti-locality, which can be overridden by the salience of the referent derived from the pragmatic inferences.

A strong influence of the structural factor on the interpretation of Russian reflexives could be explained by their lack of φ-features and their corresponding participation in the syntactic operations of Move and Agree, as suggested by most current syntactic accounts (Reuland 2018). A preference for the local interpretation of reflexives is also expected from Huang's (2000) pragmatic model, which predicts a local coreferential interpretation for the anaphoric expressions, unless the interpretation violates the disjoint reference presumption, information salience, and the general consistency constraints on conversational implicatures. For reflexives, this means that a local subject is preferred to a local object, a non-split antecedent to a split antecedent, and a c-commanding antecedent to a non-commanding one. However, Huang (2000: 237) points out that if all these NPs contradict the background assumptions about how the world stereotypically works, the next more distant clause is consulted in the same order. For the present study, it means that the non-local subject referent must have been chosen when the local subject referent was an insufficiently salient possessor given the world knowledge. Our analysis above suggests that when contextual information is minimal, as in this study, speakers seem to determine the salience of the referent based on the pragmatic inferences from the basic cognitive representations of the event. In case of a complex event described by a biclausal sentence with an object control infinitive, the basic cognitive representation likely involves an interaction between the semantics of the matrix verb, the semantics of the infinitive verb, and the semantics of a possessive relation. I believe that this complex structure of pragmatic inferences could be an explanation for a considerable item-related variation in interpretation preferences observed in this study. Moreover, it might partially explain why neither Timberlake's (2004) nor Yokoyama's (1975) classification of matrix verbs based on a very general criterion could explain the speakers' interpretation preferences.

Turning to our second question, is there interspeaker variation among Russian speakers such that there are three different strategies for interpreting pronominals and reflexives, namely the complementarity principle, free variation for reflexives and pronominals, and free variation only for reflexives?

The results suggest that different speakers may choose different strategies, but that some strategies are preferred over others. Most speakers in our study seem to follow the strategy of free variation in the interpretation of possessives in object control infinitives, allowing both possessives to refer to either the matrix subject or the matrix object. The second most common strategy is to restrict the interpretation of a possessive reflexive to the local referent, while allowing both local and non-local interpretations for a possessive pronominal. This strategy, which allows free variation only for a possessive pronominal, has not been mentioned in the literature on Russian but was

documented in an experiment on the interpretation of possessives in Czech (Mertins 2021). At the same time, our study suggests that dialect 3, described by Klenin (1974) as allowing free variation for a reflexive and only the non-local interpretation for a pronominal, is virtually nonexistent. We therefore conclude that most Russian speakers allow both local and non-local binding for possessive pronominals and reflexives in object control infinitives, but those who don't restrict the interpretation of a reflexive to the local referent while allowing both referents for a pronominal.

The present study also shows that the interpretation of possessive pronouns in Russian is less constrained than the interpretation of argumental pronouns. Most Russian speakers interpret argumental pronouns according to the principle of complementarity, reserving the local interpretation for a reflexive and the non-local interpretation for a pronominal. The second most common strategy for interpreting argumentative pronouns is the same as for possessives, i.e., a reflexive is restricted to the local referent, while a pronominal can take either the local or the non-local referent. The most common combination of strategies for interpreting possessives and argumentative pronouns is to allow free variation in the interpretation of possessives, but to restrict the interpretation of argumentative pronouns to the complementary distribution.

The difference in the interpretation strategies of argumental and possessive pronouns is in line with previous observations in Russian (Kazenin 2000; Padučeva 1983; Timberlake 2004: 251) and in Scandinavian languages (Linquist 2014). Thus, our results emphasize that the ambiguity between local and non-local interpretations assumed for reflexives and pronominals in Russian object control infinitives exists primarily in cases when the anaphor and its antecedent are not the co-arguments of the same predicate. Argumentative pronouns, on the other hand, are interpreted according to the structural constraints described for finite clauses.

## 6. Conclusion, Limitations, Future Directions

This study was the first to systematically test the referential ambiguity of possessive pronouns in Russian object control infinitives. The results showed that for the possessive reflexive, local binding is generally preferred, as predicted by structural constraints. However, when pragmatic inferences derived from the interaction between the implicit causality patterns of the matrix verb and the infinitive verb make a non-local referent salient, the syntactic preference is weakened. For the possessive pronominal, pragmatic inferences can completely override the structural constraint of non-locality and lead to a strong preference for either the local or the non-local interpretation.

The proposed explanation in terms of pragmatic inferences from implicit causality needs further verification. First, other matrix verbs occurring in ob-

ject control infinitives should be investigated in order to get a more detailed picture of the semantic and pragmatic mechanisms behind their influence on pronoun interpretation in object control infinitives. Second, it might be useful to obtain explicit opinions from speakers about how they arrived at a particular interpretation. Third, other methodological approaches are welcome that would overcome the limitations of the forced-choice task and that would capture referent choice not as a binary procedure but as a complex process in which multiple factors are weighed against each other to arrive at a final interpretation.

## Sources

[RNC] Russian National Corpus [Nacional'nyj korpus russkogo jazyka]. (2003–2023) Available at: https://ruscorpora.ru/.

## References

Apresjan, Jurij D., ed. (2004) *Novyj ob" jasnitel'nyj slovar' sinonimov russkogo jazyka*. 2nd ed. Moscow/Vena: Jazyki slavjanskoj kul'tury.

Arnold, Jennifer E. (2010) "How speakers refer: The role of accessibility". *Language and linguistics compass* 4: 187–203.

Baauw, Sergio, Shalom Zuckerman, Esther Ruigendijk, and Sergey Avrutin. (2011) "Principle B Delays as a processing problem: Evidence from task effects". Angela Grimm, Anja Müller, Cornelia Hamann, and Esther Ruigendijk, eds. *Production-comprehension asymmetries in child language*. Berlin: De Gruyter, 247–72.

Bílý, Milan. (1981) *Intrasentential pronominalization and functional sentence perspective (in Czech, Russian, and English)*. Lund: Slaviska Institutionen. [Lund Slavonic Monographs.]

Bott, Oliver, and Torgrim Solstad. (2014) "From verbs to discourse: A novel account of implicit causality". Barbara Hemforth, Barbara Mertins, and Catherine Fabricius-Hansen, eds. *Psycholinguistic approaches to meaning and understanding across languages*. New York: Springer, 213–51.

————. (2021) "Discourse expectations: Explaining the implicit causality biases of verbs". *Linguistics* 59(2): 361–416.

Brown-Schmidt, Sarah, Donna K. Byron, and Michael K. Tanenhaus. (2005) "Beyond salience: Interpretation of personal and demonstrative pronouns". *Journal of memory and language* 53(2): 292–313.

Charnavel, Isabelle, James Huang, Peter Cole, and Gabriella Hermon. (2017) "Long-distance anaphora: Syntax and discourse". Martin Everaert and Henk Van Riemsdijk, eds. *The Wiley Blackwell companion to syntax*, second edition. Hoboken, NJ: John Wiley & Sons, 1–82.

Chomsky, Noam. (1981) *Lectures on government and binding*. Dordrecht: Foris.
Dotlačil, Jakub. (2005) "Non-local binding in Slavic languages and restructuring". Sylvia Blaho, Luis Vicente, and Erik Schoorlemmer, eds. *Proceedings of ConSOLE XIII*, 1–16.
Garvey, Catherine, and Alfonso Caramazza. (1974) "Implicit causality in verbs". *Linguistic inquiry* 5(3): 459–64.
Givón, Thomas. (1980) "The binding hierarchy and the typology of complements". *Studies in language* 4(3): 333–77.
Huang, Yan. (2000) *Anaphora: A cross-linguistic study*. New York: Oxford University Press.
————. (2006) "Anaphora and the pragmatics-syntax interface". Laurence R. Horn, and Gregory Ward, eds. *The handbook of pragmatics*. Hoboken, NJ: Blackwell Publishing, Ltd., 288–314.
Kaiser, Elsi. (2003) *The quest for a referent: A crosslinguistic look at reference resolution*. PhD dissertation, University of Pennsylvania.
Kaiser, Elsi, and Monica Do. (2012) "Semantic effects on pronouns and reflexives in picture-NPs: Similarities and differences". *University of Pennsylvania working papers in linguistics* 18(1): 115–24.
Kaiser, Elsi, and John C. Trueswell. (2008) "Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution". *Language and cognitive processes* 23: 709–48.
Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman, and Michael K. Tanenhaus. (2009) "Structural and semantic constraints on the resolution of pronouns and reflexives". *Cognition* 112: 55–80.
Kazenin, Konstantin. (2000) "Infinitives and constraints on pronominals in Russian". Tracy Holloway King and Irina Sekerina, eds. *Formal approaches to Slavic linguistics 8: The Philadelphia meeting 1999*. Ann Arbor, MI: Michigan Slavic Publications, 189–213.
Klenin, Emily. (1974) *Russian reflexive pronouns and the semantic roles of noun phrases in sentences*. PhD dissertation, Princeton University.
Kuno, Susumu. (1972) "Pronominalization, reflexivization, and direct discourse". *Linguistic inquiry* 3: 161–95.
Lundquist, Björn. (2013) "On inter-individual variation and mid-distance binding in Swedish". *Working papers in Scandinavian syntax* 91: 113–46.
————. (2014) "Mid-distance binding". *Nordic atlas of linguistic structures* (NALS). Available at: https://tekstlab.uio.no/nals/chapter_text/24/binding3.htm.
McDonald, Janet L., and Brian MacWhinney. (1995) "The time course of pronoun resolution: Effects of implicit verb causality and gender". *Journal of memory and language* 34(4): 543–66.

Mertins, Barbara. (2021) "On the interpretation of possessives in Czech: An experimental approach". Catherine Fabricius-Hansen, ed. *Resolving possessive puzzles*. Oslo: University of Oslo, 97–126.

Padučeva, Elena V. (1983) "Vozvratnoe mestoimenie s kosvennym antecedentom i semantika refleksivnosti". *Semiotika i informatika* 21: 3–33.

Peškovskij, Aleksandr M. (1956) *Russkij sintaksis v naučnom osveščenii*. Moscow: Prosveščenie.

Pickering, Martin J., and Asifa Majid. (2007) "What are implicit causality and consequentiality?" *Language and cognitive processes* 22(5): 780–88.

Pitz, Anneliese, Oliver Bott, Torgrim Solstad, Robin Hoernig, Bergljot Behrens, and Cathrine Fabricius-Hansen. (2017) "An empirical L2 perspective on possessives: German/Norwegian". *Oslo studies in language* 9(2): 41–75.

Rappaport, Gilbert. (1986) "On anaphor binding in Russian". *Natural language and linguistic theory* 4: 97–120.

Reuland, Eric. (2018) "Reflexives and reflexivity". *Annual review of linguistics* 4: 81–107.

Reuland, Eric, and Jan Koster. (1991) "Long-distance anaphora: An overview". Eric Reuland and Jan Koster, eds. *Long-distance anaphora*. Cambridge, UK: Cambridge University Press, 1–26.

Reuland, Eric, and Peter Zubkov. (2022) "Agreeing to bind: The case of Russian". *Glossa: A journal of general linguistics* 7(1).

Růžička, Rudolf. (1973) "Reflexive oder nichtreflexive Pronominasierung in modernen Russischen und anderen slawischen Sprachen der Gegenwart". *Zeitschrift fűr Slawistik* 17: 636–779.

Sells, Peter. (1987) "Aspects of logophoricity". *Linguistic inquiry* 18(3): 445–79.

Sloggett, Shayne. (2017) *When errors aren't: How comprehenders selectively violate Binding Theory*. PhD dissertation, University of Massachusetts Amherst.

Stewart, Andrew J., Martin J. Pickering, and Anthony J. Sanford. (2000) "The time course of the influence of implicit causality information: Focusing versus integration accounts". *Journal of memory and language* 42(3): 423–43.

Timberlake, Alan. (1979) "Reflexivization and the cycle in Russian". *Linguistic inquiry* 10(1): 109–41.

————. (2004) *A reference grammar of Russian*. Cambridge, UK: Cambridge University Press.

van den Hoven, Emiel, and Evelyn Ferstl. (2018) "Discourse context modulates the effect of implicit causality on rementions". *Language and cognition* 10(4): 561–94.

Yokoyama, Olga. (1975) "Personal or reflexive? – A functional analysis". Susumu Kuno, ed. *Harvard studies in syntax and semantics I*. Cambridge, MA: Harvard University Linguistics Department, 75–112.

Zubkov, Peter. (2018) *The grammar of binding: A study with reference to Russian*. Utrecht: LOT. [International dissertation series.]

Tatiana Perevozchikova
Slavic Department
University of Tübingen
Tübingen, Germany
tatiana.perevozchikova@uni-tuebingen.de

# Subject Doubling in the Slovenian Dialect of Resia

**Florian Wandl*, ****

*Abstract*: One of the most curious phenomena that makes the Resian dialect of Slovenian stand out among the Slavic languages is subject doubling. Subject phrases in Resian can be doubled by clitic variants of the personal pronouns. Within Slavic, this is unknown outside the Romance-Slavic contact zone in northern Italy, which is why it is generally explained as a borrowing, most probably from Friulian (Rhaeto-Romance). Despite being such a rarity, studies dealing with subject doubling are scarce, and the phenomenon remains poorly understood. This paper aims at a description of Resian subject doubling, focusing on (1) the types of subject phrases that occur with doubling and (2) the place the subject clitics occupy in clauses with doubling. To identify cases of subject doubling, a recent translation of *The Little Prince* is used. Comparing potential cases with the French original helps to distinguish instances of subject doubling from instances of left- and right-dislocation. The analysis shows that subject clitics always precede the predicate. Apart from cases with subject-verb inversion, they follow the subject phrase but can be separated from it by adverbials. Partly in line with earlier research, it is observed that, with the exception of interrogatives and indefinite pronouns, all types of subjects (including universal quantifiers) occur with doubling. Moreover, it is shown that the lack of animacy, definiteness, and specificity do not inhibit subject doubling. Finally, subject doubling in Resian is contrasted with the use of subject clitics in Friulian as the language that, most probably, provided the example for Resian subject doubling.

*Keywords:* subject clitics, clitic doubling, subject doubling, Resian, Romance-Slavic language contact

* I am thankful to Malinka Pila and Andreas Hölzl for their comments on earlier versions of this paper. Moreover, I would like to thank Malinka Pila and Han Steenwijk for their counsel on some of the Resian data, and Uwe Junghanns for his helpful suggestions for improving the paper. Proofreading has been provided by Dana Serditova and Alexander Robert Herren, which I gratefully acknowledge. Any remaining errors or omissions are my own.

** Barbara Sonnenhauser zum 50. Geburtstag gewidmet.

## 1. Introduction

Within Slavic languages, doubling of verbal arguments is a phenomenon well-known from Bulgarian and Macedonian.[1] In these languages, direct and indirect objects can be doubled by clitic pronouns under certain circumstances (cf., for example, Koneski 1976: 335; Usikova 2003: 140; Werkmann 2015, and the contributions on Slavic languages in Kallulli and Tasmowski 2008). What is less known is that in Slavic we also find varieties that show doubling of subject phrases. Western Slovenian dialects spoken in northern Italy have acquired subject doubling in contact with Romance varieties. Thus, in the two examples in (1) taken from Steenwijk's (1992: 201, 220) transcripts of spoken Resian, the tonic third-singular-masculine personal pronoun *wón* 'he' and third-singular-feminine personal pronoun *woné* 'she' are doubled by their respective subject clitics *an* and *na*.[2]

(1) a. **Wón** **an** ma njagé mansione […]
he.NOM he.SC have.PRS.3SG his job
'He has his job […].'

b. **Woné** **na** ma pašjón […]
she.NOM she.SC have.PRS.3SG hobby.ACC.SG.M
'She has a hobby […]'

The existence of a separate set of subject clitics and their use in a doubling construction is exceptional within Slavic. Studying the phenomenon of Resian subject doubling, therefore, promises new insights into the development and functioning of clitic systems. In this paper, I intend to give a description of Resian subject doubling based on the recent translation of de Saint-Exúpery's novella *The Little Prince* (French original, *Le petit prince*). The main focus of this paper will be on the place of the subject clitic in relation to its associate, i.e., the doubled phrase (Kallulli and Tasmowski 2008), and on the parts of speech which are doubled. The pragmatics of clitic doubling are difficult to determine from the available resources and require a separate study. Since the analyzed data come from a written source, they do not allow firm conclusions to be made about the use of subject doubling in actual speech. It has been proposed that subject doubling is optional in Resian, and data from fieldwork provided by Steenwijk (1992) suggest that in spoken language, it occurs much less frequently

[1] Object doubling, moreover, occurs in the Torlak dialects of Serbian, which share a number of features with the Bulgarian and Macedonian languages (cf. most recently Escher 2021).

[2] Throughout this article, boldface type is used to highlight relevant forms in examples. A list of abbreviations is given at the end of the article, before the References.

than in the primary source of the present paper. Furthermore, there seems to be inter-speaker variation in the use of subject clitics. While questions about the pragmatics of subject doubling and about its occurrence in actual speech must remain unanswered in this paper, the investigated resource allows us to determine cases in which doubling can potentially occur. As will be demonstrated in §5.2, this, for instance, allows us to exclude the absence of semantic properties such as definiteness, specificity, and animacy of a noun phrase as factors prohibiting subject doubling in Resian. These factors have been shown to influence subject doubling in other languages. Additional data considered in this paper come from the transcripts provided by Steenwijk (1992) and from the secondary literature (e.g., Benacchio 2002; Skubic 1997; Šekli 2010a).

The paper is organized as follows: Section 2 briefly introduces the phenomenon of subject doubling. Section 3 gives some general background information about the Resian dialect. Section 4 provides a description of the Resian inventory of subject clitics, and Section 5 is dedicated to the phenomenon of subject doubling in this language. Section 6 presents the paper's conclusions.

## 2. The Phenomenon of Subject Doubling

In the previous section, we have already seen two examples with subject doubling from the Slovenian dialect of Resia (see examples 1a–b). To give some context for the more detailed discussion of subject doubling in Resian in §5, I provide a brief general introduction to this phenomenon here. In doing so, I will mainly focus on aspects that are relevant for the discussion of the Resian data.

Subject doubling can be defined as the reduplication of a subject phrase by a clitic pronoun. For instance, in example (2), a French clause is given in the standard variant (2a) and in a colloquial variant (2b). As can be seen, in the colloquial variant, the subject *le garçon* is repeated by the clitic *il*. However, according to Coveney (2005: 96), they have the same meaning and can be used in the same contexts.[3] It is this type of doubling that is of interest for the present paper.

(2) a. Le garçon mangeait une pomme.
DEF.M boy eat.IMPERF.3SG INDEF.F apple

b. Le garçon il mangeait une pomme.
DEF.M boy he.SC eat.IMPERF.3SG INDEF.F apple
'The boy ate an apple.'

[3] Note that Coveney (2005: 96) notates the example reproduced as (2b) with a comma after *le garçon*. However, he explicitly states that this is merely a convention and that, in actual speech, a break in the intonation pattern is not necessarily present.

Compared to the doubling of direct and indirect objects, subject doubling has received less attention in the scholarly literature.[4] It is mainly known from Romance and Germanic varieties, e.g., colloquial and dialectal French (Coveney 2005), Rhaeto-Romance (Haiman and Benincà 1992), dialectal Dutch (de Vogelaer and Neuckermans 2002), and Övdalian (Rosenkvist 2015), but has also been reported, for instance, from colloquial Finnish (Holmberg and Nikanne 2008) and Arabic (Aoun 1999; Jlassi 2013). As mentioned in §1, within Slavic, subject doubling is attested solely in Slovenian varieties spoken in northern Italy, most prominently in Resian (Skubic 1997: 84; Benacchio 2002; Šekli 2010a).

Like object doubling, subject doubling must be distinguished from clitic left- and right-dislocation (Anagnostopoulou 2017: 4–6). In the latter phenomenon, subjects or objects occur outside the clause boundaries. Depending on whether they precede or follow the respective clause, the phenomenon is referred to as left- or right-dislocation. Unlike subject and object doubling, dislocation is known from a vast number of languages and is sometimes considered universal (e.g., Westbury 2016 on left-dislocation). Moreover, dislocation occurs as a separate phenomenon even in languages that exhibit clitic doubling (cf., for instance, Krapova and Cinque 2008 on clitic doubling and other clitic reduplication constructions in Bulgarian). The function of dislocation is related to information structure. Left-dislocated constituents usually introduce a new topic or mark a topic-shift, while right-dislocated constituents express what is often referred to as an antitopic (e.g., Lambrecht 1994: 176–84, 199–205). For this paper, information structure is of minor relevance since it does not aim at a contribution in this domain. An important issue, however, is how doubling constructions can be distinguished from instances of dislocation. The methodology applied to do so will be introduced in §5.1.

The scope of subject doubling has been related to certain semantic features of the doubled phrase. For example, subject doubling occurs almost exclusively with personal pronouns in dialectal Dutch. Only in some varieties can proper names be doubled as well. Moreover, most of the dialects with doubling do not duplicate the neuter third-person pronoun (see de Vogelaer and Neuckermans 2002: 238 on these restrictions). According to de Vogelaer and Neuckermans (2002), variation can be accounted for by referring to the animacy hierarchy: the closer a subject is to the animate pole of the scale, the more frequently it is doubled. Moreover, doubling in Dutch can be influenced by word order and clause type (main clause vs. subclause) (cf. further, Vogelaer and Devos 2008).

In colloquial French, subject doubling is more likely to occur with subjects that are [+specific] and [+definite]. If only one of these parameters has a positive value, doubling is possible but less likely; if both of them are negative, the

[4] Cf., for example, the lengthy survey by Anagnostopoulou (2017) where subject doubling is not addressed.

subject will not be doubled.[5] Subject doubling is also more likely if the subject is [+animate] (Nadasdi 1995).

The features relevant for subject doubling in these languages, therefore, seem to be identical to those which have been identified as having an influence on the realization of object doubling: specificity, definiteness, animacy, and pronominality (cf. Anagnostopoulou 2017; Kallulli and Tasmowski 2008). One of the objectives of the present paper is to investigate whether subject doubling is prohibited in Resian if the subject phrase lacks one or several of these features. Before the Resian data will be discussed in §5, it is, however, in order to briefly introduce this dialect to familiarize the reader with the context in which it is spoken.

## 3. The Resian Dialect of Slovenian

Resian is an autochthonous Slovenian dialect spoken in the Resia Valley, located in the autonomous region Friuli-Venezia Giulia in the northeast of Italy. Its speakers are mainly distributed across the villages San Giorgio (Italian) / Bila (Resian), Prato di Resia/Ravanca, Gniva/Njiva, Oseacco/Osojane, and Stolvizza/Solbica. Their number is estimated to be 1,000, which corresponds to the number of inhabitants in the municipality of Resia. Speakers of Resian are bilingual or even trilingual: They all speak the national language, Italian, and sometimes also Friulian (Steenwijk 1992: 2). It has been observed that children often do not learn Resian as their mother tongue at home, which makes it a "definitely endangered" language according to the UNESCO scale (Moseley 2010: 24–25).

Within Slovenian, Resian belongs to the Carinthian dialect group (cf. Ramovš 1928). However, as a result of Germanic and Romance colonization of the Canale and Belska valleys in the 14th and 15th centuries, ties with the Carinthian dialect group were cut. Subsequently, contact with the Littoral dialect group intensified, which is why Resian shares a number of features with this group as well. Overall, Resian developed largely independently from other Slovenian varieties due to its remote location. According to Ramovš (1935: 32), it should thus be considered a dialect of its own.

The Resian dialect itself is usually divided into four subdialects equal to the main varieties spoken in Bila, Njiva, Osojane, and Solbica (Steenwijk 1996). The differences between these dialects are mainly phonological, which is why they will not concern us in this paper (cf. Steenwijk 1996 for some non-phonological differences).

Resian has been in intense language contact with Romance at least since the 11th century (Breu 2022: 8). In the area surrounding Resia, Carnic dialects of Friulian are spoken. Contact with these varieties lasted for several centuries

[5] Note that doubling seems to be more common in colloquial Swiss French (cf. Fonseca-Greber 2000).

and has left a profound impact on Resian.[6] Moreover, Resian has been in contact with Italian dialects and, since 1866, also with the Italian standard language.

Starting from the 1980s, efforts have been made to create a literary standard for Resian. An orthographic norm has been developed by Steenwijk (1994), who has also published several works on the grammar and lexicon of Resian (Steenwijk 2005, 1999a, 1999b). However, it should be noted that despite the small number of Resian speakers, the proposed orthography has not been generally accepted. For more information about the textual attestation of Resian, which goes back to the 18th century, and the history of research on this dialect, the reader is referred to Benacchio (2002: 71–75), Steenwijk (1992: 3–7), Dapit (2003), Malakov (2017), and Šekli (2018: 82–83). An overview of the history and sociolinguistics of Slovenian varieties in northern Italy can be found in Čermelj (1938) and Skubic (1997: 9–50).

## 4. Resian Subject Clitics

Resian as well as a few other Slovenian dialects spoken in northern Italy have developed a system of personal pronouns that is unique within Slavic. Apart from the clitic variants of the genitive, dative, and accusative forms characteristic of West and South Slavic languages (Franks and King 2000), they have developed clitic variants of the nominative forms of the personal pronoun. These nominative clitics, which are referred to as "subject clitics" in this paper,[7] appear in all three persons in the singular and the plural. Solely in the dual, we do not find clitic variants of the tonic pronouns. Instead, in the third person, the plural clitic or, according to Šekli (2010a), the neuter singular clitic is used. Tables 1 through 3 on the next page show the system of Resian subject clitics according to Steenwijk's (1992: 118) grammar of the dialect of San Giorgio. For the sake of convenience, the examples have been transcribed into the Resian standard orthography (cf. Steenwijk 1994), which is also used in the translation of *The Little Prince* (de Saint-Exupéry 2021), the primary resource of the present study.

As can be seen in tables 1 through 3, subject clitics differ from their tonic counterparts not only by their prosodic features but also segmentally. These differences are least significant in the first- and second-person personal pronouns, where, apart from the tonic first-singular form *jäs*, tonic and clitic variants are distinguished from each other solely by the quality of the vowel. In the third person, on the other hand, tonic and clitic variants differ from each other also by the lack of certain segments in the latter. The origin of these differences must largely be sought in regular sound changes, which affected either the accented or the unaccented variant of the personal pronouns (cf. Šekli 2015 for a brief

[6] Note that, of the above-mentioned Resian dialects, the dialect of San Giorgio shows the most influence from Romance (Steenwijk 1996).

[7] This term is used in parallel to the nomenclature applied in Romance linguistics.

**Table 1.** Nominative forms of the first-person personal pronouns in Resian

| 1st person | Tonic | Clitic |
|---|---|---|
| SG | *jäs/jä* | *ja* |
| DU | *midwa* (M), *midvi* (F) | not attested |
| PL | *mï* | *mi* |

**Table 2.** Nominative forms of the second-person personal pronouns in Resian

| 2nd person | Tonic | Clitic |
|---|---|---|
| SG | *tï* | *ti* |
| DU | *vidwa* (M), *vidvi* (F) | not attested |
| PL | *vï* | *vi* |

**Table 3.** Nominative forms of the third-person personal pronouns in Resian

| 3rd person | M | | F | |
|---|---|---|---|---|
| | Tonic | Clitic | Tonic | Clitic |
| SG | *wun* | *an* | *wonä* | *na* |
| DU | *wonedva* | not attested | *wonedvi* | not attested |
| PL | *wuny* | *ni* | *wone* | *ni* |

overview of Resian sound changes). Finally, in the third-person plural form, gender differences have been eliminated in the clitic variant (Šekli 2010a: 145–46).

In the singular, Resian also has a neuter personal pronoun. However, according to Šekli (2010a: 146), the tonic third-singular-neuter pronoun *onö* 'it' is only rarely used. Instead, speakers employ the clitic *to* based on the demonstrative stem *t-*. The investigated text does not contain the pronoun *onö*.

There are currently no detailed studies on the use and placement of pronominal clitics in Resian. The matter is merely addressed as part of works with a more general scope (e.g., Skubic 1997; Benacchio 2002). The most informative study on the use of personal pronouns in Resian is Šekli (2010a). However, the author is not concerned with the place that subject clitics occupy in the clause. Therefore, the information on the position of subject clitics in the clause reported

in the following stems from a separate study on pronominal clitics in Resian that is currently in preparation (Wandl, forthcoming).

While standard Slovenian is a pro-drop language, in Resian pronominal subjects are usually expressed by subject clitics (Šekli 2010a: 150–51). At least in some cases, the expression of pronominal subjects, however, is not obligatory. Example (3) contains two largely identical clauses taken from Šekli (2010a: 150 and cf. also Šekli 2010b: 161) (accent marks have been adapted for the sake of convenience). These clauses differ from each other in the presence or absence of a subject clitic. When, exactly, pro-drop occurs in Resian has yet to be properly studied. Šekli (2010a: 150) suggests that it is related to information structure. For the present paper, this question is not of immediate relevance since it is concerned mainly with the types of subjects that are doubled and with the place of the subject clitics in constructions with doubling.

It is interesting to note that subject clitics are placed according to different rules than object clitics. They always precede the highest-ranking verb in the clause (Wandl, forthcoming, and less specifically, Skubic 1997 and Benacchio 2002). Usually, this is the finite verb, including clitic and tonic auxiliaries. However, in past tense and modal constructions, the auxiliary can also be omitted, in which case the subject clitics precede the past participle (i.e., the so-called *l*-participle). Since Resian subject clitics are proclitic to the verb, they can appear in clause-initial position.

(3) a. […] kə to orę́ w Rę́zijə
because this.ACC.SG.F hour.ACC.SG.F in Resia.LOC.SG.F

**ja** si jin dál wžę́
I.SC AUX.1SG they.DAT.PL give.PST.PTCP.SG.M already

den lípə azək […]
INDEF.ACC.SG.M beautiful.ACC.SG.M language.ACC.SG.M

'[…] because in Resia I already gave them a beautiful language […]'

b. […] kə **Ø** si jin dál
because **Ø** AUX.1SG they.DAT.PL give.PST.PTCP.SG.M

den lípə azək […]
INDEF.ACC.SG.M beautiful.ACC.SG.M language.ACC.SG.M

'[…] I already gave them a beautiful language […]'

Although subject clitics have to precede the verb, they do not have to do so immediately. We find examples with both tonic and clitic elements occurring between subject clitics and verbs (more on this in §5.3 below). Moreover, subject clitics do not form a clitic cluster with other pronominal clitics. While the former must precede the highest-ranking verb in the clause, the latter usually precede their immediate lexical head verb. Therefore, whenever the lexical verb

is preceded by auxiliaries, adverbs, or particles, subject and object clitics do not occur subsequent to each other.

In sum, the investigated data justify the qualification of Resian subject clitics as clitics if they are understood as "elements with some of the properties characteristic of independent words and some characteristic of affixes" (Zwicky 1994: xii). Compared to their tonic counterparts, they not only lack accent but also differ in their phonological structure. Especially in the third person, segmental differences are so significant that they cannot simply be explained by means of synchronic rules. Furthermore, subject clitics are more restricted with regard to their placement than their tonic counterparts. While the latter can occur both before and after the verb depending on information structural requirements, the former must precede the verb.[8]

From what has been said above, it is also clear that in the broader Slavic context, Resian belongs to the group of languages with verb-oriented clitics. In this regard, it is more similar to the eastern South Slavic languages Bulgarian and Macedonian than to standard Slovenian, which belongs to the group of languages with second-position (Wackernagel position) clitics (Franks and King 2000).

## 5. Subject Doubling in Resian

Having provided some general information about subject doubling (§2) and introducing the Resian language (§3), in this section, I discuss the phenomenon of subject doubling in Resian. To do so, I first describe the resource and the methodology that has been applied to identify constructions with doubling (§5.1). In line with the objectives of this paper, I then provide a survey of collected data with a focus on the types of associates found with doubling (§5.2) and on the place of subject clitics in clauses with subject doubling (§5.3).

### 5.1. Resource and Methodology

As mentioned in §2, a major issue in analyzing doubling constructions is distinguishing them from instances of dislocation, i.e., instances in which subjects or objects are detached from the clause that contains the co-referential pronoun. This problem is especially salient when working with low-resource languages,

[8] An anonymous reviewer asks whether Resian subject clitics could be classified as weak pronouns in the sense of Cardinaletti and Starke (1999). However, Cardinaletti and Starke (1999: 169) conclude that "[d]oubling is always clitic-doubling, in the sense that doubling must always involve (at least) one clitic, no combination of weak and strong pronoun is possible", which seems to exclude the very analysis the reviewer asks about. See Testelets (2003) for a further analysis that casts doubt on the general applicability of the concept of weak pronouns.

which often lack detailed descriptions of the phenomena involved. However, I propose that identifying relevant cases can still be possible based on the existence of parallel texts in doubling and non-doubling languages.

The parallel text I use as the main resource for this investigation is the recent translation of *The Little Prince* by Antoine de Saint-Exupéry (de Saint-Exupéry 2021) into Resian. It has been prepared by a native speaker of Resian, Silvana Paletti, and a linguist, Malinka Pila, based on the French original. The native dialect spoken by Silvana Paletti is that of Varcota, which is largely identical to the dialect of San Giorgio.

The way that doubling constructions are identified in the Resian version of *The Little Prince* is similar to Coveney's (2005) approach to subject doubling in colloquial French. This author argues that structures can be counted as an example of subject doubling if non-doubled structures can replace them without resulting in a change in the semantics or pragmatics of the expression (Coveney 2005: 100). Although the application of these diagnostics leads Coveney (2005: 99–102) to also include instances of dislocation in his study, we can apply them more strictly to identify cases of subject doubling in Resian (see below).

Since examples such as those from standard and colloquial French in (2) are semantically and pragmatically equal, they can be regarded as translations of each other. Mutatis mutandis, it can be argued that whenever Resian doubling constructions are used to translate non-doubling constructions in the French original of *The Little Prince*, they are semantically and pragmatically equal because they occur in exactly the same context. Needless to say, this argument is only plausible if the number of examples in which non-doubling structures are conveyed by doubling structures is large enough. With regard to the investigated resource, this is clearly the case since about one-third of the 1,500 sentences (in total) of the Resian translation of *The Little Prince* exhibit subject doubling.

As mentioned above, Coveney (2005), in his investigation of spoken French, also includes instances of dislocation, although he does not use a specific term to distinguish them from other doubling constructions. Since the present study is not interested in instances of dislocation, several examples had to be excluded from the analysis. On one hand, this concerns instances where the French original contains structures consisting of a subject and a coreferential subject pronoun. In these cases, it cannot be excluded that the use of the doubling construction in Resian is influenced by the original text. On the other hand, I have excluded examples which contain some indication of an intonation break in the Resian translation. This concerns instances where subjects are separated from doubling clitics by commas or ellipses. To verify whether structures such as those investigated in this paper indeed do not require intonation breaks, I additionally checked the pronunciation of several example recordings which were generously provided to me by Malinka Pila (including examples with and without subject-verb inversion, cf. §5.3). Therefore, I believe that the diagnostics applied in this paper are sufficiently reliable for identifying instances of subject

doubling in Resian, even though the limitations of working with written text must be acknowledged.

## 5.2. Types of Associates

In §1, we have already encountered two examples for the doubling of tonic personal pronouns. However, the phenomenon is not limited to this part of speech. The purpose of this section is to determine the different associates attested with doubling in the investigated text. In doing so, I not only give examples for the different parts of speech that are doubled but also consider different genders and try to provide examples for animate and inanimate subjects if attested. Most of the existing studies provide only a limited number of examples (see Benacchio 2002; Skubic 1997; Steenwijk 1992; but cf. Šekli 2010a), which is why I believe that a larger set of examples will be beneficial for future research. For instance, it has been noted by Poletto (2008) that across northern Italian dialects, some tonic pronouns are more readily doubled than others. Therefore, providing examples for doubling with different tonic pronouns is relevant for identifying the source language of the doubling in Resian. Moreover, I provide examples for noun phrases that are modified by definite or indefinite articles. It is important to include these features because clitic doubling has been reported to be influenced by specificity, definiteness, and animacy in other languages (see §2).

Before I discuss doubling with different parts of speech in separate sections below (sections 5.2.1–5.2.3), a few general notes on agreement in Resian subject doubling are in order. Resian subject clitics agree with their associates in number, and in the third-person singular also in gender (cf. the examples cited throughout this paper).[9] In the dual, where no separate subject clitics are available, the plural form of the third-person clitic is used (see §5.2.3, and see also §5.4). The same form is also used with coordinated phrases. Thus, in example (4), *ni* 'they' doubles the coordinated noun phrases *dizenj nümër 1 anü nümër 2* 'drawing number 1 and number 2'.[10]

[9] Gender differences have been lost in the plural form of the third-person subject clitic (see Table 3 on p. 303).

[10] Page numbers given in parentheses after the English translations of glossed examples refer to the Resian translation of *The Little Prince*.

(4) […] ko si vïdël da **möj**
when AUX.1SG see.PST.PTCP.SG.M COMP my.NOM.SG.M

**dizenj** **nümër** **1**
drawing.NOM.SG.M number.NOM.SG.M one

**anö** **nümër** **2** **ni** ni grejo.
and number.NOM.SG.M two they.SC NEG go.PRS.3PL

'[…] when I saw that my drawings number 1 and number 2 did not work out.' (p. 10)

Apart from grammatical gender and number agreement, we also find one instance of plural agreement *ad sensum*. In example (23), the quantifier phrase *malö njyh* 'few of them' is doubled by the nominative plural subject clitic *ni* 'they'. Moreover, there are several instances of the neuter third-person subject clitic *to* occurring with masculine and feminine noun phrases. In these cases, the subject clitic most probably functions as an expletive (cf. Runić 2014: 11, and further, Schuster-Šewc 1974: 349–50).

### 5.2.1. Nouns

In the investigated text, subject doubling is attested with masculine and feminine nouns, regardless of whether they refer to animate or inanimate subjects. Representative examples for doubling with unmodified nouns can be found in examples (5a–f). Example (5b) further shows that doubling occurs with foreign nouns.

(5) a. **Krej** z no roko **an**
king.NOM.SG.M with INDEF.ACC.SG.F hand.ACC.SG.F he.SC

mu löpö pokazal swöj
he.OC.DAT meaningfully show.PST.PTCP.SG.M his.ACC.SG.M

planët […]
planet.ACC.SG.M

'And with a meaningful gesture, the king showed him his planet […]' (p. 38)

b. **Aštronomo** **an** spet naredil
astronomer.NOM.SG.M he.SC again hold.PST.PTCP.SG.M

swo dimunstracjun tu-w lëtö 1920 […]
his.ACC.SG.F lecture.ACC.SG.F in year.LOC.SG.N 1920

'The astronomer held his lecture again in the year 1920 […]' (p. 19)

(5) c. **Pïsak** **an** jë to barbo
sand.NOM.SG.M he.SC be.PRS.3SG DEF.NOM.SG.N color.NOM.SG.N
od stärde.
of honey.GEN.SG.F
'The sand has the color of honey.' (p. 81)

d. Ma **lisïca** **na** paršlä spet na
but fox.NOM.SG.F she.SC come.PST.PTCP.SG.F again on
nji idëjo.
her idea.ACC.SG.F
'But the fox came back to its idea.' (p. 68)

e. Dan din **roža** **na**
day.ACC.SG.M INDEF.ACC.SG.M flower.NOM.SG.F she.SC
bila vïdala prijtyt no
AUX.PST.PTCP.SG.F see.PST.PTCP.SG.F pass.INF INDEF.ACC.SG.F
rïjo wözöw.
row.ACC.SG.F wagon.GEN.PL.M
'One day the flower had seen a caravan passing by.' (p. 62)

f. **Utrucy** **ni** mäjo mët
child.NOM.PL.M they.SC have.PRS.3SG have.INF
pacincjo ziz timi vlïkimi
patience.ACC.SG.F with DEF.INS.PL grown.up.INS.PL
judin.
people.INS.PL.M
'Children must have patience with grown-ups.' (p. 20)

Instances of doubled neuter noun phrases are more difficult to identify because Resian uses the neuter subject clitic *to* also as an expletive (Schuster-Šewc 1974: 349–50). Thus, in example (6), *to* is ambiguous. The only conclusion that can be drawn is that the use of the subject clitic is not the result of a loan translation. It does not have a counterpart in the French original.

(6) Anö **hawfanjë** za-me **to** jë na
and yawning.NOM.SG.N for-me it.SC be.PRS.3SG INDEF.NOM.SG.F
rič kurjöws.
thing.NOM.SG.F curious
'And yawning, for me it is something interesting.' (p. 37)

It has been argued on the basis of Romance and Germanic languages that subject doubling is related to definiteness (de Vogelaer and Neuckermans 2002). In Resian, which has developed a definite article from a demonstrative pronoun and an indefinite article from the numeral 'one' (Steenwijk 1992: 125–26), this does not seem to be the case, at least not when it comes to the mere realization of doubling. We find instances of doubling both with noun phrases modified by definite articles and noun phrases modified by indefinite articles. Representative cases of subject doubling with noun phrases modified by a definite article can be found in examples (7a–c). In examples (7d–e), noun phrases modified by indefinite articles are doubled. This includes instances of doubling with animate and inanimate subjects.

(7) a. Alore **te** **marćadänt** **an**
so DEF.NOM.SG.M merchant.NOM.SG.M he.SC
kapyl […]
understand.PST.PTCP.SG.M
'So the merchant understood […]' (p. 47)

b. Ma **te** **bašlën** **an** naprid
but DEF.SG.M shrub.NOM.SG.M he.SC soon
žmëtinal rast […]
stop.PST.PTCP.SG.M grow.INF
'But the shrub soon stopped growing […]' (p. 30)

c. Dobro nuć! **na** mu
good.ACC.SG.F night.ACC.SG.F she.SC he.OC.DAT
raklä **ta** **kača**.
say.PST.PTCP.SG.F DEF.NOM.SG.F snake.NOM.SG.F
'Good night!, said the snake.' (p. 59)

d. Ko **dan** **aštronomo** **an**
if INDEF.NOM.SG.F astronomer.NOM.SG.M he.SC
škuprïjë dan planët
discover.PRS.3SG INDEF.ACC.SG.M planet.ACC.SG.M
od njyh an mu danë za
of they.GEN.PL.M he.SC it.OC.DAT give.PRS.3SG for
jïmë dan nümër.
name.ACC.SG.N INDEF.ACC.SG.M number.ACC.SG.M
'When an astronomer discovers a planet from them, he gives it a number as a name.' (p. 18)

(7) e. **Dan** **din** izdë **an** dura
day.NOM.SG.M INDEF.NOM.SG.M here he.SC last.PRS.3SG
dan minüt.
INDEF.ACC.SG.M minute.ACC.SG.M
'A day lasts one minute here.' (p. 52)

There are also numerous examples in which the head nouns of the doubled subject noun phrases are modified by a definiteness marker and an adjective.[11] Although this may not be surprising given the occurrence of doubling with other parts of speech, it is still noteworthy. Runić's (2014: 138–78) formal analysis suggests that Resian definiteness markers, referred to as weak pronouns, possess all the functions typically associated with definite articles only when occurring with modified nouns. The investigated data show that these differences are not significant enough to influence the general possibility of the doubling. Examples (8a–c) contain instances of doubling with modified noun phrases marked as definite, and in examples (8d–f), the doubled noun phrases contain indefinite articles.

(8) a. **Te** **mali** **prïncip** **an**
DEF.SG.M little.NOM.SG.M prince.NOM.SG.M he.SC
nï rišpundäl nikar.
NEG.AUX.3SG answer.PST.PTCP.SG.M nothing.ACC
'The little prince did not answer anything.' (p. 60)

b. **Te** **petnji** **planët** **an** bil
DEF.SG.M fifth.NOM.SG.M planet.NOM.SG.M he.SC be.PST.PTCP.SG.M
fys kurjöws.
really curious
'The fifth planet was very curious.' (p. 50)

c. **Ti** **vlïki** **jüdi** **ni** ni
DEF.NOM.PL big.NOM.PL people.NOM.PL they.SC NEG
kapïjo mej nikar sami […]
understand.PRS.3PL never nothing.ACC self.NOM.PL
'The big people never understand anything by themselves […]' (p. 10)

[11] However, note that in most cases, this concerns the noun phrase *te mali prïncip* 'the little prince'.

(8) d. […] anö ka **na** **mala** **wöwca**
and that INDEF.NOM.SG.F small.NOM.SG.F sheep.NOM.SG.F

**na** mörë dan sami nur
she.SC can.PRS.3SG one.ACC.SG.M self.ACC.SG.M time.OC.ACC

jo pužrët […]
she devour.INF

'[…] and that a small sheep can devour her at once […]' (p. 29)

e. […] ka **dan** **türski** **ditatör**
when INDEF.NOM.SG.M Turkish.NOM.SG.M dictator.NOM.SG.M

**an** wkazal svën judin, da
he.SC order.PST.PTCP.SG.M his.DAT.PL people.DAT.PL COMP

wsi mäjo se ublić tej ti
all.NOM.PL have.PRS.3PL REFL dress.INF as DEF.NOM.PL

jüdi tu-w Ewropi […]
people.NOM.PL in Europe.LOC.SG.F

'[…] when a Turkish dictator ordered his people to dress like the people in Europe […]' (p. 19)

f. **Na** **taka** **möćnöst** **na**
INDEF.NOM.SG.F such.NOM.SG.F power.NOM.SG.F she.SC

čüdila toga malaga prïncipa.
surprise.PST.PTCP.SG.F DEF.ACC.SG.M little.ACC.SG.M prince.ACC.SG.M

'Such power surprised the little prince.' (p. 38)

Subject doubling also occurs with substantivized adjectives. The only two instances of this kind that I came across in *The Little Prince* are examples (9a–b). In both cases, the adjective is modified by a definite article. However, it is not clear to me whether substantivization of adjectives modified by indefinite articles is possible outside predicative constructions. Steenwijk (1992: 114) only provides examples where the adjective is part of the predicate.

(9) a. **Ti** **vlïki** **ni** so mi
DEF.NOM.PL big.NOM.PL they.SC AUX.3PL I.OC.DAT

konsajali […]
advise.PST.PTCP.PL

'The grown-ups advised me […]' (p. 10)

(9) b. […] **ti** **drügi** **ni** so maköj
DEF.NOM.PL other.NOM.PL they.SC AUX.3PL only
za je hwalit.
for they.OC.ACC praise.INF
'[…] they are there only to praise them.' (p. 43)

Instances of subject doubling with subject phrases containing nouns modified by demonstratives are given in examples (10a–e). Unsurprisingly, they include instances where the doubled phrase, apart from a demonstrative, contains an attributive adjective; cf. (10e).

(10) a. **Isi** **šlovëk** **an** bil itakö
this.NOM.SG.M man.NOM.SG.M he.SC be.PST.PTCP.SG.M so
wjet […]
busy.NOM.SG.M
'This man was so busy […]' (p. 46)

b. […] anö **isi** **baobab** **ni** so privić […]
and this.NOM.PL baobab they.SC be.PRS.3PL too.many
'[…] and these baobabs are too many […]' (p. 23)

c. Ma **isa** **roža** **na** bocala
but this.NOM.SG.F flower.NOM.SG.F she.SC come.up.PST.PTCP.SG.F
dan din od ne saminče […]
day.ACC.SG.M one.ACC.SG.M from INDEF.GEN.SG.F seed.GEN.SG.F
'But this flower came up one day from a seed […]' (p. 30)

d. **Isö** **mëstö** **to** jë tej iti
this.NOM.SG.N place.NOM.SG.N it.SC be.PRS.3SG like that.NOM.PL
od ite paǵine prit […]
from that.GEN.SG.F page.GEN.SG.F before
'This place is the same as those on the page before […]' (p. 94)

e. **Ise** **privlïke** **dižmirïne** **ni**
this.NOM.PL very.big.NOM.PL boa.constrictor.NOM.PL they.SC
žrëjo cëlaga brawa […]
swallow.PRS.3PL entire.ACC.SG.M prey.ACC.SG.M
'These huge boa constrictors swallow the entire prey […]' (p. 9)

Noun phrases modified by possessive pronouns are also doubled; cf. (11a–g). This includes phrases modified by the non-agreeing third-person possessive pronouns; cf. (11d and 11g).

(11) a. **Möj** **amïg** **an** se nasmëjnöl […]
my.NOM.SG.M friend.NOM.SG.M he.SC REFL smile.PST.PTCP.SG.M
'My friend smiled […]' (p. 14)

b. **Ma** **hïša** **na** skrïwala
my.NOM.SG.F house.NOM.SG.F she.SC hide.PST.PTCP.SG.F
no bogatïjo tu-w svin särcö…
INDEF.ACC.SG.F riches.ACC.SG.F in its.LOC.SG.N heart.LOC.SG.N
'My house hid riches in its heart…' (p. 77)

c. **Waš** **špinjulët** **an** jë
your.NOM.SG.M cigarette.NOM.SG.M he.SC be.PRS.3SG
wgasnën.
go.out.NOM.SG.M
'Your cigarette has gone out.' (p. 46)

d. Ma **njagä** **planët** **an** jë fys mäsa
but his planet.NOM.SG.M he.SC be.PRS.3SG really too
dan mali.
INDEF.NOM.SG.M small.NOM.SG.M
'But his planet was too small.' (p. 53)

e. **Swa** **rožica** **na** bila mu
his.NOM.SG.F flower.NOM.SG.F she.SC AUX.PST.PTCP.SG.F I.OC.DAT
raklä, da […]
say.PST.PTCP.SG.F COMP
'His flower had told him […]' (p. 64)

f. Anö **twujy** **amïgavi** **ni** ćejo se
and your.NOM.PL friend.NOM.PL they.SC AUX.FUT.3PL REFL
fys čüdit za te vïdët […]
really be.surprised.INF for you.OC.ACC.SG see.INF
'Your friends will be really surprised to see you […]' (p. 87)

(11) g. Itakü **nji** **žïwjöst** **na** bo
thus their life.NOM.SG.F she.SC be.FUT.3SG
wezana na two juštïcijo.
bind.PST.PTCP.PASS.SG.F to your.ACC.SG.F jurisdiction.ACC.SG.F
'Thus their life will be bound to your justice.' (p. 41)

Subject doubling does not only occur with subjects specified by adjectival attributes. In *The Little Prince*, we find the following two instances of subject doubling with noun phrases modified by attributes expressed by prepositional phrases (12a–b). A doubled subject noun phrase containing an apposition, i.e., *müj dizenj nümër 2* 'my drawing number 2', is given in example (12c).

(12) a. **Zëmja** **od** **planëta** **na** bila
earth.NOM.SG.F of planet.GEN.SG.M she.SC be.PST.PTCP.SG.F
basana.
full.NOM.SG.F
'The ground of the planet was full.' (p. 23)

b. **Idëja** **od** **tröp** **lafantöw** **na** se
idea.NOM.SG.F of herd elephant.GEN.PL.M she.SC REFL
zdëlala smëjat timu malimö
make.PST.PTCP.SG.F laugh.INF DEF.DAT.SG.M little.DAT.SG.M
prïncipö.
prince.DAT.SG.M
'The idea of a herd of elephants made the little prince laugh.' (p. 22)

c. **Möj** **dizenj** **nümër** **2** **an**
my.NOM.SG.M drawing.NOM.SG.M number.NOM.SG.M 2 he.SC
jë bil itakö:
AUX.3SG be.PST.PTCP.SG.M like.that
'My drawing number 2 was like that:' (p. 10)

### 5.2.2. Personal Pronouns

In §1, two examples containing doubling of masculine and feminine forms of the tonic third-person personal pronoun were given in (1). However, doubling is not limited to the third person. We find reliable instances also in the second-person singular and plural. Examples containing the third- and second-person personal pronouns are given in (13a–e). I have not come across an instance of doubling with the feminine plural pronoun *wune* 'they', which

occurs only once in the investigated text. The neuter singular pronoun *onö* 'it' is absent from the text.

(13) a. Ma **tï** **ti** boš ji
but you.NOM.SG you.SC AUX.FUT.2SG she.OC.DAT
naredil wsaki vijäč no
make.PST.PTCP.SG.M every.ACC.SG.M time.ACC.SG.M INDEF.ACC.SG.F
gračijo za jo sparnjät.
grace.ACC.SG.F for she.OC.ACC save.INF
'But you will reprieve her every time to save her.' (p. 41)

b. Anö **wun** **an** dï.
and he.NOM he.SC say.PRS.3SG
'And he says.' (p. 90)

c. Anö **wonä** **na** jë tulïkö žboh.
and she.NOM she.SC be.PRS.3SG so weak.NOM.SG.M
'And she is so weak.' (p. 90)

d. Ma **vï** **vi** se ni samajata fys
but you.NOM.PL you.SC REFL NEG resemble.PRS.2PL really
nikar mimu gorofulö […]
nothing.ACC my.DAT.SG.M rose.DAT.SG.M
'But you do not resemble my rose at all […]' (p. 72)

e. **Wuny** **ni** ni durijavawajo si mïslit
they.NOM they.SC NEG be.capable.PRS.3PL REFL imagine.INF
po itëj hïši.
about this.LOC.SG.F house.LOC.SG.F
'They are not capable of picturing this house.' (p. 20)

Instances of doubling with first-person personal pronouns are uncertain. In *The Little Prince*, examples in the singular are limited to cases where the tonic variant *jäs* is separated from the rest of the clause by intonation breaks marked with commas. Moreover, *jäs* in these instances translates French *moi*, which is also detached from the clause (e.g., example 14).

(14) **Ja** dilan, **jäs**, poštëne rëči!
I.SC do.PRS.1SG I.NOM important.ACC.PL thing.ACC.PL
'I do important things!' (p. 28)

The analysis of these structures as non-doubling additionally seems to be confirmed by example (15). Here, we find the tonic personal pronoun repeated while no subject clitic is present in the clause.[12]

(15) Ma **jäs** si poštën, **jäs**.
but I.NOM be.PRS.1SG serious.NOM.SG.M I.NOM
'But I am serious.' (p. 47)

However, it should be noted that similar instances of doubling with the first-person personal pronoun are also attested in Steenwijk's (1992: 201, 203) transcripts. Therefore, further research is needed to determine whether doubling of this form might not be possible after all.

The sole instance of the nominative plural *mï* 'we' does not provide a clear example of doubling. In example (16), *mï* is repeated by the subject clitic *mi*, but this is true not only for the Resian text but also the French original (cf. *nous… nous*). Therefore, following the methodology laid out in §5.1, example (16) cannot be classified as an instance of doubling.

(16) Ma, sigür, **mï** ki kapïmo žïwjöst, **mi**
but of.course we.NOM REL understand.PRS.1PL life.ACC.SG.F we.SC
nïsimö interešani na nümarje!
NEG.be.PRS.1PL interested.NOM.PL on numbers.ACC.PL
'But, of course, we who do understand life are not interested in numbers!' (p. 20)

### 5.2.3. Other Parts of Speech

Other parts of speech that occur with doubling in *The Little Prince* are demonstrative and relative pronouns, quantifiers, and numerals. Instances of doubling with demonstratives are given in (17).

(17) a. […] **isi** **an** ražona nu
this.NOM.SG.M he.SC be.similar.PRS.3SG INDEF.ACC.SG.N
malö tej möj pijancar.
little.ACC.SG.N like my.NOM.SG.M drunkard.NOM.SG.M
'[…] this one thinks like my drunkard a little bit.' (p. 48)

[12] An anonymous reviewer suggested that *jä(s)* in the above examples could be compared to emphatic pronouns in Italian, cf. *Giovanni interviene lui* 'Giovanni himself intervenes' (see Burzio 1986: 109–15 and Cardinaletti 1999 for two different interpretations of this phenomenon).

(17) b. **Isa** **na** jë wžë karjë bulna.
this.NOM.SG.F she.SC be.PRS.3SG already very ill.NOM.SG.F
'This one is already very ill.' (p. 14)

c. **Isö** **to** jë sćatula.
this.NOM.SG.N it.SC be.PRS.3SG box.NOM.SG.F
'This is a box.' (p. 14)

Subject clitics further occur with the two uninflected relative markers *ki* and *ka*; cf. (18a–g).[13] In this case, subject clitics agree with the antecedents of the relative clauses in number and gender. It is interesting to note that this use of subject clitics with uninflected relative markers resembles the standard Slovenian model of employing resumptive pronouns for marking oblique cases with the relative marker *ki* (see Herrity 2016: 146, and cf. also §5.5 below).

(18) a. Anö pa ešploradür, **ka** **an** pïjë
and also explorer.NOM.SG.M REL he.SC drink.PRS.3SG
mäsa muć.
too much
'And also an explorer who drinks too much.' (p. 55)

b. Itö jë stal dan stari
there AUX.PRS.3SG live.PST.PTCP.SG.M INDEF.NOM.SG.M old.NOM.SG.M
muž, **ka** **an** pïsal ne
man.NOM.SG.M REL he.SC write.PST.PTCP.SG.M INDEF.ACC.PL
privlïke lïbrine.
very.big.ACC.PL book.ACC.PL
'There lived an old man who wrote voluminous books.' (p. 53)

c. Si vïdël no hïšo, **ki**
AUX.PRS.1SG see.PST.PTCP.SG.M INDEF.ACC.SG.F house.ACC.SG.F REL
**na** vala stu mijarjüw
she.SC be.worth.PRS.3SG hundred.NOM.SG.N thousand.GEN.PL
franköw.
francs.GEN.PL
'I saw a house that is worth one hundred thousand francs.' (p. 20)

[13] Note that the latter element also introduces other subordinate clauses (Steenwijk 2005: 56).

(18) d. Na rožica, **ka** **na** nï
INDEF.NOM.SG.F flower.NOM.SG.F REL she.SC NEG.AUX.3SG
valala nikar…
be.worth.PST.PTCP.SG.F nothing.ACC
'A flower that was not worth anything…' (p. 62)

e. Anü ja mürën dovantät tej ti vlïki
and I.NOM can.PRS.1SG become.INF like DEF.NOM.PL big.NOM.PL
jüdi, **ki** **ni** se interašajo köj
people.NOM.PL REL they.SC REFL be.interested.PRS.3PL only
na nümarje.
in number.ACC.PL
'And I can become like the grown-up people which are interested only in numbers.' (p. 20)

f. Mujy amïgavi, **ki** **ni** so spet
my.NOM.PL.M friend.NOM.PL.M REL they.SC AUX.3PL again
me vïdali […]
I.OC.ACC see.PST.PTCP.PL
'My friends which saw me again […]' (p. 91)

g. An löpö spücinal te vulkane,
he.SC properly sweep.PST.PTCP.SG.M DEF.ACC.PL volcano.ACC.PL
**ki** **ni** gurïjo ta-znutrë.
REL they.SC burn.PRS.3PL inside
'He properly sweeped the volcanos which burn inside.' (p. 34)

As these examples show, doubling does not depend on the type of relative clause introduced by the relative marker. We find doubling in both restrictive and non-restrictive relative clauses.

In one instance, the components of the correlative construction 'one… the other' are doubled; cf. (19). However, the correlatives *dan* 'one' and *drügi* 'other' are used attributively in this example since they modify the noun *dizenj* 'drawing'. It could not be verified based on the investigated resource whether doubling is possible with bare correlatives.

(19) **Dan** **dizenj** **an** gre prow,
one.NOM.SG.M drawing.NOM.SG.M he.SC go.PRS.3SG right.NOM.SG.M
ma **te** **drügi** **an** se ni samaja
but DEF.NOM.SG.M other.NOM.SG.M he.SC REFL NEG resemble.PRS.3SG
vić.
more
'One drawing goes along alright, but the other does already not resemble the subject.' (p. 21)

There are several instances of doubling with quantifier phrases in *The Little Prince*. Example (20) contains cases with the quantifiers *väs* 'all; whole' and *wsaki* 'every(one)', together with a noun, cf. (20a–b, 20d), or a demonstrative (20c).

(20) a. **Väs** **svit** **an** ga
whole.NOM.SG.M world.NOM.SG.M he.SC he.OC.GEN
nïmël.
NEG.have.PST.PTCP.SG.M
'The whole world does not have him.' (p. 20)

b. **Wse** **zwïzde** **ni** ćejo byt ne
all.NOM.PL star.NOM.PL they.SC AUX.FUT.3PL be.INF INDEF.NOM.PL
päče ziz no rïjawo karïkulo.
wells.NOM.PL with INDEF.INS.SG.F rusty.INS.SG.F winch.INS.SG.F
'All the stars will be wells with rusty winches.' (p. 89)

c. Alore **wsë** **isö** **to** nareja
so all.NOM.SG.N this.NOM.SG.N it.SC make.PRS.3SG
petstu […]
five.hundred.ACC.SG.N
'So all this makes five hundred […]' (p. 47)

d. […] da **wsaki** **šlovëk** dan
COMP every.NOM.SG.M human.NOM.SG.M day.ACC.SG.M
din **an** mörë nalëst swo.
one.ACC.SG.M he.SC can.PRS.3SG find.INF his.ACC.SG.F
'[…] so that one day every man can find his.' (p. 60)

There is also one possible instance of doubling of a bare universal quantifier. However, it should be noted that the quantifier in example (21) is referring to *mak* 'pile' and thus could be interpreted as elliptic. Note further that the object clitic genitive-plural *jih* 'they' is not an attribute of the quantifier but a genitive

argument of the verb *mä* 'has'. The attribute use of the tonic genitive-plural *njyh* can be found in (23).

(21) Ja divïdinän jüdi w make, anö
I.NOM sort.PRS.1SG people.ACC.PL in pile.ACC.PL and
**wsaki an** jih mä dän
every.NOM.SG.M he.SC they.OC.GEN have.PRS.3SG one.ACC.SG.M
mijar.
thousand.ACC.SG.M
'I sort the people in piles, and every pile has one thousand people.' (p. 74)

A clear example of doubling of a universal quantifier is cited by Runić (2018: 8) from the Resian local newspaper *Näš glas* (cf. also the example given by Šekli 2010a: 149). It is reproduced in (22). Here, the quantifier *wsak* 'everyone' is further specified by the prepositional phrase *ta-mi nami* 'among us'.

(22) **[W]sak ta-mi nami an** ma wojo
everyone.NOM.SG.M among we.INS he.SC have.PRS.3SG wish.ACC.SG.F
si wdäržat itö ka an si čüjë
REFL conserve.INF that.ACC.SG.N REL he.SC REFL feel.PRS.3SG
'[E]veryone of us has the wish to conserve what he is feeling'[14]

The only instance with a quantifier other than *väs* and *wsaki* that I came across is example (23). Here, the quantifier *malö* 'few', which is further specified by genitive plural *njyh* 'they', is doubled by the subject clitic *ni* 'they'. The subject clitic agrees with the quantifier *ad sensum* since the latter has a neuter singular ending.

(23) Ma **malö njyh ni** se spomänjajo.
but few they.GEN.PL they.SC REFL remember.PRS.3PL
'But few of them remember.' (p. 5)

Two instances of doubled bare numerals have been encountered in *The Little Prince*. In example (24a), the numeral 'one' translates French *on* 'one' and thus acts as the subject of an impersonal clause. Therefore, this example does not present a case of doubling with a true numeral. In the second instance, the masculine form of the numeral *dwa* 'two' is doubled; cf. (24b). Due to the

[14] Cf. Runić's (2018: 8) Croatian translation: "Svi među nama imaju želju da sačuvaju ono što osećaju".

lack of a dual subject clitic, the plural form *ni* is employed in this construction. Likewise, the finite verb of the clause occurs in the third-person plural.

(24) a. To karjë pomaga, ći **dän** **an**
that.NOM.SG.N a.lot help.PRS.3SG when one.NOM.SG.M he.SC
se zgubï tu-w nićy.
REFL be.lost.PRS.3SG in night.LOC.SG.F
'That helps a lot when you are lost in the night.' (p. 10)

b. **Dwa** **ni** fys gurïjo, anö
two.NOM.DU.M they.SC.PL proper burn.PRS.3PL and
dän jë wgasnën.
one.NOM.SG.M be.PRS.3SG expire.PST.PTCP.PASS
'Two which are properly burning and one which is expired.' (p. 56)

## 5.3. The Place of the Doubling Clitic

Having discussed the types of phrases that occur with doubling in the previous section, I now turn to the place that the subject clitic occupies in clauses with subject doubling. The reference points for determining the position of the clitics will be (i) the finite verb and (ii) the associate of the doubling clitic. As for the latter position, it is not expected that there will be any deviations from the placement of subject clitics in clauses without doubling if they are indeed oriented towards the verb, and this is exactly what we find in the investigated text. Subject clitics precede the highest-ranking verb even in clauses with subject-verb inversion. Examples for this placement are given in the discussion of the relative position of subject clitics and their associates; cf. (32a–c, 33a–e, 34, 35, and 36). They clearly show that subject clitics are more restricted in their placement compared to their tonic counterparts as well as compared to lexical subjects which follow the verb in clauses with inversion. Although I did not encounter differences in the relative placement of subject clitics and the finite verb in clauses with and without doubling, in the following, I provide several examples attesting various positions of the clitics in clauses with doubling. Since there are at present no studies dealing with clitic placement in Resian specifically (cf. §4), the provided examples have a documentary value. Moreover, it is not excluded that other resources might reveal differences between the placement of subject clitics in clauses with and without doubling.

As mentioned in §4, Resian subject clitics do not have to precede the finite verb immediately. In clauses with and without doubling alike, they can be separated from it by particles (PTCL), reflexive markers (REFL), object clitics (OC),

and negation markers (NEG). Examples with these elements occurring between the subject clitic and the finite verb are given in (25a–c). Note that the particle *ba* is glossed with the label ATT since, according to Steenwijk (1992: 186), it has attenuative meaning.

(25) a. […] anö pa ći **wun** **an** ba
and EMPH if he.NOM.SG he.SC ATT
vignäl dan cëli
drive.out.PST.PTCP.SG.M INDEF.ACC.SG.M entire.ACC.SG.M
tröp lafantöw […]
flock.ACC.SG.M elephant.GEN.PL.M
'[…] and even if he would drive out an entire flock of elephants […]' (p. 22)

b. […] da **wonä** **na** se rädi härći […]
COMP she.NOM she.SC REFL gladly praise.PRS.3SG
'[…] that she gladly praises herself […]' (p. 31)

c. **Wčinïca** **na** mi ni sarvija.
wheat.NOM.SG.F she.SC I.OC.DAT NEG serve.PRS.3SG
'Wheat is useless for me.' (p. 69)

Starting with subject clitics and ending with the finite verb or the *l*-participle in past-tense constructions lacking auxiliaries, the tentative sequence in (26) can be proposed. However, it must be noted that this sequence should not be considered a clitic cluster since it can be split at various places by tonic elements. I intend to address these issues in more detail elsewhere (Wandl, forthcoming). For now, it is sufficient to state that subject clitics always precede the highest-ranking verb, regardless of whether the subject is doubled or not.

(26) SC + PTCL + OC + REFL + NEG + finV/*l*-PTCP

The relative position of subject clitics and their associates depends on the order of the subject and the verb. In unmarked declarative sentences, subject clitics follow their associates. To be precise, they follow the entire subject phrase, including postponed genitive or prepositional attributes (e.g., example (27) with a prepositional attribute further modified by a genitive attribute).

(27) **Gnüwanjë** **od** **ise** **tropade** **judi**
movement.NOM.SG.N of this.GEN.SG.F army.GEN.SG.F people.GEN.PL
**to** bilö däržanö […]
it.SC be.PST.PTCP.SG.N hold.PST.PTCP.PASS.NOM.SG.N
'the movement of this army of people was organized […]' (p. 58)

However, we also find instances where subject clitics and their associates are separated from each other by constituents that are not part of the subject phrase. More specifically, it is adverbials that occur between the associates and the subject clitics. Examples (28a–d) contain adverbials with clausal scope in the discussed position. In example (28e), on the other hand, the associate and the subject clitic are separated from each other by the prepositional phrase *z no roko* 'with one hand', which serves as an adjunct of the verb *pokazati* 'show'.

(28) a. **Te** **mali** **princïp** injän **an**
DEF.NOM.SG.M little.NOM.SG.M prince.NOM.SG.M now he.SC
bil fys smlid za wümon.
be.PST.PTCP.SG.M really pale.NOM.SG.M from anger.INS.SG.M
'Now, the little prince was really pale from anger.' (p. 29)

b. **Rožice** prid **ni** ti rumunïjo
flower.NOM.PL at.first they.SC.PL you.OC.DAT.SG tell.PRS.3PL
dnö, döpö ni ti ričëjo
one.ACC.SG.N then they.SC.PL you.OC.DAT.SG say.PRS.3PL
nö drügë.
INDEF.ACC.SG.N other.ACC.SG.N
'First, the flowers tell you one thing and then, they say something else.' (p. 33)

c. **Te** **mali** **prïncip** wsako
DEF.NOM.SG.M little.NOM.SG.M prince.NOM.SG.M every.ACC.SG.F
nuć **an** diwa to rožo
night.ACC.SG.F he.SC put.PRS.3SG DEF.ACC.SG.F flower.ACC.SG.F
ta-pod dan verinavi zwun!
under INDEF.ACC.SG.M of.glass.ACC.SG.M bell.ACC.SG.M
'Every night the little prince puts the flower under a bell made from glass!' (p. 93)

(28) d. ka **te** **mali** **prïncip** s
because DEF.NOM.SG.M little.NOM.SG.M prince.NOM.SG.M with
potjo **an** me baral, tej ba an
way.INS.SG.F he.SC I.OC.acc ask.PST.PTCP.SG.M as if he.SC
ni bëšë sigür.
NEG be.IMPERF.3SG sure.NOM.SG.M
'that the little prince, in passing, asked me as if he wasn't sure.' (p. 21)

e. **Krej** z no roko **an**
king.NOM.SG.M with INDEF.INS.SG.F hand.INS.SG.F he.SC
mu löpö pokazal swöj
he.OC.DAT meaningfully show.PST.PTCP.SG.M his.ACC.SG.M
planët […]
planet.ACC.SG.M
'And with a meaningful gesture, the king showed him his planet […]' (p. 38)

Yet another constituent occurring between associates and doubling subject clitics are prepositional phrases acting as verbal arguments. In example (29), *na te poštëne rëči* 'on these important things' is an argument of the verbal noun *vïdanjë* 'seeing'. Verbal nouns are not recognized as forming a distinct verb form in Resian by Steenwijk (1992: 127) due to their potentially unpredictable meaning. However, *vïdanjë* clearly demonstrates that some nouns which at least originally presented verbal nouns have retained their ability to govern prepositional phrases. Whether more prototypical arguments such as direct and indirect objects can be placed between associates and subject clitics could not be determined based on the investigated resource or the available secondary literature.

(29) **Te** **mali** **prïncip** na te
DEF.NOM.SG.M little.NOM.SG.M prince.NOM.SG.M on DEF.ACC.PL
poštëne rëči **an** mël fys
important.ACC.PL thing.ACC.PL he.SC have.PST.PTCP.SG.M entirely
nö drügë vïdanjë […]
INDEF.ACC.SG.N different.ACC.SG.N view.ACC.SG.N
'The little prince had an entirely different view on important things […]' (p. 49)

The lexical subject is also split from the subject clitic in example (30a). Here, the combination of the preposition *za* 'for' and the bound morpheme accusative

*me* 'me' is part of the predicative construction 'to be an interesting thing for somebody'. Thus, we are dealing with a different function compared to the prepositional phrases in examples (28d–e). The placement of the combination of preposition and bound pronominal form between the subject phrase and the subject clitic is, however, not obligatory. In example (30b), it follows the copula verb. A possible explanation for the placement of *za-me* between the subject phrase and the subject clitic in example (30a) is that *za-me* is highlighted to emphasize the contrast between the opinion of the speaker, the king, and others.

(30) a. Anö **hawfanjë** za-me **to** jë na
and yawning.NOM.SG.N for-me it.SC be.PRS.3SG INDEF.NOM.SG.F
rič kurjöws.
thing.NOM.SG.F curious
'And yawning, for me it is something interesting.' (p. 37)

b. **Ma** **zvïzda** **na** bo za-te
my.NOM.SG.F star.NOM.SG.F she.SC be.FUT.3SG for-you
dnä od zvizdi.
one.NOM.SG.F of star.GEN.PL
'My star will be one among many stars.' (p. 87)

Splitting of subject clitics and their associates seems to be less common with personal pronouns. The only instance I found in the investigated text is example (31). Here, the third-person personal pronoun *wun* 'he' is separated from the subject clitic *an* by the adverb *kadä* 'once'. Whether the scarcity of examples with pronouns is incidental or whether tonic subject pronouns and subject clitics are more prone to being grouped together could not be determined based on the investigated data.

(31) […] anö ka **wun** kadä **an** hudyl jïskat […]
and REL he.NOM once he.SC turn.PST.PTCP.SG.M search.INF
'[…] and which he once had gone to seek […]' (p. 52)

The subject phrase and the subject clitic are also separated from each other in cases of subject-verb inversion. This follows from the different degrees of freedom in the placement of subject clitics compared to their tonic counterparts and other subject phrases: While the latter can be placed before or after the verb, the former must precede it. Inversion is mostly attested in subordinate clauses and with direct speech complements. However, I can also report a few instances of inversion in main clauses. These comprise example (32a) and examples (32b–c), which show inversion in the second of two coordinated

clauses. As can be seen in example (32b), combinations of prepositional phrase and bound personal pronoun, e.g., *za-me* 'for me', can precede the subject clitic also in inversion.

(32) a. Möj lipi mali
my.NOM.SG.M beautiful.NOM.SG.M little.NOM.SG.M
mužyć, izdë **na** ni čentra vić
little.man.NOM.SG.M here she.SC NEG be.in.focus.PRS.3SG more
**lisïca**!
fox.NOM.SG.F
'My beautiful little man, the fox is not in the focus here anymore!' (p. 76)

b. Anö si jo pargnäl amïg anö
and be.PRS.1SG she.OC.ACC bend.PST.PTCP.SG.M friend and
injän za-me **na** jë köj **wonä** na
now for-me she.SC be.PRS.3SG only she.NOM in
svëtö.
world.LOC.SG.M
'And I have made her my friend and now she is the only one in the world for me.' (p. 72)

c. Itö, kë si stal, si
there where be.PRS.1SG live.PST.PTCP.SG.M be.PRS.1SG
mël no rožo: **na**
have.PST.PTCP.SG.M INDEF.ACC.SG.F flower.ACC.SG.F she.SC
römunila rüdi **wonä** ta pärwa…
speak.PST.PTCP.SG.F always she.NOM DEF.NOM.SG.F first.NOM.SG.F
'Where I lived, I had a flower: she always spoke first…' (p. 64)

Examples of inversion in subordinated clauses can be found in (33a–e). They include the relative conjunction *skod* 'where', the temporal conjunction *ko* 'when', the comparative conjunction *tej*, the complementizer *da* 'that', and the conditional conjunction *ći* 'if'.

(33) a. […] da planët, skod **an** parhaja
that planet.NOM.SG.M from.where he.SC come.PRS.3SG
**te** **mali** **prïncip** […]
DEF.NOM.SG.M little.NOM.SG.M prince.NOM.SG.M
'[…] that the planet from which the little prince came […]' (p. 19)

(33) b. […] ko **an** dušyl **te**
when he.SC arrive.PST.PTCP.SG.M DEF.NOM.SG.M

**mali** **prïncip**.
little.NOM.SG.M prince.NOM.SG.M

'[…] when the little prince arrived.' (p. 46)

c. Anö, tej **an** bil mi rëkal
and as he.SC AUX.PST.PTCP.SG.M I.OC.DAT say.PST.PTCP.SG.M

**te** **mali** **prïncip**, si
DEF.NOM.SG.M little.NOM.SG.M prince.NOM.SG.M AUX.1SG

dizanjäl iti planët.
draw.PST.PTCP.SG.M that.ACC.SG.M planet.ACC.SG.M

'And I drew this planet as the little prince had told me.' (p. 24)

d. Swa rožica na bila
his.NOM.SG.F flower.NOM.SG.F she.SC AUX.PST.PTCP.SG.F

mu raklä, da **na** jë
he.OC.DAT say.PST.PTCP.SG.F COMP she.SC be.PRS.3SG

sama **wonä** od sve sjorte
alone.NOM.SG.F she.NOM of its.GEN.SG.F kind.GEN.SG.F

ta na wsimö svëtö…
there in all.LOC.SG.M world.LOC.SG.M

'His little flower had told him that she was the only one of her kind in the entire world…' (p. 64)

e. Ći **an** bëšë jo mël
if he.SC AUX.IMPERF.3SG she.OC.ACC have.PST.PTCP.SG.M

**wun**, an mörëšë vïdët suncë jtyt
he.NOM he.SC can.IMPERF.3SG see.INF sun.ACC.SG.N go.INF

za göro ni štredi nu štikrat
behind mountain.ACC.SG.F NEG forty and four

čas, ma trïkrat dwisti nu dwanijst
time.GEN.PL.M but three-times twenty and twelve

čas […]
time.GEN.PL.M

'If he had had her, he could see the sun go behind the mountain not 44 times but 72 times […]' (p. 38)

An unusual case of inversion in a relative clause can be found in example (34). Here, the uninflected relative marker *ka* is followed by a resumptive subject clitic, which should be sufficient for expressing the syntactic relationship of the relative marker (cf. §5.2.3 for more examples of doubled relative markers). However, in this specific case, the relative clause also contains the tonic pronoun *wonä* 'she', following the predicative construction *jë maköj* 'be the only one'. Example (36) shows that the use of subject clitics is not obligatory in this type of construction. If it is further correct that, in example (36), *za-me* 'for me' and *za-te* 'for you' are moved to the beginning of the sentences to contrast them with each other, then the appearance of both a subject clitic and its tonic counterpart in example (34) may serve the purpose of highlighting the subject.

(34) Anö ći ja znan no rožo, ka
and if I.SC know.PRS.1SG INDEF.ACC.SG.F flower.ACC.SG.F REL
**na** jë maköj **wonä** na svëtö […]
she.SC be.PRS.3SG only she.NOM on earth.LOC.SG.M
'And if I know a flower which is the only one in the world […]' (p. 29)

Examples for quotative inversion can be found in examples (35a–c).

(35) a. "Hum! Hum", **an** mu rišpundal
hm hm he.SC he.OC.DAT answer.PST.PTCP.SG.M
**krej**.
king.NOM.SG.M
'Hm, hm, the king replied to him.' (p. 37)

b. "Ti si dan čüdni
you.SC be.PRS.2SG INDEF.NOM.SG.M peculiar.NOM.SG.M
braw", **an** ji
animal.NOM.SG.M he.SC she.OC.DAT
rëkal **wun** nejzad […]
say.PST.PTCP.SG.M he.NOM finally
'"You are a peculiar animal," he finally said to her […]' (p. 60)

c. "Bonǵorno", **an** mu rëkal **isi**.
good-day he.SC he.OC.DAT say.PST.PTCP.SG.M this.NOM.SG.M
'"Good day," he said to him.' (p. 46)

Moreover, inversion seems to occur in contrastive constructions. In example (36), *za-me* 'for me' is contrasted with *za-te* 'for you' in the next sentence. In both sentences, the combination of *za* plus bound personal pronoun (i.e., *-me*

and *-te*) is placed in initial position, only proceeded by a conjunction in the second sentence.

(36) Za-me **ti** ći byt maköj **tï** na
for-me you.SC.SG AUX.FUT.2SG be.INF only you.NOM.SG in
svëtö. Anö za-te ćon byt maköj jä
world.LOC.SG.M and for-you AUX.FUT.1SG be.INF only I.NOM
na svëtö…
in world.LOC.SG.M
'You will be the only one in the world for me. And I will be the only one in the world for you…' (p. 68)

It is interesting to note that in example (36), doubling of the tonic personal pronoun occurs only in the first of the two clauses. One possible explanation for this discrepancy is that the second-person subject clitic is more likely to occur with the future auxiliary than the first-person personal pronoun. This is because the form of the auxiliary is ambiguous in the case of the former, cf. second/third auxiliary *ći* but first auxiliary *ćon* (cf. Steenwijk 1992: 133 for the auxiliary paradigm).

In sentences with inversion, the following parts of speech have been found to occur between the subject clitic and the lexical verb: NEG, REFL, OC, AUX, and adverbials. Adverbials can occur both between subject clitics and verbs and between verbs and the associates of the subject clitics. Additionally, we find prepositional phrases acting as verbal arguments in the latter position. For further examples of the different word orders, the reader is referred to the examples given in §5.2.

## 5.4. Intermediate Summary

Before discussing Resian subject doubling as a case of language contact in §5.5, it is in order to give a summary of the findings so far. This is the purpose of the present section.

First of all, it is worth noting that although subject doubling has been described as optional in Resian (Šekli 2010a), at least in written language, it is far from a marginal phenomenon. We find repetition of subjects by the corresponding clitics in about one-third of the 1,500 sentences of *The Little Prince*. As for the phrases that function as associates of the doubling clitics, we find them to contain nouns (including substantivized adjectives), pronouns, and quantifiers as their heads (cf. also Šekli 2010a). Doubling of subjects containing nouns occurs regardless of whether they are modified by definite or indefinite articles, possessives, or demonstratives. Secure instances of doubling with personal pronouns are limited to the second- and third-person singular and

plural. Examples of co-occurring tonic and clitic nominative *jäs*/*jä* and subject clitic *ja* 'I', cf. (14), are rather to be explained as cases of dislocation. First, in these examples, the tonic pronoun is separated from the rest of the clause by commas indicating intonation breaks, and second, they translate French constructions with *moi*, which, according to Coveney (2005), are not equal to doubling constructions.

As regards the first-person plural pronoun, neither *The Little Prince* nor the secondary literature provides us with secure examples for doubling (Benacchio 2002; Skubic 1997; Steenwijk 1992). The repetition of the pronoun in example (16) mirrors the French original, which is why it is not a secure instance of doubling. However, it should be noted that this is the only occurrence of the tonic pronoun *mï* 'we' in the text. Doubling with second-person personal pronouns, on the other hand, is reliably attested in *The Little Prince*.

Doubling of demonstrative pronouns occurs in numerous examples, e.g., (17), regardless of their deixis (*isi* 'this', *iti* 'that'), gender, and number (apart from the dual). The same is true for the uninflected relatives *ki* and *ka*. Subject clitics occurring with these markers are co-referential with their antecedents. They resemble standard Slovenian resumptives used in relative clauses introduced by uninflected *ki*, except that in Standard Slovenian, resumptives are naturally limited to the oblique cases due to the lack of subject clitics in this language.

For determining the status of Resian subject doubling, examples with relatives are of special interest. As stated by Haiman and Benincà (1992: 172–73), relative pronouns cannot be dislocated. Thus, examples such as those cited in (18) provide evidence for the assumption that Resian does present a doubling language.

The other part of speech that is of special interested for the analysis of subject doubling in Resian are quantifiers. Runić (2018 with reference to Rizzi 1986; cf. also Brandi and Cordin 1989) argues that the doubling of quantifiers, cf. (22), proves that subject doubling in Resian is a genuine doubling phenomenon because quantifiers cannot be dislocated.[15] In *The Little Prince*, we find instances of doubling with both universal and existential quantifiers; cf. (20, 21, 23). In the latter case, the neuter singular quantifier *malö* 'few' modified by the genitive plural *njyh* 'they' is doubled by the plural subject clitic *ni* 'they'. Therefore, this example shows that subject clitics in doubling constructions do not necessarily have to agree with their associate in number. Agreement can also be *ad sensum*.

[15] However, note that this does not seem to be generally accepted since Poletto (2008: 45) states that universal quantifiers are more easily left dislocated than existential and negative quantifiers.

Doubling of quantifiers also includes numerals. In *The Little Prince*, it is attested with the numeral 'two'.[16] "This is of special interest because Resian (as standard Slovenian) has preserved the dual throughout all parts of speech apart from pronominal clitics. Accordingly, no specific dual clitic is available for doubling in this number. Šekli (2010a: 146) argues that the subject clitic used instead is the neuter singular form *to*. However, in example (24b), *dwa* 'two' is doubled by the plural clitic *ni* 'they'. Considering that *to* also acts as an expletive in Resian (cf. Schuster-Šewc 1974), the question arises whether it does not rather have this function in the examples analyzed by Šekli (2010a: 146). Thus, Runić (2014: 11) interprets the clitic *to* in the example repeated in (37) (glossing and translation adapted) as an expletive. Due to a lack of studies dealing with Resian expletives, this question has to remain unanswered here. At any rate, what example (24b) shows is that subjects in the dual can be doubled by the plural clitic *ni* 'they'.

(37) […] dwa mlada, muž anu
two.NOM.DU young.NOM.DU.M husband.NOM.SG.M and
žana, to bilu šlo kopat […]
wife.NOM.SG.F it.SC AUX.PST.PTCP.SG.N go.PST.PTCP.SG.N dig.INF
'[…] two young people, husband and wife, had gone to dig […]'

The large variety of different subject phrases that occur with doubling gives the impression that any type of phrase can be doubled in Resian. However, there are also parts of speech that do not seem to allow for doubling. In accordance with Šekli's (2010a) findings, no instances of doubling with interrogative and negative pronouns can be reported from *The Little Prince*. Examples (38a–b) show the lack of doubling with these parts of speech.

(38) a. Du stä?
who be.PRS.2PL
'Who are you?' (p. 64)

b. Ma nišći nï bil se
but nobody.NOM.SG.M NEG.AUX.3SG be.PST.PTCP.SG.M REFL
mu vërwal […]
he.OC.DAT believe.PST.PTCP.SG.M
'But nobody did believe him […]' (p. 19)

[16] In the sole instance of doubling with the bare numeral 'one' (24a), the numeral acts as a generalizing pronoun (see §5.2.3).

Thus, unlike, for example, in dialectal Dutch (cf. §2), subject doubling is much less restricted in Resian with regard to the parts of speech that occur with doubling. Among the semantic properties shown to influence the realization of doubling in some languages (cf. §2), none appear to inhibit the phenomenon in Resian. The lack of animacy, definiteness, or specificity of subjects does not prevent the doubling, nor did I find a restriction of the doubling to main or subordinate clauses. The same can be said about different word order patterns. At the least, subject-verb inversion does not inhibit doubling. However, it should be stressed that the present study was not designed to make any claims about the frequency of subject doubling in Resian. In a further step, it should be investigated whether there is a correlation between the frequency of doubling and the above-mentioned semantic and clausal properties.

As for the placement of subject clitics in doubling constructions, I have not found any differences compared to constructions without doubling. Subject clitics always precede the highest-ranking verb in the clause, including auxiliaries. The presence of subject phrases in the clause does not have an influence on this placement. Regarding this, clauses with subject-verb inversion are especially telling. In these clauses, subject phrases follow the entire verbal complex, while subject clitics remain in a position before the highest-ranking verb. Therefore, we can conclude that the placement of subject clitics is independent of the subject phrase (cf. also (2)).

### 5.5. Resian Subject Doubling as a Case of Language Contact

It has already been noted that from a broader Slavic perspective, subject doubling is an exceptional phenomenon. Apart from Resian, it only occurs in several other Slovenian dialects spoken in northern Italy (cf. Skubic 1997: 84). However, it is no accident that doubling is attested exactly here in Slavic. While standard Italian is a pro-drop-language (e.g., Dryer 2013), several Romance varieties of the area exhibit subject clitics, which they can in many cases use together with overt subjects. Therefore, it is reasonable to assume that the Slovenian dialects of the area have developed subject doubling as a result of language contact (cf. Skubic 1997: 84; Benacchio 2002; Šekli 2010a). As for Resian, the most likely source for the doubling is the Rhaeto-Romance language Friulian. Before the introduction of the Italian standard language, it had been the largest language of the area, which is, for instance, also reflected in the numerous lexical loans found in Resian (e.g., Steenwijk 1990). Concerning the phenomenon of Resian subject doubling specifically, Friulian is the most likely source because of the similarity between the two languages. For instance, in the Venetian dialects discussed by Benincà (2014: 43), the scope of doubling is much more limited. Therefore, the purpose of the present section is to compare the observed Resian patterns with the example of Friulian.

When contrasting subject doubling in Resian with the same phenomenon in Friulian, it should first of all be remembered that in the former language, the use of subject clitics is not obligatory. As mentioned in §4, we find many instances of pro-drop in Resian. In Friulian, on the other hand, subject clitics are almost always obligatory (e.g., Marchetti 1977: 224; Benincà and Vanelli 2016: 150). Moreover, subject clitics in Friulian "are well on the way to becoming further reduced to the status of bound affixes on the verb" (Haiman and Benincà 1992: 170 on subject clitics in Rhaeto-Romance and northern Italian dialects). Aside from the placement before the negative marker in the third-person plural reported by Gaglia (2010), they are always adjacent to the verb and, therefore, largely non-selective with regard to their host. In Resian, on the other hand, subject clitics exhibit a greater degree of positional freedom, as shown by the fact that they can be separated from the verb by tonic elements.

Another difference between the placement of subject clitics in Resian and Friulian doubling constructions concerns questions. While in Resian the subject clitics always precede the verb regardless of the clause type, in Friulian they do so only in declarative sentences (cf. examples 39a–b). In questions, on the other hand, subject clitics follow the verb while their associates can precede or follow the verb as is shown in examples (39c–d) (Haiman and Benincà 1992: 170).[17]

(39) a. Jo o feveli.
I.NOM I.SC speak.PRS.3SG

b. O feveli jo.
I.SC speak.PRS.3SG I.NOM
'I speak.' (assertive)

c. Jo feveli-o?
I.NOM speak.PRS.3SG-I.SC
'Do I speak?'

d. Feveli-o jo?
speak.PRS.3SG-I.SC I.NOM
'Maybe I speak?'

However, it should be acknowledged that while the restriction of Resian subject clitics to the preverbal position is generally very well attested in the investigated source, I have come across only examples (40a–b) as potential instances of doubling in binary questions.

[17] Note that in the Gorizia dialect of Friulian, subject clitics are confined to the preverbal position (see Roseano 2015: 179). However, this is not true for the central and Carnic dialects adjacent to the Resian speech area (cf. the map in Frau 1984).

(40) a. Wöwca **na** snëdla ǵö o aliböj
sheep.NOM.SG.F she.SC eat.PST.PTCP.SG.F yes or or
në to rožo?
no DEF.ACC.SG.F flower.ACC.SG.F
'Did the sheep eat the flower, yes or no?' (p. 93)

b. **Na** ni wridna ta wöjska
she.SC NEG noteworthy.NOM.SG.F DEF.NOM.SG.F war.NOM.SG.F
ta-mi wöwce anö rože?
between sheep.ACC.PL and flower.ACC.PL
'Is the war between sheep and flowers not important?' (p. 29)

Analyzing example (40a) as an instance of subject doubling does not provide any problems. In the case of (40b), we find a construction with *ce* 'this, that' in the French original. Therefore, one could ask whether the Resian doubling construction mimics the *ce*-construction in French. However, the argument against interpreting it as a simple calque is the gender agreement between the subject clitic *na* and the subject phrase *ta wöjska*, which is absent in the original, cf. *ce* and *la guerre*. Considering that Resian would have had the subject clitic *to* in its function as an expletive at disposal to translate French *ce*, this deviation speaks for a genuine Resian case of doubling (cf. Schuster-Šewc 1974: 349 for the use of *to* as an expletive).

Thus, while the sequential position of the subject clitics in Resian is fixed—they always precede the finite verb—in Friulian, it is dependent on the type of phrase: proclitic in declarative sentences, but enclitic in interrogative and also in optative clauses. At the same time, the subject phrase itself is flexible with regard to its position in the clause both in Resian and in Friulian: It can occur before or after the verb regardless of the placement of the subject clitic.

As for the subject phrases that occur with doubling, we find that their number is greater in Friulian. Example (41), reproduced from Benincà (2014: 43), displays the use of Friulian subject clitics in clauses with different types of subjects, including a sentence which lacks an overt subject phrase (example (41a)) and a sentence with *al* 'he' acting as an expletive (example (41g)). Sentences marked by an asterisk are ungrammatical due to the absence of subject clitics.[18]

[18] Note that the realization of subject clitics is limited in some instances (Marchetti 1977: 221–26; Madriz and Roseano 2003: 74–78; Šekli 2010a: 144; Gaglia 2010; Benincà and Vanelli 2016: 150–52; Poletto 1991: 56–62). Since this does not concern cases with doubling specifically, they are not discussed here.

(41) a. *Rive / Al rive (*Arrives / He arrives)
b. Toni al rive / *Toni rive (Toni he arrives / *Toni arrives)
c. Nissun al rive / *Nissun rive (Nobody he arrives / *Nobody arrives)
d. Cui rivi-al? / *Cui rive? / Rivi-al? (Who arrives-he? / *Who arrives? / Arrives-he?)
e. Il fantat k-al rive… / *Il fantat ke rive… (The boy that-he arrives… / *The boy that arrives…)
f. Al rive Toni / *Rive Toni (He arrives Toni / *Arrives Toni)
g. Al è masse tart / *È masse tart (It is too late / *Is too late)

As can be seen, there are no restrictions with regard to the elements that can be doubled in Friulian. Notably, subject clitics occur with interrogative and indefinite pronouns, cf. (41c–d), where no doubling is attested in Resian; cf. (38a–b). In the northern Italian context, a similarly extended use of subject clitics can be found in most Piedmontese dialects and in some Ligurian and Lombard varieties (Poletto 2008: 142). Elsewhere, the realization of subject clitics is more restricted. Based on an analysis of these restrictions, Poletto (2000: 140) proposed that subject doubling proceeds according to an implicational scale. According to her, DPs are doubled in a language only if tonic pronouns are doubled, and QPs are doubled only if DPs and tonic pronouns are doubled. Finally, if relative, interrogative, and cleft structures are doubled, doubling is obligatory with all the other types of subjects. The resulting scale is given in (42) (Poletto 2008: 40–41).

(42) Pronouns
Pronouns, DPs
Pronouns, DPs, QPs
Pronouns, DPs, QPs, Variables

Although Poletto (2000) developed this scale based on Romance data, there are no indications that she perceives it as restricted to them. Therefore, it is interesting to note that Resian does not fit into the hierarchy. On one hand, Resian shows doubling of relative markers (18) but not of interrogative pronouns (38a), and on the other hand, it exhibits doubling of universal quantifiers, cf. (22), but not with negative pronouns (38b). Poletto (2008: 40–54) acknowledges that there are differences in the proneness to doubling also between subjects belonging to the same rank on the scale. For instance, certain case forms of pronouns may be more readily doubled than others. Similarly, universal quantifiers appear with doubling more often than existential or negative quantifiers. Therefore, Poletto (2008: 44) suggests that doubling is more frequently realized with elements that

have more functional information. She then reformulates the implicational scale above as a probability scale: "the more an element has features, the more it will be prone to […] double" (Poletto 2008: 44).

Applying Poletto's (2008: 46) adapted scale to Resian does not, however, resolve the above-stated contradiction. The functional feature that differentiates QPs and Variables from each other in Poletto's (2008: 46) scale is number. The Resian negative pronoun *nišći* 'nobody' has the feature number since it occurs only with verbs in the singular.[19] The uninflected relative markers *ki* and *ka*, on the other hand, do not. Following Poletto's argumentation, we would thus rather expect doubling with negative pronouns than with the relative markers. Therefore, Resian contradicts the proposed doubling scale. An explanation of this deviation, however, can be provided if the origin of subject doubling in the Romance languages and in Resian are considered.

Doubling in the Romance languages spoken in northern Italy is assumed to have originated as the result of language-internal developments (Poletto 1991). On the other hand, Resian adopted subject doubling from Romance varieties (most notably from Friulian) due to language contact. Accordingly, there may have been certain structural conditions in Resian that, for some reason, facilitated the adaptation of doubling with certain elements regardless of their place in the doubling scale. One such structure could have been exactly relative clauses. To express case with uninflected relative markers, Slovenian has developed resumptive pronouns identical to the clitic variants of personal pronouns. As can be seen from the standard Slovenian and Resian examples in (43), the uninflected relative marker *ki* is accompanied by the resumptives Slovenian *jim* and Resian *jin*, which both express the dative plural (for the standard Slovenian example, see Herrity 2016: 148).

(43) a. Učenci, ki ste (Standard Slovenian)
pupils.NOM.PL.M REL AUX.2PL
**jim** dali knjige, so
they.OC.DAT give.PST.PTCP.PL book.ACC.PL.F AUX.3PL
naredili izpit.
pass.PST.PTCP.PL exam.ACC.SG.M
'The pupils to whom you gave the books have passed the exam.'

[19] Poletto (2008: 45) argues that the fact that universal quantifiers in northern Italian dialects are doubled by subject clitics in the plural while negative pronouns are doubled by subject clitics in the default singular shows that the former have the feature number while the latter do not.

(43) b. Män trï vulkane, ka (Resian)
have.PRS.1SG three volcanoes.ACC.PL REL
**jin** pücinan ćamyn
they.OC.DAT.PL sweep.PRS.1SG chimney.ACC.SG.M
wsaki tëdan.
every.ACC.SG.M weekACC.SG.M
'I have three volcanoes which I sweep every week.' (p. 49)

Structurally, this pattern is identical to the instances of subject doubling with relative markers (cf. §5.2.3). Therefore, the reason why Resian adopted the doubling of relative markers but not of negative pronouns could be found in the fact that a corresponding structure already existed in the language. Hence, the borrowing consisted in the extension of the pattern to a further cell of the paradigm. If this explanation is correct, it shows that when formulating implicational scales or universals, it is important to consider the origin of the patterns at hand. If it is the result of language contact, other factors, such as the similarity of certain structures in the involved languages, might have played a role in the borrowing process.

In sum, we may conclude that it can hardly be doubted that subject doubling was introduced into Resian as a result of the prolonged language contact with Romance, most notably with Friulian. The fact that we do not find this phenomenon in Slavic languages outside the northern Italian contact zone, as well as the significant similarities between the two languages, clearly speaks for this assumption.

## 6. Conclusions

In this paper, I aimed at a comprehensive description of subject doubling in the Slovenian dialect of Resia with a focus on (1) the types of associates that occur with doubling (cf. §5.2) and (2) the place of the subject clitic relative to its associate and the verb (cf. §5.3). Based on the translation of *The Little Prince*, we can conclude that subject doubling in Resian is far from a sporadic phenomenon. Numerous examples of doubling in a variety of different contexts and positions have been found in the text.

The assumption that Resian has developed subject doubling as a result of language contact with Romance can hardly be contested. However, we also find interesting differences between the Resian and Romance languages spoken in the area. Notably, Resian does not fit into the doubling scale proposed by Poletto (2000, 2008). It has been argued that this can be explained by considering language-internal factors in Resian, which facilitated the adoption of doubling with certain elements. This highlights the importance of considering the origins of phenomena under scrutiny when establishing or contesting implicational scales.

Another conclusion that can be drawn from both Resian and, to a lesser extent, also from Friulian concerns the relationship between clitic doubling and the complexity of inflectional systems. It has often been assumed that clitic doubling is linked to the loss of nominal declension. The idea behind this is that clitic doubling develops as a consequence of the necessity to distinguish subject and object. However, as pointed out by Guentchéva (2008: 207–08 with references) with regard to object doubling in Bulgarian, syntactic relations can remain entirely clear without clitic doubling. According to her, the fact that indirect objects which are marked by the preposition *na* can be doubled in Bulgarian suggests that clitic doubling is not related to the grammatical function of the constituents. With its complex six-case declensional system, Resian seems to confirm the assumption that the rise of clitic doubling is not necessarily linked to the loss of declension (Šekli 2010a: 152). Based on Šekli's (2010a) suggestion that the realization of subject doubling in Resian is related to information structure, it could, therefore, be proposed that factors from this domain can play a role in the introduction as well. However, it must be acknowledged that, at present, we do not know the exact circumstances that induce the doubling in Resian. In this paper, it could be shown that the lack of animacy, specificity, and definiteness in subjects is not an inhibiting factor. However, we do not know whether doubling is more likely to occur when one of these features is present. Therefore, the next step in the research on Resian subject doubling should be to test whether semantic and pragmatic factors indeed play a role in the realization of the doubling. The most promising way to do so seems to be by conducting a large-scale quantitative study, which, however, at the current stage, is not possible because there exist no annotated electronic corpora of Resian.

## Abbreviations

| | | | |
|---|---|---|---|
| ADV | adverb | M | masculine |
| ACC | accusative | N | neuter |
| ATT | attenuative | NEG | negation marker |
| COMP | complementizer | NOM | nominative |
| DAT | dative | OC | object clitic |
| DEF | definite | PASS | passive |
| DU | dual | PL | plural |
| EMPH | emphasizer | PRS | present |
| F | feminine | PST | past |
| GEN | genitive | PTCL | particle |
| IMPERF | imperfect | PTCP | participle |
| INDEF | indefinite | REFL | reflexive |
| INF | infinitive | REL | relative |
| INS | instrumental | SC | subject clitic |
| LOC | locative | SG | singular |

## References

Anagnostopoulou, Elena. (2017) "Clitic doubling". Martin Everaert and Henk C. van Riemsdijk, eds. *The Wiley Blackwell companion to syntax*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Ltd., 1–56.

Aoun, Joseph. (1999) "Clitic doubled arguments". Kyle Johnson and Ian Roberts, eds. *Beyond principles and parameters: Essays in memory of Osvaldo Jaeggli*. Dordrecht: Springer, 13–42.

Benacchio, Rosanna. (2002) *Contatti slavo-romanzi nei dialetti sloveni del Friuli*. Udine, Italy: Societè Filologica Friulana, 63–103.

Benincà, Paola. (2014) "Friulian linguistics". Rose Mucignat, ed. *The Friulian language: Identity, migration, culture*. Newcastle, UK: Cambridge Scholars Publishing, 30–52.

Benincà, Paola, and Laura Vanelli. (2016) "Friulian". Adam Ledgeway and Martin Maiden, eds. *The Oxford guide to the romance languages*. 1st ed. Oxford: Oxford University Press, 139–53. [Oxford Guides to the World's Languages.]

Brandi, Luciana, and Patrizia Cordin. (1989) "Two Italian dialects and the null subject parameter". Osvaldo Jaeggli and Kenneth J. Safir, eds. *The null subject parameter*. Dordrecht/Boston/London: Kluwer Academic Publisher, 111–42. [Studies in Natural Language and Linguistic Theory 15.]

Breu, Walter. (2022) "Romance in contact with Slavic in southern and south-eastern Europe". *Oxford research encyclopedias of linguistics*. Oxford: Oxford University Press.

Burzio, Luigi. (1986) *Italian syntax: A government-binding approach*. Dordrecht/Boston/Lancaster/Tokyo: D. Reidel.

Cardinaletti, Anna. (1999) "Italian emphatic pronouns are postverbal subjects". *University of Venice working papers in linguistics* 9(1–2): 59–92.

Cardinaletti, Anna, and Michal Starke. (1999) "The typology of structural deficiency: A case study of the three classes of pronouns". Henk van Riemsdijk, ed. *Clitics in the languages of Europe, Part 1*. Vol. 5 of *Eurotyp.* Berlin/New York: De Gruyter Mouton, 145–234.

Coveney, Aidan. (2005) "Subject doubling in spoken French: A sociolinguistic approach". *The French review* 79(1): 96–111.

Čermelj, Lavo. (1938) *La minorité slave en Italie (les slovènes et croates de la marche Julienne).* Ljubljana: Delniška tiskarna d. d.

Dapit, Roberto. (2003) "Nastajanje krajevnih knjižnih jezikov pri slovencih v furlaniji". Ada Vidovič-Muha, ed. *Slovenski knjižni jezik—aktualna vprašanja in zgodovinske izkušnje: ob 450-letnici izida prve slovenske knjige*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani, 301–12.

de Saint-Exupéry, Antoine. (2021) *Te mali prïncip. Ziz dizinje od itogä, ki ga napïsal. Vobračën po rozajanskin od Silvane Paletti anö Malinke Pila*. Neckarsteinach, Germany: Tintenfass.

de Vogelaer, Gunther, and Annemie Neuckermans. (2002) "Subject doubling in Dutch: A dialect phenomenon in cross-linguistic perspective". *Sprachtypologie und Universalienforschung* 55(3): 234–58.

Dryer, Mathew S. (2013) "Expression of pronominal subjects (v2020.3)". Matthew S. Dryer and Martin Haspelmath, eds. *The world atlas of language structures online*. Zenodo.

Escher, Anastasia. (2021) "Double argument marking in Timok dialect texts (in Balkan Slavic context)". *Zeitschrift für Slawistik* 66(1): 61–90.

Fonseca-Greber, Bonnibeth Beale. (2000) *The change from pronoun to clitic and the rise of null subjects in spoken Swiss French*. PhD thesis, University of Arizona.

Franks, Steven, and Tracy H. King. (2000) *A handbook of Slavic clitics*. New York: Oxford University Press. [Oxford Studies in Comparative Syntax.]

Frau, Giovanni. (1984) *I dialetti del friuli*. Udine, Italy: Societa' filologica friulana.

Gaglia, Sascha. (2010) "The omission of preverbal subject clitics in Friulian: Methodology and constraint-based analysis". *Corpus* 9: 191–220.

Guentchéva, Zlatka. (2008) "Object clitic doubling constructions and topicality in Bulgarian". Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 203–23. [Linguistik Aktuell/Linguistics Today, 130.]

Haiman, John, and Paola Benincà. (1992) *The Rhaeto-Romance languages*. London/New York: Routledge.

Herrity, Peter. (2016) *Slovene: A comprehensive grammar*. 2nd ed. London: Routledge. [Routledge Comprehensive Grammars.]

Holmberg, Anders, and Urpo Nikanne. (2008) "Subject doubling in Finnish: The role of deficient pronouns". Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margaret van der Ham, eds. *Microvariation in syntactic doubling*. Leiden, Netherlands: Brill, 147–89.

Jlassi, Mohamed. (2013) *The multiple subject construction in Arabic: Evidence from subject doubling in Tunisian Arabic*. PhD thesis, School of English Literature, Language and Linguistics, Newcastle University.

Kallulli, Dalina, and Liliane Tasmowski, eds. (2008) *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins. [Linguistik Aktuell/Linguistics Today, 130.]

Koneski, Blaže. (1976) *Gramatika na makedonskiot literaturen jazik*. Vols. 1 and 2. Skopje, North Macedonia: Kultura.

Krapova, Iliyana, and Guglielmo Cinque. (2008) "Clitic reduplication constructions in Bulgarian". Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 257–86. [Linguistik Aktuell/Linguistics Today, 130.]

Lambrecht, Knud. (1994) *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge, UK: Cambridge University Press.

Madriz, Anna, and Paolo Roseano. (2003) *Scrivere in friulano*. Udine, Italy: Societât filologjiche furlane.

Malakov, Milen. (2017) "Rezijanskijat knižoven ezik". *Balkanistic Forum* 3: 95–103.

Marchetti, Giuseppe. (1977) *Lineamenti di grammatica friulana*. Udine, Italy: Societa' filologica friulana.

Moseley, Christopher, ed. (2010) *Atlas of the world's languages in danger*. Paris: Unesco. [Memory of Peoples Series, 3rd ed.]

Nadasdi, Terry. (1995) "Subject NP doubling, matching, and minority French". *Language variation and change* 7(1): 1–14.

Poletto, Cecilia. (1991) "The diachronic development of subject clitics in North Eastern Italian dialects". *Working papers in linguistics*. Vol. 1. Venice: Centro Linguistico Interfacoltà, Università degli studi di Venezia.

————. (2000) *The higher functional field: Evidence from northern Italian dialects*. New York/Oxford: Oxford University Press

————. (2008) "Doubling as splitting". Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margaret van der Ham, eds. *Microvariation in syntactic doubling*. Bingley, West Yorkshire, UK: Emerald, 37–68. [Syntax and Semantics, 36.]

Ramovš, Franc. (1928) "Karakteristika slovenskega narečja v Reziji". *Časopis za slovenski jezik, književnost in zgodovino* 7(1–4): 107–21.

Ramovš, Franc. (1935) *Historična gramatika slovenskega jezika.* Vol. 7, *Dialekti*. Ljubljana, Slovenia: Učiteljska tiskarna.

Rizzi, Luigi. (1986) "On the status of subject clitics in Romance". Osvaldo Jaeggli and Carmen Silva-Corvalán, eds. *Studies in romance linguistics*. Dordrecht: Foris, 391–419.

Roseano, Paolo. (2015) "Suddivisione dialettale del friulano". Sabine Heinemann and Luca Melchior, eds. *Manuale di linguistica friulana*. Berlin/Boston: de Gruyter, 155–86.

Rosenkvist, Henrik. (2015) "The syntax and meaning of subject doubling in Övdalian". Kristine Bentzen, Henrik Rosenkvist, and Janne Bondi Johannessen, eds. *Studies in Övdalian morphology and syntax: New research on a lesser-known Scandinavian language*. Amsterdam/Philadelphia: John Benjamins, 107–35. [Linguistik Aktuell Linguistics Today, 221.]

Runić, Marija. (2014) *Issues in the syntax of nominals*. PhD thesis, Dipartimento di Studi Linguistici e Letterari, Università degli Studi di Padova, Padua, Italy.

————. (2018) "Slovensko-romanski jezički kontakt na primeru dve pojave iz rezijskog". *Slavia meridionalis* 18: 1–16.

Schuster-Šewc, Heinz. (1974) "Sätze mit fiktivem Subjekt vom Typ os. *wono so dešćuje* / ns. *to se pada* 'es regnet' und ihre Stellung in der slawischen Syntax". *Zeitschrift für Slawistik* 19(3): 340–52.

Skubic, Mitja. (1997) *Romanske jezikovne prvine na zahodni slovenski jezikovni meji*. Ljubljana, Slovenia: Znanstveni inštitut Filozofske fakultete.

Steenwijk, Han. (1990) "The nominal declension of Friulian loans in the Slovene dialect of Val Resia". *Slovene studies* 12(1): 23–31.

————. (1992) *The Slovene dialect of Resia: San Giorgio.* Amsterdam: Rodopi. [Studies in Slavic and General Linguistics, 18.]

————. (1994) *Ortografia resiana.* Padua, Italy: CLEUP.

————. (1996) "Der romanisch-slavische Sprachkontakt und die interne Differenzierung des Resianischen". Jože Toporišič, ed. *Kopitarjev sbornik: Mednarodni simpozij v Ljubljani, 29. junij do 1. julij 1994*. Ljubljana: Filozofska fakulteta, Oddelek za slovanske jezike in književnosti, Seminar slovenskega jezika, literature in kulture, Znanstveni inštitut, 553–66.

————. (2005) *Piccolo dizionario ortografico resiano.* Padua, Italy: CLEUP.

————, ed. (1999a) *Fondamenti per una grammatica pratica resiana*. Padua, Italy: CLEUP.

————, ed. (1999b), *Grammatica pratica resiana: il sostantivo*. Padua, Italy: CLEUP.

Stolz, Thomas. (2007) "*Harry Potter* meets *Le petite prince*—on the usefulness of parallel corpora in crosslinguistic investigation". *Sprachtypologie und Universalienforschung* 60(2): 100–17.

Šekli, Matej. (2010a) "Zaimkovno podvajanje predmeta in osebka v rezijanskem narečju slovenščine : (s stališča jezikovnega stika s furlanščino)". *Treta makedonsko-slovenečka naučna konferencija: makedonsko-slovenečki jazični, književnini i kulturni vrski = makedonsko-slovenske jezikoslovne, književne in kulturne zveze (Ohrid, 12–15 septemvri 2007).* Skopje: Filološki fakultet "Blaže Koneski". Univerzitet "Sv. Kiril i Metodij", 133–55.

————. (2010b) "Furlanščina in rezijansko narečje slovenščine v besedilih". *Treta makedonsko-slovenečka naučna konferencija: makedonsko-slovenečki jazični, književnini i kulturni vrski = makedonsko-slovenske jezikoslovne, književne in kulturne zveze (Ohrid, 12–15 septemvri 2007).* Skopje: Filološki fakultet "Blaže Koneski". Univerzitet "Sv. Kiril i Metodij", 157–79.

————. (2015) "Rezijanščina: geneolingvistična in sociolingvistična opredelitev". *Poznańskie studia slawistyczne* 8: 199–213.

————. (2018) "Slovenski jezik v Reziji". Andreja Žele and Matej Šekli, eds. *Slovenistika in slavistika v zamejstvu—Videm*. Ljubljana, Slovenia: Zveza društev Slavistično društvo Slovenije, 82–94.

Testelets, Yakov. (2003) "Are there strong and weak pronouns in Russian?" Wayles Browne, Ji-Yung Kim, Barbara H. Partee, and Robert A. Rothstein, eds. *Annual Workshop on Formal Approaches to Slavic Languages: The Amherst Meeting 2002*. Ann Arbor, MI: Michigan Slavic Publications, 515–38.

Usikova, Rina Pavlovna. (2003) *Grammatika makedonskogo literaturnogo jazyka*. Moscow: Muravej.

Vogelaer, Gunther D., and Magda Devos. (2008) "On geographical adequacy, or: How many types of subject doubling in Dutch". Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margaret van der Ham, eds. *Microvariation in syntactic doubling*. Leiden, Netherlands: Brill, 251–76. [Syntax and Semantics 36.]

Wandl, Florian. (forthcoming) "Pronominal clitics in the Slovenian dialect of Resia".

Werkmann, Valja. (2015) *Objektklitika im Bulgarischen*. Berlin: Akademie Verlag. [Studia Grammatica, 57; reprint 2015 ed.]

Westbury, Josh. (2016) "Left dislocation: A typological overview". *Stellenbosch papers in linguistics plus* 50: 21–45.

Zwicky, Arnold M. (1994) "What is a clitic?". Joel A. Nevis, Brian D. Joseph, Dieter Wanner, and Arnold M. Zwicky, eds. *Clitics: A comprehensive bibliography (1892–1991).* Amsterdam: John Benjamins, xii–xx. [Library and Information Sources in Linguistics 22.]

Florian Wandl
University of Zurich
Department of Slavonic Languages and Literatures
Zurich, Switzerland
florian.wandl@uzh.ch

# Quasi-Participial *mówiąc* 'speaking' with Adverbial Complements in Polish: A Quantitative Corpus-Based Study

**Jarosław Wiliński**

*Abstract*: This article employs frame semantics, a usage-based model of construction grammar, along with a quantitative corpus-based approach, to examine the characteristics of the construction involving adverbial complements and the quasi-participle *mówiąc* 'speaking' in Polish. The author analyzes instances of this construction within the National Corpus of Polish to ascertain its structural, semantic, distributional, and discourse-functional features. Furthermore, the study identifies adverbial complements with a strong affinity for the construction. The investigation reveals that the construction is frequently associated with distinct categories of adverbial complements that invoke diverse semantic frames. Additionally, this construction appears in various registers and serves multiple functions in discourse.

*Keywords:* adverbial complements, construction, usage-based construction grammar, frame semantics, the National Corpus of Polish, quantitative corpus-based study

## 1. Introduction and Literature Review

The phenomenon of Polish metatextual expressions, particularly quasi-participles preceded by adverbial phrases (e.g., *ściśle mówiąc* 'strictly speaking' or *ogólnie mówiąc* 'generally speaking'), has gained renewed attention due to the availability of computerized corpora and increased interest in grammatical and functional vocabulary (Stępień 2014; Birzer 2015, 2017; Kubicka 2017; Żabowska 2020). Previous studies have examined their origins (Birzer 2017), lexical status (Bednarek 1989; Weiss 2005; Stępień 2014, 2015), the distinction between *[Adv] speaking* and *speaking [Adv]* (Żabowska 2020; Wiliński 2025), their functions and prosodic features (Stępień 2014), and the formal and semantic properties of adverbial expressions (Kubicka 2017; Moroz 2007), as well as the conditions for the independent usage of adverbial phrases (Birzner 2012; Kleszczowa 2015; Kubicka 2017). Additionally, some researchers have briefly discussed these expressions, classifying them as parentheticals (Moroz 2010) or particles (Grochowski 2002, 2007; Kisiel and Żabowska 2011). Wróbel (1975), Weiss (2005), and Żabowska (2009) have offered broader analyses, identifying certain expressions as phraseological units or metatextual commentaries.

Dictionary resources have also provided relevant information about these adverbial complement (AC)-*speaking* constructions. For example, Bogusławski and Wawrzyńczyk (1993) identified *delikatnie mówiąc* 'to put it mildly' as a form of mitigating metatextual commentary. The SWJP, ISJP, USJP, and WSJP PAN Polish dictionaries provided lexical classifications, categorizing some expressions as particles, phraseological units, or metatextual operators that aid in text creation (SWJP 1996; ISJP 2000; USJP 2004; WSJP PAN 2006–24).

While traditional research has significantly contributed to our understanding of quasi-participles, it has primarily focused on formal and semantic classifications, frequently neglecting usage dynamics, cognitive processing, and the interconnectedness of constructions. Consequently, adopting a usage-based construction grammar approach could reveal new insights into the status of the AC-*mówiąc* construction. Such a perspective emphasizes actual usage contexts, cognitive representations, relationships with other constructions, frequency-driven entrenchment, and corpus-based evidence, thereby effectively capturing subtle nuances of usage and providing a more comprehensive understanding.

Nevertheless, to date, there has been a noticeable absence of studies analyzing the AC-*mówiąc* construction from a usage-based perspective, particularly through quantitative corpus methodologies. Existing research, such as Kubicka's (2017) semantic classification of ACs, has only provided a basic division of adverbial phrases into evaluative communication markers (e.g., *szczerze* 'sincerely/honestly', *uczciwie* 'honestly', or *otwarcie* 'openly') and communication-altering markers (e.g., *krótko* 'briefly', *ogólnie* 'generally', or *inaczej* 'differently'). While such classifications are informative, they do not fully capture the complexity of constructional usage that usage-based analysis reveals. Therefore, this article aims to (1) confirm the existence of the AC-*mówiąc* construction as a symbolic unit within a constructional cline; (2) provide a comprehensive description of its structural, semantic, pragmatic, distributional, and discourse-functional characteristics; and (3) quantitatively identify ACs that are strongly or weakly associated with this construction.

The rest of the article is structured as follows: Section 2 details the theoretical and methodological frameworks, along with the data sources and analytical procedures. Section 3 analyzes the structural, semantic, distributional, and discourse-functional characteristics of the AC-*mówiąc* construction. Section 4 presents and discusses the quantitative results related to ACs. Finally, Section 5 concludes by summarizing the main findings of the study.

## 2. Theoretical and Methodological Frameworks

This section outlines the theoretical and methodological frameworks used in this study.

## 2.1. Theory

This subsection addresses the theoretical issues related to construction grammar and frame semantics.

### 2.1.1. Construction Grammar

This article applies a usage-based model of construction grammar (see Barlow and Kemmer 2000; Bybee and Beckner 2010; Diessel 2017; Perek 2023) and utilizes the theoretical principles of frame semantics (Fillmore 2006) to identify the formal and semantic characteristics of the construction under study. Construction Grammar (Goldberg 2006, 2013) rests on three main assumptions: grammar comprises pairings of form and meaning (or function), known as constructions; constructions can encompass various levels of complexity and generality; and constructions are interconnected within an extensive network (Diessel 2019). All grammatical units can be defined as constructions as long as they possess non-compositional meanings, non-predictable formal properties, or a high enough frequency to be memorized as such (Goldberg 2006: 5). From a usage-based perspective, the representation of constructions in the minds of individual speakers is believed to emerge from their frequent real-life language usage. Linguistic signs that occur with sufficient frequency become entrenched in a speaker's or hearer's linguistic system, thus acquiring the status of constructions (Croft and Cruse 2004: 292–93).

### 2.1.2. Frame Semantics

Frame semantics (Fillmore 2006; Fillmore and Baker 2010) is a theory that explores how language relates to conceptual knowledge. It posits that words and phrases derive meaning by activating mental structures called *semantic frames*. These frames represent our understanding of specific situations, events, or domains of knowledge. For example, consider the verb *speak*, which invokes the meaning associated with the STATEMENT frame. In this context, a speaker conveys a particular message on a specific topic to an addressee. The frame includes both core and non-core frame elements (FEs), representing participant roles signaled by particular lexical items that activate this frame. Core FEs, such as *speaker*, *message*, *medium*, and *topic*, are essential in defining the frame's meaning. Conversely, non-core FEs, including *addressee*, *time*, *place*, or *manner*, serve a more peripheral and descriptive function within the frame.

This study utilizes frame semantics to clarify the meanings of the examined construction and the adverbial complements linked to the participle *mówiąc*. The semantic frames of STATEMENT and CANDIDNESS, along with their updated descriptions, are derived from the FrameNet lexical database

(Fillmore, Johnson, and Petruck 2003; hereafter referred to as the "FrameNet project" in data references). These descriptions offer paraphrases that include core-frame elements, typical participants, and roles specific to a particular context.

The author has developed additional frames, including SUMMARIZATION, ASIDE, CONFIDENTIALITY, REPHRASING, UNDERSTATEMENT, ACCURACY, GENERALITY, SIMPLIFICATION, NEGATIVE EVALUATION, POSITIVE EVALUATION, BLUNTNESS, FIGURATIVITY, LITERAL DESCRIPTION, CONVENTIONALITY, COLLOQUIAL STATEMENT, SERIOUSNESS, RELATABILITY, SELF-PROMOTION, PRACTICALITY, ELEVATION, CONTRADICTION, OBJECTIVITY, JEST, BOASTING, and IRONY. These frames aim to define the meanings associated with the ACs and the participle *mówiąc* more precisely. They highlight the connection between the STATEMENT frame and the meaning conveyed by each AC. Consequently, while they inherit the general properties of the STATEMENT frame, their semantic focus is limited to specific styles or manners in which the speaker communicates the message.

A contextualized semantic test was employed to identify the semantic frames associated with ACs in the construction. This test involved analyzing the whole discourse surrounding each instance of the construction to determine whether a particular AC evoked a specific frame, based on its interaction with contextual cues and background knowledge. The identified frames were grouped into two categories: *content-related* and *wording/phrasing-related*. Content-related frames capture the underlying message, intent, or evaluative stance of the utterance, encompassing categories such as SUMMARIZATION, CANDIDNESS, ASIDE, UNDERSTATEMENT, ACCURACY, GENERALITY, and both NEGATIVE and POSITIVE EVALUATION. In contrast, wording/phrasing-related frames focus on the style or structure of the utterance and include categories such as REPHRASING, SIMPLIFICATION, FIGURATIVITY, LITERAL DESCRIPTION, CONVENTIONALITY, and COLLOQUIAL STATEMENT.

However, the distinction between these two frame types is not always clear-cut. Due to the fluid and context-sensitive nature of language, many expressions convey both content and stylistic functions simultaneously, resulting in overlaps between categories. For example, *krótko mówiąc* 'briefly speaking' can both summarize content and rephrase it concisely; *szczerze mówiąc* 'sincerely/honestly speaking' expresses sincerity while also presenting information in a direct, unembellished manner. Similarly, the ASIDE frame often overlaps with COLLOQUIAL STATEMENTS or REPHRASING, complicating the interpretation of its primary function. These overlaps reflect the inherent ambiguity and multifunctionality of metatextual expressions, which are influenced by discourse context, speaker intent, and stylistic conventions. Therefore, the semantic frame analysis of these constructions necessitates a flexible, context-aware approach that accommodates the dynamic interplay between meaning and form in actual language use.

## 2.2. Methodology

This subsection discusses the corpus-based methodology, the origins of the data, the processes of data extraction and quantification, and the tools and procedures used for retrieval and statistical evaluation.

### 2.2.1. Method

Regarding the methodology, this study employs the attraction and reliance measures introduced by Schmid (2000: 54–55) and further refined by Schmid and Küchenhoff (2013). These measures address the issue of directionality by distinguishing how much a specific lexeme attracts a grammatical construction and how much a lexeme depends on a particular construction. The former proportion metaphorically represents the attraction that the construction exerts on the lexeme, while the latter reflects the lexeme's dependence on the construction. Attraction is calculated by dividing the frequency of a lexeme's occurrence in a construction by the frequency of that construction in the corpus. Conversely, reliance is determined by dividing the frequency of a lexeme's occurrence in a construction by its overall frequency in the entire corpus. Both values are then multiplied by 100 to express the scores as percentages. The calculation is shown in Table 1 on p. 351, which includes examples illustrating the scores for the AC *szczerze* 'sincerely/honestly'. The resulting percentage values obtained from these statistical measures serve as indicators of attraction and reliance. A higher percentage indicates a stronger level of attraction and reliance, while a lower percentage signifies a weaker level of attraction and reliance.

### 2.2.2. Corpus Data

The data source for this study is a balanced version of the National Corpus of Polish (NKJP). This version comprises an approximately 300 million-word text database containing samples of written and spoken Polish from diverse sources, including classic literature, daily newspapers, specialist periodicals and journals, conversation transcripts, and various transient and internet texts. The PELCRA corpus search engine (Pęzik 2012) was utilized to search the balanced version of the corpus, which encompasses over 250 million textual words, equivalent to approximately 300 million segments.

### 2.2.3. Data Retrieval and Statistical Assessment

The data retrieval process involved multiple steps. Initially, the frequencies of the participle *mówiąc* were automatically obtained from the NKJP corpus

using the PELCRA search engine. This software system was employed to search all texts within the corpus for instances of the participle within a five-word context and a sample size of 10,000 words. Through this corpus search, concordance lines were directly retrieved and then manually inspected to identify valid combinations of the participle *mówiąc* with adverbs and adverbial phrases. These valid combinations refer to instances that do not indicate the simultaneous occurrence of events with the same subject in two clauses; instead, they represent characteristics of current utterances related to the loss of categorical features of the participle.

Any false combinations indicating co-reference between the subject of the participle and the first argument of the matrix verb were excluded from further analysis. For example, the expression *nieśmiało mówiąc*, meaning 'shyly speaking', in the sentence *Daniel pocieszał ją, nieśmiało mówiąc o szacunku dla każdego, nawet upośledzonego życia* 'Daniel comforted her, shyly speaking about respect for everyone, even for a disabled life', was considered a false hit because it represented a standard use of the active verb participle *mówiąc*.

The observed frequencies, with "a" representing the frequency of the AC *szczerze* 'sincerely/honestly' in the AC-*mówiąc* construction, and "x" representing the total frequency of all ACs in the construction, as shown in Table 1, were manually calculated by examining concordance lines. In contrast, the total frequencies of ACs in NKJP, denoted as "y" for the total frequency of the AC *szczerze*, were automatically computed by the software program.

Subsequently, the observed frequencies (a, x, and y) in Table 1 were entered into an Excel worksheet and used to calculate Schmid's measures of attraction and reliance, following the methodology described by Wiliński (2021). Attraction was determined by dividing the raw frequency of each AC in the AC-*mówiąc* construction by the total frequency of all ACs in this construction, while reliance was calculated by dividing the frequency of an AC's occurrence in the construction under study by its frequency of occurrence in NKJP, in accordance with Schmid's (2000: 54) approach. The scores for attraction and reliance were then expressed as percentages by multiplying the raw frequency of a specific AC in the AC-*mówiąc* construction by one hundred. These percentage results were treated as indices of attraction or reliance, with a higher percentage indicating a stronger attraction to and reliance on the AC-*mówiąc* construction (cf. Wiliński 2021).

A quick look at Table 1 immediately reveals that the word *szczerze* 'sincerely/honestly' appears 1,786 times in the AC-*mówiąc* pattern. Consequently, its attraction score is also high, at 16.29%, indicating that this adverb accounts for 16.29% of all AC uses in the specific construction being discussed. In simpler terms, *szczerze* is strongly associated with the AC-*mówiąc* construction. However, as indicated in the last column, its reliance score is higher than its attraction score, standing at 19.22%. This percentage suggests that only 19.22% of occurrences of *szczerze* can be observed in the AC-*mówiąc* construction com-

**Table 1.** Frequency data for a statistical analysis

| AC | a | x | y | attraction | reliance |
|---|---|---|---|---|---|
| szczerze | 1786 | 10965 | 9291 | 16.29 % | 19.22 % |

( a = The frequency of the AC *szczerze* 'sincerely' in the AC-*mówiąc* construction;
x = The total frequency of all ACs in the construction;
y = The total frequency of the AC *szczerz*e in NKJP)

pared to other contexts within the corpus. Consequently, it can be inferred that *szczerze* depends on other patterns for most of its usage, accounting for 80.78%. In other words, it appears more frequently in contexts other than the AC-*mówiąc* construction. The calculation results were sorted based on the attraction score and subjected to qualitative analysis.

## 3. The AC-*mówiąc* Construction

Extensive empirical evidence derived from corpus analysis, including over 10,000 attested instances of *mówiąc* 'speaking' combined with various ACs, supports the recognition of the AC-*mówiąc* construction as a distinct and cognitively entrenched unit in contemporary Polish. Frequent and recurrent forms such as *krótko mówiąc* 'briefly speaking' (1,826 occurrences) provide quantitative grounding for the construction's status. The rationale for positing the AC-*mówiąc* construction as a symbolic unit rests on three interrelated factors: semantic and functional specialization, structural idiosyncrasy, and partial schematicity.

First, the construction's meaning and function are not strictly predictable from its component parts. It possesses specific functional characteristics associated with a metatextual expression (Ożóg 1990; Birzer 2017), parenthetical expression (Moroz 2007; Stępień 2014), or discourse structuring element (Birzer 2012, 2015). Thus, it fulfills a discourse-managing, rather than purely descriptive, role. This behavior can be explained through the principle of coercion (Michaelis 2004: 25), whereby semantically incongruent lexical items are adapted to fit the morphosyntactic context. In the AC-*mówiąc* construction, ACs such as *nawiasem* ''parenthetically/incidentally', *krótko* 'briefly', and *szczerze* 'sincerely/honestly' are often coerced into functioning not merely as modifiers but also as discourse markers, shifting their role from conveying adverbial information to structuring discourse. These expressions typically signal transitions, introduce supplementary remarks, or frame the speaker's stance, resulting in a semantic enrichment that exceeds the sum of their parts. However, the strength of coercion is not uniform across all ACs. While some

ACs (*krótko* 'briefly', *szczerze* 'sincerely/honestly', *delikatnie* 'delicately/mildly', and *inaczej* 'differently') combine freely with different forms of *mówić* 'speak', others exhibit clear restrictions (e.g., *mówiąc nawiasem* 'speaking incidentally/parenthetically' seems ungrammatical), supporting a clinal (graded) model of coercibility. This variation lends empirical weight to Żabowska's (2020) distinction between *unitées concrètes*—conventionalized, idiomatic expressions—and the results of syntactic operations.

Complementing coercion, the principle of preemption (Boyd and Goldberg 2011) explains why certain expressions become dominant. Highly frequent and idiomatic combinations, such as *nawiasem mówiąc* 'parenthetically speaking' and *szczerze mówiąc* 'sincerely speaking', "preempt" alternative constructions, becoming the preferred means of conveying specific discourse functions due to their clarity, rhetorical efficiency, and conventional status. Through repeated usage, these expressions become entrenched in speakers' mental grammars, contributing to the formation of a partially schematic construction.

Secondly, the AC-*mówiąc* construction exhibits structural irregularity. It is quasi-participial (Stępień 2015) and displays formal and syntactic constraints that distinguish it from related participial expressions, such as the reverse-order *mówiąc*-AC construction (e.g., *mówiąc żargonem piłkarskim* 'speaking in football jargon', *mówiąc słowami Pascala* 'speaking in Pascal's words', or *mówiąc slangiem* 'speaking in slang'). Unlike standard participles, the AC-*mówiąc* construction typically resists transformation into subordinate clauses, is prosodically marked off, and refers to the speech act itself rather than to the events described in the discourse (cf. Weiss 2005: 182–88).

Thirdly, the construction is partially schematic, consisting of a fixed element—the participle *mówiąc*—and a variable slot accommodating a range of adverbial complements. These complements include adverbs (e.g., *szczerze* 'sincerely/honestly'), nouns in the instrumental or accusative case (e.g., *nawiasem* 'parenthetically/incidentally', *prawdę* 'truth'), and prepositional phrases (e.g., *między nami* 'between us', *w skrócie* 'in short'), selected according to specific semantic and syntactic criteria. The construction's usage in discourse is illustrated by the following examples from NKJP, with the participial unit italicized:

(1) *Nawiasem mówiąc*, Kaligula to przydomek Gajusza Cezara.
'*By the way*, Caligula is a nickname of Gaius Caesar.'
(*Gazeta Wyborcza*)[1]

(2) Ja nie pamiętam, *szczerze mówiąc*.
'I don't remember, *to be honest*.' (*O jurze i górach*)

[1] Details regarding the sources of the examples are given in the Appendix (p. 377).

(3) Ja, *prawdę mówiąc*, ociągałem się z jej potępieniem.
'*In truth*, I procrastinated her condemnation.' (*Oni*)

These examples reveal a consistent bipartite structure: one component consists of the participial construction, which can be positioned at the beginning, inserted medially, or placed at the end of the clause; the second component is the main finite clause. Notably, the participial construction requires an overt adverbial complement: the omission of the AC results in an ungrammatical or highly unnatural structure (**Ja nie pamiętam, mówiąc*, with the intended meaning 'I do not remember, speaking'). Thus, the selection of appropriate ACs within this construction is semantically and pragmatically constrained.

In sum, empirical evidence, frequency, semantic specialization through coercion and preemption, structural idiosyncrasy, and partial schematicity converge to support the recognition of the AC-*mówiąc* construction as a cognitively entrenched, functionally specialized construction in modern Polish.

### 3.1. Structural, Semantic, and Functional Characteristics

The prototypical syntactic structure of this construction is symbolized as [AC *mówiąc*, FINITE CLAUSE], as portrayed in Table 2 on p. 354. The adverbial phrase and the participle introduce an adverbial participle clause, which serves as a disjunct (Quirk et al. 1985)—also referred to as a stance adverbial (Biber et al. 1999) or modal adjunct (Halliday 2004)—encompassing the speaker's intention, interpersonal meaning, and perspective regarding the message. Hence, the principal communicative function of this adverbial participle construction is to indicate the speaker's stance or attitude regarding the message and to comment on the style or form of the utterance.[2] This frequently provides insight into how the speaker conveys the message or how the utterance should be understood.

This construction may be considered optional within discourse segments, as it operates beyond the structural confines of a complete sentence, as illustrated in examples (1–3). Nonetheless, this participial construction significantly influences the entire sentence in most cases, such as the aforementioned examples. Essentially, the adverbial clause introduces an interpersonal comment on the content or style of the comprehensive proposition. A comma indicates a pause between sentence components and separates the adverbial participle clause from the finite clause. The finite clause commonly employs a topic-comment structure, signifying states or occurrences. It begins with a defined subject, presenting familiar information, followed by a comment (predicate) introducing new information.

[2] For similar functions of disjuncts, see Quirk et al. (1985) and Biber et al. (1999).

Semantically, diverse instances (1–3) evoke the STATEMENT frame, where a speaker imparts a distinct message about a specific topic to an implied addressee in a particular manner or style. The participle construction denotes the manner of speaking, while the finite clause conveys the message regarding the specific topic. The finite clause's subject (topic) and predicate (comment) communicate a precise message. The addressee is implicitly included in the entire sentence. These distinctive features of the construction are succinctly summarized in Table 2.

**Table 2.** Features of the prototypical AC-*mówiąc* construction

| | | | |
|---|---|---|---|
| **Form** | an adverbial participle clause: AC + the present participle *speaking* | (,) | finite clause (subject + predicate) |
| **Meaning** | the STATEMENT frame | a comma | states/activities/ achievements/ accomplishments/ |
| **Function** | a disjunct, stance adverbial, or mood adjunct | a punctuation mark indicating a slight pause or transition | topic/comment; given/new |

### 3.2. Alternative Structures

Two distinct alternative structures can be observed in the corpus. The first structure is characterized by the omission of the adverbial participle clause, where specific adverbial phrases are sporadically used independently as stylistic elements to indicate a manner of speech or convey a speaker's attitude. Examples include *krótko* 'briefly', *szczerze* 'sincerely/honestly', *nawiasem* 'parenthetically/incidentally', *inaczej* 'differently', *ściślej* 'more strictly', *ogólnie* 'generally', and *dokładniej* 'more accurately', among others. These forms primarily appear at the beginning of sentences, suggesting a structural limitation or preference in their usage. This positioning could be significant in setting the tone, context, or subject of the following sentence. These adverbials function more as linkages to previous discourse rather than as disjuncts, implying they are more integrative in nature. Rather than standing apart from the main sentence to provide separate commentary or perspective (as disjuncts typically do), these forms connect more directly to the preceding context or discourse. This suggests that they are used to build upon, contrast with, or otherwise

relate to what has already been said. These characteristics might affect the meaning conveyed by these single adverbials compared to their counterparts with *mówiąc*. The absence of *mówiąc* could lead to a more seamless integration into the flow of discourse, potentially altering the emphasis, tone, or interpretive nuances. Although such instances are relatively infrequent, the following examples are noticeable in the corpus:

(4) *Krótko*: doświadczenia z lat chłopięcych sprawiły, że umiałem się bić i to naprawdę skutecznie.

'*Briefly*: experiences from my boyhood years made it so that I knew how to fight, and quite effectively at that.' (*Tkacz iluzji*)

(5) *Szczerze*, w ogóle w tym temacie nie pracowałem.

'*Honestly*, I haven't worked on this topic at all.' (*Gazeta Wyborcza*)

(6) *Nawiasem*, rodzice Edwarda są wyznania prawosławnego, ale z powodu braku cerkwi chodzą do katolickiego kościoła.

'*By the way*, Edward's parents are of the Orthodox faith, but due to the lack of an Orthodox church, they attend a Catholic church.' (*Tygodnik Podhalański*)

(7) *Inaczej*: mój pościg świadczy o mnie.

'*Differently put*: my pursuit speaks for me.' (*Narrenturm*)

Such independent occurrences may stem from the high frequency of each adverbial expression and its meaning (cf. Kubicka 2017). Bybee's (2006; see also Bybee and Hopper 2001) research showed that expressions with high frequency tend to undergo phonological reduction more quickly than those with low frequency. This results from the brain's improved recognition through repetition (cf. Bybee 2006: 714). Morphologically complex forms that are frequently used might lose their inherent structure and free themselves from their original forms (Bybee 2006: 715). The increasing frequency of altered forms, in turn, arises from their utility within the communication system, essentially their meaning (cf. Schmid 2010: 117). Additionally, according to Birzer (2012: 242), the omission of the quasi-participle is possible when the adverb relates to the form of the utterance (*dokładniej* 'more accurately' or *ściślej* 'more strictly') or the speaker's attitude towards it (*szczerze* 'sincerely/honestly' or *uczciwiej* 'more honestly').

Furthermore, nearly all of the adverbial phrases that fulfill the role of disjuncts possess corresponding structures employing the participle 'speaking' within the corpus. Nevertheless, in contrast to the standard configuration of the participle construction (e.g., *krótko mówiąc* 'briefly speaking' or *szczerze mówiąc* 'sincerely speaking'), the present participle in these similar structures

precedes the adverbial phrases, thereby forming an adverbial clause featuring a present participle, as exemplified in (8) and (9):

(8) *Mówiąc szczerze*, ja też nie mam zbyt wielu doświadczeń.
'*Honestly speaking*, I also don't have too many experiences.' (*Tabu*)

(9) *Mówiąc krótko*—odczułem, co to jest strach przed faktami.
'In short, I felt what it's like to be afraid of facts.' (*Nie tylko biegun*)

These alternative configurations exhibit significantly lower occurrence rates compared to the participle constructions denoted as *krótko mówiąc* 'briefly speaking' or *szczerze mówiąc* 'sincerely speaking'. To illustrate, the phrase *krótko mówiąc* is found 1,826 times within the NKJP corpus, whereas *mówiąc krótko* appears only 344 times. Similarly, *szczerze mówiąc* 'sincerely speaking' is much more prevalent than *mówiąc szczerze*, with respective frequencies of 1,786 and 384 instances. This pattern applies to other phrases such as *prawdę mówiąc* 'truth speaking', *nawiasem mówiąc* 'parenthetically speaking', *inaczej mówiąc* 'differently speaking', *własciwie mówiąc* 'actually speaking', *ściślej mówiąc* 'more strictly speaking', *delikatnie mówiąc* 'delicately speaking', *ogólnie mówiąc* 'generally speaking', *najkrócej mówiąc* 'most briefly speaking', or *najogólniej mówiąc* 'most generally speaking', all of which occur more frequently than the reverse-order *mówiąc prawdę* 'speaking truth' (24 occurrences), *mówiąc nawiasem* 'speaking incidentally/parenthetically' (49 occurrences), *mówiąc inaczej* 'speaking differently' (301 occurrences), *mówiąc właściwie* 'speaking actually' (1 occurrence), *mówiąc ściślej* 'speaking more strictly' (72 occurrences), *mówiąc delikatnie* 'speaking delicately' (160 occurrences), *mówiąc ogólnie* 'speaking generally' (93 occurrences), *mówiąc najkrócej* 'speaking most briefly' (73 occurrences), or *mówiąc najogólmiej* 'speaking most generally' (69 occurrences). In addition to the distributional properties, Żabowska's (2020) study revealed that both constructions differ in their syntactic and semantic properties. These differences include the set and form of permissible adverbials, acceptable adverbial transformations, coexistence with negation, and their function. Specifically, [Adv] *mówiąc* functions as a metatextual unit of the system, while *mówiąc* [Adv] results from a quasi-participial operation.

Wiliński's (2025) in-depth, distinctive collexeme analysis of both the *mówiąc*-AC construction and the AC-*mówiąc* construction has provided new insights into their usage, roles in discourse, and semantic limitations. The comparative quantitative study highlights significant distinctions between the two. The *mówiąc*-AC construction, for instance, is less common in the corpus, recorded 4,434 times, yet it exhibits a broader variety of AC associations, totaling 208 distinct types. Impressively, 78 of these ACs are uniquely identified with this construction, suggesting that the *mówiąc*-AC construction is

more productive than the AC-*mówiąc* construction, as evidenced by its diverse range of ACs. Furthermore, the research reveals marked differences in the semantic restrictions applied to the ACs within each construction. While there is some overlap in ACs shared by both constructions, the study's focus on analyzing the most strongly associated lexemes of these constructions uncovers their different semantic patterns and preferences.

The *mówiąc*-AC construction, for example, does not typically collocate with *lekko* 'lightly' and is seldom used with *prawdę* 'truth' (24 occurrences), *nawiasem* 'parenthetically/incidentally' (49 occurrences), and *właściwie* 'actually' (1 occurrence). However, it frequently occurs with terms like *językiem* 'in the language of' (e.g., *sportowym*) 'in (sports) terms' (350 occurrences), *wprost* 'directly' (324 occurrences), *obrazowo* 'figuratively' (135 occurrences), *poważnie* 'seriously' (134 occurrences), *w skrócie* 'in short' (125 occurrences), *prościej* 'more simply' (103 occurrences), *serio* 'seriously' (74 occurrences), *kolokwialnie* 'colloquially' (74 occurrences), *słowami* 'in one's words' (61 occurrences), *w uproszczeniu* 'in simplification' (54 occurrences), *po prostu* 'simply' (53 occurrences), *bez ogródek* 'bluntly' (25 occurrences), *żargonem* (e.g., *prawniczym*) 'in (legal) jargon' (16 occurrences), *w przenośni* 'figuratively' (12 occurrences), *banalnie* 'banally' (11 occurrences), *ostrożnie* 'cautiously' (5 occurrences), and many others. Therefore, the primary role of this construction in discourse is to highlight the use of specialized jargon or terminology, make complex concepts more understandable, introduce figurative language, stress the importance of seriousness or sincerity in communication, offer further explanations or concise summaries, create a casual, friendly, or informal atmosphere, and approach sensitive or delicate subjects with tact and thoughtfulness.

In contrast, the AC-*mówiąc* construction is more selective in its ACs and is generally used to emphasize honesty, introduce side remarks, summarize points, or moderate language. This construction avoids certain terms such as *słowami* 'in one's words', *bez ogródek* 'bluntly', *żargonem* 'in jargon', *w przenośni* 'figuratively', *banalnie* 'banally', *pół żartem pół serio* 'half-jokingly, half-seriously', and *jaśniej* 'more clearly'. Moreover, it is rarely used with expressions like *językiem* 'in the language of' (4 occurrences), *wprost* 'directly' (3 occurrences), *serio* 'seriously' (3 occurrences), *w uproszczeniu* 'in simplification' (6 occurrences), *po prostu* 'simply' (6 occurrences), *precyzyjnie* 'precisely' (9 occurrences), *prosto* 'simply' (5 occurrences), *skrótowo* 'concisely' (4 occurrences), *fachowo* 'professionally' (1 occurrence), *dosadnie* 'bluntly/frankly' (2 occurrences), *żartem* 'jokingly' (2 occurrences), and several others. However, it shows a distinct preference for others, including *prawdę* 'truth', *nawiasem* 'parenthetically/incidentally', *krótko* 'briefly', *szczerze* 'sincerely/honestly', *inaczej* 'differently', *właściwie* 'actually', *ściślej* 'more strictly', *delikatnie* 'delicately/mildly', *lekko* 'lightly', *generalnie* 'generally', *ściśle* 'strictly', and *najłagodniej* 'most gently/mildly'. The five most common terms (*prawdę*, *nawiasem*, *krótko*, *szczerze*, and *inaczej*) account for 8,177 occurrences within this construction.

**Table 3.** The distribution of the twenty ACs across six registers

| Rank | ACs | Translation | jour | qu-sp | lit | Net-int | sc-di | conv |
|---|---|---|---|---|---|---|---|---|
| 1. | krótko | briefly | 614 | 534 | 244 | 44 | 67 | 7 |
| 2. | prawdę | truth | 591 | 187 | 655 | 63 | 20 | 2 |
| 3. | szczerze | sincerely/ honestly | 684 | 295 | 228 | 334 | 8 | 58 |
| 4. | nawiasem | in parentheses/ incidentally | 635 | 232 | 207 | 86 | 28 | 3 |
| 5. | inaczej | differently | 452 | 301 | 84 | 28 | 163 | 1 |
| 6. | delikatnie | delicately | 329 | 135 | 46 | 34 | 1 | 0 |
| 7. | ściślej | more strictly | 114 | 59 | 45 | 1 | 28 | 0 |
| 8. | ogólnie | generally | 78 | 38 | 11 | 20 | 11 | 3 |
| 9. | dokładniej | more accurately | 35 | 19 | 15 | 5 | 8 | 0 |
| 10. | najkrócej | most briefly | 85 | 30 | 3 | 11 | 5 | 0 |
| 11. | najogólniej | most generally | 78 | 31 | 1 | 4 | 15 | 0 |
| 12. | właściwie | actually | 5 | 1 | 66 | 0 | 0 | 0 |
| 13. | najprościej | most simply | 60 | 9 | 6 | 1 | 7 | 0 |
| 14. | ściśle | strictly | 19 | 15 | 25 | 1 | 5 | 0 |
| 15. | między nami | between us | 13 | 8 | 47 | 0 | 0 | 1 |
| 16. | oględnie | cautiously | 46 | 11 | 9 | 2 | 4 | 0 |
| 17. | łagodnie | mildly | 33 | 7 | 4 | 1 | 1 | 0 |
| 18. | uczciwie | honestly | 13 | 13 | 13 | 1 | 1 | 0 |
| 19. | brzydko | crudely/ bluntly | 13 | 14 | 1 | 4 | 1 | 4 |
| 20. | najdelikat-niej | most delicately | 20 | 12 | 4 | 0 | 0 | 0 |

(jour = journalism; qu-sp = quasi-spoken; lit = literature; Net-int = Internet-interactive; sc-di = scientific-didactic; conv = conversation)

Consequently, these expressions are not merely regarded as instances of the AC-*mówiąc* construction, but as distinct constructions in their own right, characterized by their high frequency and standard usage, which reinforces their role in specific communication contexts as established linguistic units.

### 3.3. Distributional Properties

Regarding the distribution of the AC-*mówiąc* construction across distinct registers within NKJP, Table 3 on p. 358 presents the recorded frequencies of the 20 most noteworthy ACs collocating with the participle *mówiąc* across six selected registers that comprise the corpus: journalism, literature, spoken conversation, quasi-spoken texts (including parliamentary transcripts), Internet sources (forums, chatrooms, mailing lists, etc.), and academic writing and scientific-didactic textbooks. Additionally, Figure 1 on the following page, created using Matplotlib in Python from the data in Table 3, visually illustrates the distribution of these ACs across different genres within the construction in question. The heatmap in Figure 1 employs varying color intensities, from white (lower frequencies) to darker blues (higher frequencies), to indicate the frequency of each AC in each genre, providing clear insights into which ACs are most commonly used in each genre and identifying genres where specific ACs occur more frequently.

A more thorough examination of the frequencies in Table 3 and Figure 1 reveals that the aforementioned construction predominantly appears in journalism, is relatively frequent in quasi-spoken texts, tends to occur often in literature, is less common in Internet sources and scientific-didactic texts, and is exceptionally rare in conversational contexts. Additionally, Table 3 and Figure 1 illustrate considerable disparities in the distribution of the discussed constructions in NKJP. For instance, pairings involving *krótko* 'briefly', *szczerze* 'sincerely/honestly', *nawiasem* 'parenthetically/incidentally', *inaczej* 'differently', *delikatnie* 'delicately/mildly', and many others show a higher prevalence in journalism and quasi-spoken discourse compared to literature. In contrast, the usage patterns involving *prawdę* 'truth', *właściwie* 'actually', *ściśle* 'strictly', and *między nami* 'between us' exhibit significantly greater frequencies in literature than in other linguistic registers. Furthermore, combinations such as *prawdę*, *szczerze*, *nawiasem*, and *delikatnie* appear more frequently in Internet sources than in scholarly and conversational contexts. Finally, configurations involving *krótko* 'briefly', *inaczej* 'differently', *ściślej* 'more strictly', *dokładniej* 'more accurately', *najogólniej* 'most generally', or *najprościej* 'most simply' are relatively more prevalent in scientific discourse compared to Internet sources and conversational language.

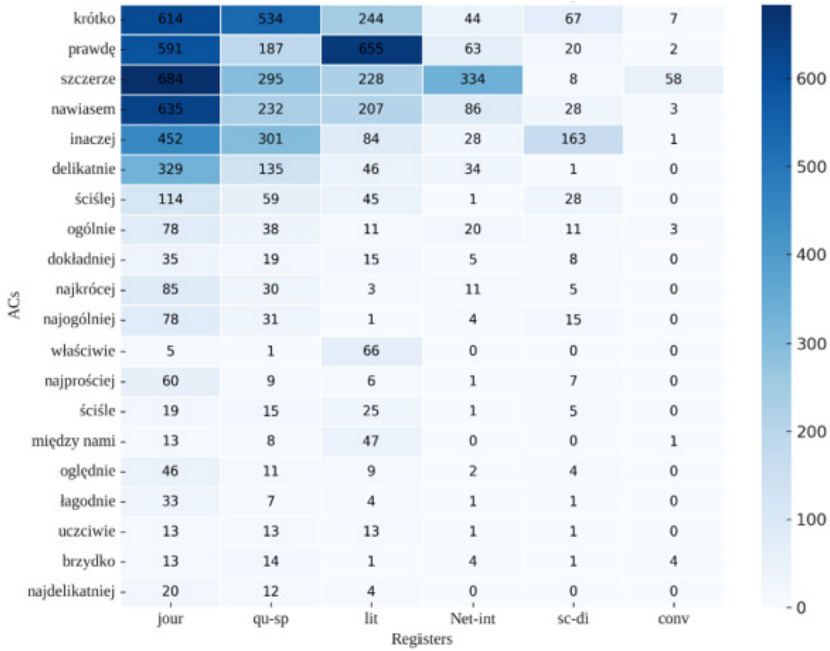


**Figure 1.** Visualization of the distribution of ACs across different genres

## 4. Results and Discussion

The initial frequencies of ACs extracted from the corpus included 130 distinct AC types, of which 46 appeared only once within the AC-*mówiąc* construction. However, due to space limitations in this context, this section will primarily analyze the quantitative findings regarding the 60 most highly attracted ACs within the construction. Nonetheless, some less attractive ACs and their frequencies of occurrence will also be mentioned in passing to provide a more comprehensive overview of the construction's usage patterns. Presented in Table 4 on p. 362 are the attraction and reliance scores for the 60 most noteworthy ACs, including the raw frequency of ACs within the construction, the overall frequency of all ACs within the construction, and the comprehensive frequency of these ACs across the entire corpus.

The quantitative data and results in both tables are organized and ranked according to the measure of attraction. These findings support the hypothesis that specific ACs are strongly associated with the AC-*mówiąc* construction. A brief analysis of the results indicates that the upper portion of Table 4 consists of ACs with notably high frequencies, such as *krótko* 'briefly', *prawdę* 'truth', *szczerze* 'sincerely/honestly', and *nawiasem* 'parenthetically/incidentally'. The most plausible explanation for the prominent positioning of these ACs in the

ranking is their substantial overall frequency within the NKJP corpus, which presumably significantly influences the statistical likelihood of their occurrence in the specific construction. For instance, due to their higher frequencies within the construction being studied, *krótko* 'briefly' (attraction score 16.65%) and *prawdę* 'truth' (attraction score 16.32%) achieved significantly higher attraction scores compared to *konkretnie* 'concretely/specifically' (attraction score 0.17%) and *potocznie* 'colloquially' (attraction score 0.16%).

In comparison, excluding *nawiasem* 'parenthetically/incidentally' (reliance score 84.21%), the ranking list for reliance reveals significantly higher scores associated with less common ACs that collocate with the construction. Typical examples include *najogólniej* 'most generally' (reliance score 24.58%), *najdelikatniej* 'most delicately' (reliance score 25.50%), and *najogłędniej* 'most cautiously' (reliance score 45.65%, as seen in Table 4). The underlying rationale for these substantial percentages stems from the statistical analysis employed to gauge reliance, which considered the overall frequency of each lexeme within the NKJP corpus. To elaborate, despite the higher frequency of *krótko* 'briefly' compared to *najogłędniej* 'most cautiously' within the specific construction, the latter secures a notably higher reliance score due to its lower frequency in NKJP (46 occurrences). Consequently, the reliance of *najogłędniej* on the AC-*mówiąc* construction holds greater significance (45.65%). *Nawiasem*, on the other hand, relies extensively on this pattern, appearing frequently within the construction (1,525 occurrences) and less regularly in alternative contexts across the corpus (286 occurrences).

### 4.1. ACs Introducing a Summary

The ACs listed in Table 4 can be categorized into distinct semantic groups. The first group consists of ACs that evoke the SUMMARIZATION frame. In this frame, a speaker concisely conveys a message to a recipient, providing a summary or succinct explanation of the key points or ideas previously discussed, as demonstrated in the examples below:

(10) W telewizji oglądałem sport, ale, [*krótko*]$_{\text{MANNER}}$ *mówiąc*, [często wydawał mi się nudny.]$_{\text{MESSAGE}}$

'I was watching sports on TV, but, *in short*, it often seemed boring to me.' (*Polityka*)

(11) [*W skrócie*]$_{\text{MANNER}}$ *mówiąc*, [celem jest kształtowanie ludzi, a nie fabrykowanie uczniów.]$_{\text{MESSAGE}}$

'*In short*, the goal is to shape individuals, not manufacture students.' (*Dziennik Zachodni*)

**Table 4.** The 60 most strongly attracted ACs of the AC-*mówiąc* construction

| Rank | ACs | Translation | a | x | y | Attraction | Reliance |
|---|---|---|---|---|---|---|---|
| 1. | krótko | briefly | 1826 | 10965 | 14599 | 16.65% | 12.51% |
| 2. | prawdę | truth | 1789 | 10965 | 10861 | 16.32% | 16.47% |
| 3. | szczerze | sincerely/honestly | 1786 | 10965 | 9291 | 16.29% | 19.22% |
| 4. | nawiasem | parenthetically/ incidentally | 1525 | 10965 | 1811 | 13.91% | 84.21% |
| 5. | inaczej | differently | 1251 | 10965 | 42079 | 11.41% | 2.97% |
| 6. | delikatnie | delicately | 588 | 10965 | 4560 | 5.36% | 12.89% |
| 7. | ściślej | more strictly | 320 | 10965 | 1744 | 2.92% | 18.35% |
| 8. | ogólnie | generally | 179 | 10965 | 7490 | 1.63% | 2.39% |
| 9. | dokładniej | more accurately | 153 | 10965 | 2698 | 1.40% | 5.67% |
| 10. | najkrócej | most briefly | 150 | 10965 | 770 | 1.37% | 19.48% |
| 11. | najogólniej | most generally | 146 | 10965 | 594 | 1.33% | 24.58% |
| 12. | właściwie | actually | 129 | 10965 | 33736 | 1.18% | 0.38% |
| 13. | najprościej | most simply | 97 | 10965 | 1039 | 0.88% | 9.34% |
| 14. | ściśle | strictly | 93 | 10965 | 7827 | 0.85% | 1.19% |
| 15. | między nami | between us | 86 | 10965 | 2881 | 0.78% | 2.99% |
| 16. | oględnie | cautiously | 76 | 10965 | 341 | 0.69% | 22.29% |

(**a** = frequency of the AC *krótko* in the AC-*mówiąc* construction; **x** = total frequency of all ACs in the construction; **y** = total frequency of the AC *krótko* in NKJP)

*—continued on next page—*

—*continued*—

| Rank | ACs | Translation | a | x | y | Attraction | Reliance |
|---|---|---|---|---|---|---|---|
| 17. | łagodnie | mildly | 59 | 10965 | 2193 | 0.54% | 2.69% |
| 18. | uczciwie | honestly | 47 | 10965 | 2645 | 0.43% | 1.78% |
| 19. | brzydko | crudely/bluntly | 38 | 10965 | 714 | 0.35% | 5.32% |
| 20. | najdelikatniej | most delicately | 38 | 10965 | 149 | 0.35% | 25.50% |
| 21. | poważnie | seriously | 38 | 10965 | 11327 | 0.35% | 0.34% |
| 22. | dokładnie | accurately | 37 | 10965 | 31063 | 0.34% | 0.12% |
| 23. | generalnie | generally | 32 | 10965 | 8005 | 0.29% | 0.40% |
| 24. | obrazowo | figuratively | 25 | 10965 | 566 | 0.23% | 4.42% |
| 25. | umownie | conventionally | 23 | 10965 | 493 | 0.21% | 4.67% |
| 26. | najoględniej | most cautiously | 21 | 10965 | 46 | 0.19% | 45.65% |
| 27. | w skrócie | in short | 20 | 10965 | 2691 | 0.18% | 0.74% |
| 28. | kolokwialnie | colloquially | 20 | 10965 | 156 | 0.18% | 12.82% |
| 29. | konkretnie | specifically | 19 | 10965 | 4993 | 0.17% | 0.38% |
| 30. | potocznie | colloquially | 17 | 10965 | 1130 | 0.16% | 1.50% |
| 31. | niełądnie | improperly | 15 | 10965 | 576 | 0.14% | 2.60% |
| 32. | precyzyjniej | more precisely | 14 | 10965 | 242 | 0.13% | 5.79% |
| 33. | brutalnie | brutally/bluntly | 13 | 10965 | 1864 | 0.12% | 0.70% |
| 34. | prościej | more simply | 13 | 10965 | 794 | 0.12% | 1.64% |

(**a** = frequency of the AC *krótko* in the AC-*mówiąc* construction; **x** = total frequency of all ACs in the construction; **y** = total frequency of the AC *krótko* in NKJP)

—*continued on next page*—

—*continued*—

| Rank | ACs | Translation | a | x | y | Attraction | Reliance |
|---|---|---|---|---|---|---|---|
| 35. | popularnie | popularly | 12 | 10965 | 718 | 0.11% | 1.67% |
| 36. | eufemistycznie | euphemistically | 12 | 10965 | 160 | 0.11% | 7.50% |
| 37. | po ludzku | humanly | 12 | 10965 | 963 | 0.11% | 1.25% |
| 38. | lekko | lightly | 11 | 10965 | 11005 | 0.10% | 0.10% |
| 39. | najłagodniej | most mildly | 11 | 10965 | 70 | 0.10% | 15.71% |
| 40. | szerzej | more broadly | 10 | 10965 | 3274 | 0.09% | 0.31% |
| 41. | precyzyjnie | precisely | 9 | 10965 | 4002 | 0.08% | 0.22% |
| 42. | żartobliwie | jokingly | 9 | 10965 | 977 | 0.08% | 0.92% |
| 43. | innymi słowy | in other words | 8 | 10965 | 2435 | 0.07% | 0.33% |
| 44. | nieskromnie | immodestly | 7 | 10965 | 249 | 0.06% | 2.81% |
| 45. | konkretniej | more specifically | 7 | 10965 | 187 | 0.06% | 3.74% |
| 46. | po prostu | simply | 6 | 10965 | 63360 | 0.05% | 0.01% |
| 47. | w uproszczeniu | in simplification | 6 | 10965 | 515 | 0.05% | 1.17% |
| 48. | z grubsza | roughly | 6 | 10965 | 1166 | 0.05% | 0.51% |
| 49. | otwarcie | openly | 5 | 10965 | 10864 | 0.05% | 0.05% |
| 50. | prosto | simply | 5 | 10965 | 9864 | 0.05% | 0.05% |
| 51. | elegancko | elegantly | 4 | 10965 | 1364 | 0.04% | 0.29% |
| 52. | dosadniej | more bluntly/frankly | 4 | 10965 | 100 | 0.04% | 4.00% |

(**a** = frequency of the AC *krótko* in the AC-*mówiąc* construction; **x** = total frequency of all ACs in the construction; **y** = total frequency of the AC *krótko* in NKJP)

—*continued on next page*—

—*continued*—

| Rank | ACs | Translation | a | x | y | Attraction | Reliance |
|---|---|---|---|---|---|---|---|
| 53. | jednym słowem | in one word | 4 | 10965 | 2539 | 0.04% | 0.16% |
| 54. | skrótowo | concisely | 4 | 10965 | 342 | 0.04% | 1.17% |
| 55. | praktycznie | practically | 4 | 10965 | 16316 | 0.04% | 0.02% |
| 56. | patetycznie | pathetically | 4 | 10965 | 225 | 0.04% | 1.78% |
| 57. | paradoksalnie | paradoxically | 4 | 10965 | 2267 | 0.04% | 0.18% |
| 58. | ogólniej | more generally | 4 | 10965 | 95 | 0.04% | 4.21% |
| 59. | obiektywnie | objectively | 4 | 10965 | 1457 | 0.04% | 0.27% |
| 60. | metaforycznie | metaphorically | 4 | 10965 | 157 | 0.04% | 2.55% |

(**a** = frequency of the AC *krótko* in the AC-*mówiąc* construction; **x** = total frequency of all ACs in the construction; **y** = total frequency of the AC *krótko* in NKJP)

The most prominent term within this semantic category is *krótko* 'briefly', which ranks at the top. It exhibits an attraction to the pattern in 16.65% of cases and relies on it in 12.51%. It is accompanied by the following ACs: *najkrócej* 'most briefly' (ranked 10th), *w skrócie* 'in short' (ranked 27th), *jednym słowem* 'in one word' (ranked 53rd), and *skrótowo* 'concisely' (ranked 54th). Additionally, four other ACs can be identified in the corpus: *lapidarnie* 'succinctly' (two occurrences), *lakonicznie* 'laconically' (two occurrences), *krócej* 'more briefly' (two occurrences), *hasłowo* 'in a slogan-like manner' (two occurrences), and *pokrótce* 'briefly' (one occurrence). All these ACs condense complex or lengthy content into a more manageable and digestible form, capturing the main points and essence of what was previously stated.

## 4.2. ACs Introducing a Candid Statement

The second category consists of ACs invoking the CANDIDNESS frame. In this frame, a speaker conveys their message about a specific topic honestly, considering the veracity and sincerity of the message, as exemplified in (2), (3), (5), and (12).

(12) [*Uczciwie*]$_{\text{MANNER}}$ *mówiąc*, [nie da się tego czytać.]$_{\text{MESSAGE}}$
'*Honestly speaking*, it's unreadable.' (*Gazeta Poznańska*)

The term *prawdę* 'truth' in rank 2, preceding *szczerze* 'sincerely/honestly' in rank 3 and *uczciwie* 'honestly' in rank 18, serves as the primary expression invoking this particular semantic frame and is distinguished as the most strongly attracted lexeme within this semantic category (with an attraction score of 16.32%). Additionally, *otwarcie* 'openly' ranks 49th, and *dosadniej* 'more bluntly/frankly' ranks 52nd, both of which also appear to evoke this frame. Two other words conveying similar meanings are *prawdziwie* 'truly' and *najuczciwiej* 'most honestly', which display singular occurrences within the structure. These phrases signal that the upcoming statement is meant to be forthright and truthful, often providing insight into the speaker's inner thoughts or beliefs. They emphasize the intention of transparent communication and are commonly used to establish a sincere and open tone in discourse.

## 4.3. ACs Introducing an Aside

The lexeme *nawiasem* 'parenthetically/incidentally' holds the fourth position in the ranking list. This is followed by *między nami* 'between us' and *na marginesie* 'by the way', which occupy ranks 15 (86 occurrences) and 61 (four occurrences), respectively. *Nawiasem* constitutes 13.91% of all occurrences of the construction being investigated within the corpus. Furthermore, it demonstrates

a significant dependence on this pattern (with a reliance score of 84.21%), indicating its use in other contexts to a relatively minor extent, comprising 16.47% of its usage.

Collectively, these lexemes evoke the ASIDE frame. In this frame, the speaker provides additional information, shares personal comments, or makes asides that may not be directly related to the main topic of conversation but are still relevant or interesting to share, as shown in (1) above and (13) below.

(13) [*Między nami*]$_{\text{MANNER}}$ *mówiąc*, [to niektórzy są teraz trochę rozczarowani.]$_{\text{MESSAGE}}$

'*Between us*, some are now slightly disappointed.' (*Początek*)

The phrases *na marginesie mówiąc* 'by the way speaking', *między nami mówiąc* 'between us speaking', and *nawiasem mówiąc* 'parenthetically speaking' frequently introduce a shift in focus or provide context for the upcoming statement, helping to create a more personal or conversational tone in discourse. Notably, the phrase *między nami mówiąc* seems to invoke the CONFIDENTIALITY frame. This frame involves discreetly or secretively sharing information, typically intended solely for those engaged in the conversation rather than for public or widespread dissemination.

## 4.4. ACs Introducing a Restatement

The next lexeme in the ranking is *inaczej* 'differently', which occupies the fifth position. A similar meaning is conveyed by *innymi słowy* 'in other words' at rank 43. Both of these lexemes evoke the semantic frame of REPHRASING. In this frame, the speaker or writer selects alternative words, sentence structures, or expressions to convey the same information or idea more clearly, concisely, or stylistically differently, as illustrated in (14).

(14) [*Inaczej*]$_{\text{MANNER}}$ *mówiąc*, [bez cechy planowości nie ma budżetu.]$_{\text{MESSAGE}}$

'*Put differently*, without the attribute of planning, there is no budget.' (*Finanse publiczne*)

The phrase *inaczej mówiąc* 'differently speaking' is used to enhance comprehension, avoid repetition, emphasize a particular point, or adapt the language to suit the context or audience.

## 4.5. ACs Introducing an Understatement

Another semantic category comprises ACs that evoke the semantic frame of UNDERSTATEMENT. This frame involves a speaker presenting a statement in a

way that deliberately downplays the significance, impact, or severity of a situation, often employed for rhetorical purposes or to soften the tone of the expression, as exemplified in (15).

(15) [*Delikatnie*]$_{\text{MANNER}}$ *mówiąc*, [on mija się z prawdą i nie wspomina o kosztach.]$_{\text{MESSAGE}}$
'*To put it mildly*, he is not being truthful and doesn't mention the cost.' (*Trybuna*)

Ranked sixth, *delikatnie* 'delicately/mildly' is the most significant lexeme in this category. It is followed by *oględnie* 'cautiously', *łagodnie* 'mildly', *najdelikatniej* 'most delicately', *najoględniej* 'most cautiously', *lekko* 'lightly', *najłagodniej* 'most gently/mildly', *subtelnie* 'subtly', *łagodniej* 'more mildly', and *subtelniej* 'more subtly', which are positioned at ranks 16, 17, 20, 26, 38, 39, 86, 92, and 99, respectively. These ACs imply that the truth or situation is more extreme or important than what is explicitly stated, thereby creating a sense of understated emphasis or irony.

## 4.6. ACs Introducing an Accurate Statement

The seventh position is occupied by *ściślej* 'more strictly'. It is accompanied by a set of synonymous ACs, including *dokładniej* 'more accurately', *właściwie* 'actually', *ściśle* 'strictly', *dokładnie* 'accurately', *konkretnie* 'concretely/specifically', *precyzyjniej* 'more precisely', *precyzyjnie* 'precisely', and *konkretniej* 'more specifically/concretely', which are ranked at the 9th, 12th, 14th, 22nd, 29th, 32nd, 41st, and 45th positions, respectively. These ACs occupy subsequent positions in the ranking list, signifying varying degrees of precision and specificity. Consequently, they can be defined relative to the ACCURACY frame. In this frame, a speaker introduces a statement that aims to provide a more detailed, exact, or specific description, emphasizing a high level of correctness or precision in the information conveyed, as provided in (16).

(16) [*Ściślej*]$_{\text{MANNER}}$ *mówiąc:* [mój wstyd i moja wina.]$_{\text{MESSAGE}}$
'*To be more precise*: my shame and my fault.' (*Kultura*)

## 4.7. ACs Introducing a General Statement

The ranking list also includes ACs such as *ogólnie* 'generally' (ranked 8th), *najogólniej* 'most generally' (ranked 11th), *generalnie* 'generally' (ranked 23rd), *popularnie* 'popularly' (ranked 35th), *szerzej* 'more broadly' (ranked 40th), *z grubsza* 'roughly' (ranked 48th), *ogólniej* 'more generally' (ranked 58th), *szeroko* 'broadly' (ranked 73rd, with two occurrences), and *ogółem* 'overall' (ranked

113th, with one occurrence), invoking the GENERALITY frame. In this frame, a speaker provides a broad or general overview or statement before delving into specific details, as illustrated in (17). It signals that the upcoming statement or explanation will cover a wide scope and may not address every possible exception or nuance.

(17) [*Ogólnie*]$_{\text{MANNER}}$ *mówiąc*, [jest źle, bardzo źle.]$_{\text{MESSAGE}}$
'*Generally speaking*, it's bad, very bad.' (*Tygodnik Tucholski*)

## 4.8. ACs Introducing a More Straightforward Explanation

The next category comprises ACs that evoke the SIMPLIFICATION frame. This frame accentuates the speaker's intent to break down complex ideas into more easily understandable elements, thereby improving access to information and fostering clarity and simplicity in communication, as illustrated in (18). Among these, *najprościej* 'most simply' ranks 13th and holds the utmost significance within this category. It is followed by *prościej* 'more simply', *po prostu* 'simply', *w uproszczeniu* 'in simplification', *prosto* 'simply', *trywialnie* 'trivially' (three occurrences), and *najbanalniej* 'most simply' (two occurrences), which are ranked at 34th, 46th, 47th, 50th, 66th, and 80th, respectively.

(18) [Zajmujemy się, [*najprościej*]$_{\text{MANNER}}$ *mówiąc*, zmianami pola grawitacyjnego ziemi.]$_{\text{MESSAGE}}$
'We deal with, *simply put*, changes in the Earth's gravitational field.'
(*Gazeta Wrocławska*)

## 4.9. ACs Introducing a Positive or Negative Evaluation

Another group comprises ACs related to both the NEGATIVE AND POSITIVE EVALUATION frames. The first frame is invoked by *brzydko* 'crudely/bluntly', *niełądnie* 'improperly', *ordynarnie* 'rudely' (one occurrence), *ordynarniej* 'more rudely' (one occurrence), ranked 19th, 31st, 110th, and 111th, respectively. On the other hand, the second frame is represented by *eufemistycznie* 'euphemistically', *elegancko* 'elegantly', *ładnie* 'nicely' (three occurrences), and *pięknie* 'beautifully' (one occurrence), ranked 36th, 51st, 64th, and 89th, respectively. In the former frame, a speaker introduces a statement that may be considered tactless, rude, impolite, offensive, or morally questionable due to its bluntness or lack of refinement, as exemplified in (19). In contrast, the latter frame introduces a more polite, euphemistic, or diplomatic way of expressing something that might otherwise be perceived as impolite, harsh, or harmful. This indicates that the speaker opts for milder or less direct phrasing for the sake of tact or

courtesy, even if the actual situation may warrant a stronger description, as demonstrated in (20).

(19) [Po drugie, mają na niego [*brzydko*]$_{\text{MANNER}}$ *mówiąc*, haka.]$_{\text{MESSAGE}}$
'Secondly, they have, *to put it crudely*, dirt on him.' (*Nakielski Czas*)

(20) Handel żywym towarem, czyli—[*elegancko*]$_{\text{MANNER}}$ *mówiąc*—transfery, stały się zjawiskiem powszechnym.
'Human trafficking, or—*to put it elegantly*—transfers, has become a common phenomenon.' (*CKM*)

These ACs offer nuanced ways to introduce statements based on factors such as directness, propriety, diplomacy, harshness, and sophistication, each contributing a specific tone, connotation, or context to the communication. For example, in (19), *brzydko mówiąc* introduces a direct, straightforward, or potentially harsh description. It implies that the speaker is about to express something in a more candid or unvarnished manner, without euphemisms or softened language. In contrast, in (20), *elegancko mówiąc* introduces a statement that might involve a graceful, refined, or stylish manner of expression. A negative evaluation is also expressed by *brutalnie* 'brutally/bluntly', ranked 33rd. *Brutalnie mówiąc* evokes the BLUNTNESS semantic frame, indicating that the following statement will be delivered in a direct, honest, and straightforward manner, without concern for softening or sugar-coating the content. It implies that the speaker intends to be forthright, even if it might sound harsh or uncomfortable.

## 4.10. ACs Introducing a Figurative or Literal Explanation

The next group includes ACs that can be interpreted relative to the FIGURATIVITY and LITERAL DESCRIPTION frames. The FIGURATIVITY frame is represented by *obrazowo* 'figuratively' (ranked 24th), *metaforycznie* 'metaphorically' (ranked 60th), *symbolicznie* 'symbolically' (ranked 74th, with two occurrences), and *przenośnie* 'figuratively' (ranked 104th, with one occurrence). Meanwhile, the LITERAL DESCRIPTION is invoked by *dosłownie* 'literally' (ranked 125th, with one occurrence).

In the first frame, a speaker uses figurative or symbolic language, often incorporating metaphors and vivid imagery to convey a more imaginative and illustrative message. For example, the phrase *obrazowo mówiąc* 'figuratively speaking' in (21) suggests that the ensuing statement should be interpreted beyond its literal meaning, emphasizing creative language to evoke a specific mental image or emotional response.

(21) [*Obrazowo*]$_{\text{MANNER}}$ *mówiąc*, [jesteśmy jak chłopak, który w garażu rozkręcił motocykl.]$_{\text{MESSAGE}}$

'*Figuratively speaking*, we are like a boy who disassembled a motorcycle in the garage.' (*Lato nieśmiertelnych*)

In the second frame, the speaker aims to describe a concept or situation clearly and accurately without incorporating imaginative or metaphorical elements. For example, the phrase *dosłownie mówiąc* 'literally speaking' presents information straightforwardly and factually, lacking imagination and avoiding metaphorical interpretation, as demonstrated in *Nagle, dosłownie mówiąc, budynek zawalił się* 'Suddenly, the building, literally speaking, collapsed'.

## 4.11. ACs Introducing Conventional Usage

The next lexeme following *obrazowo* 'figuratively' in the ranking list is *umownie* 'conventionally', which is ranked 25th. This lexeme, along with three others—*tradycyjnie* 'traditionally' (ranked 94th, occurring once), *zwyczajnie* 'ordinarily/customarily' (ranked 93rd, occurring once), and *normalnie* 'normally' (ranked 115th, occurring once)—can be associated with the CONVENTIONALITY frame. In this frame, a speaker uses language that aligns with established norms, standard practices, widely accepted meanings, or agreed-upon understandings within a particular culture or community, as illustrated in (22).

(22) [W kolejnym pojedynku Unia pokonała, na [*umownie*]$_{\text{MANNER}}$ *mówiąc* swoim torze Polonię Piła 72:19.]$_{\text{MESSAGE}}$

'In the next match, Unia defeated Polonia Piła with a score of 72:19, *conventionally speaking*, on their home track.' (*Gazeta Pooznańska*)

In this context, *umownie mówiąc* implies that the term *swój tor* 'home track' may not refer to the team's literal home but is used in a way that is generally understood or accepted within the context of speedway or the conversation.

## 4.12. ACs Introducing Colloquial Usage

The bottom of the ranking list in Table 4 also includes *kolokwialnie* 'colloquially' and *potocznie* 'colloquially' at ranks 28 and 30. The meanings of these adverbs can be defined relative to the COLLOQUIAL STATEMENT frame. In this frame, a speaker introduces a sentence or utterance crafted in a casual, conversational style, employing language typical of informal spoken communication. For example, *kolokwialnie mówiąc* in (23) indicates that the upcoming information reflects natural conversation or everyday speech. This adds a touch of familiarity and relatability to the sentence while acknowledging that the

description might not be the most precise or detailed; rather, it serves as an accessible and simplified means of conveying the idea.

(23) [Nie wygraliśmy tego meczu, myśmy go, [*kolokwialnie*]$_{\text{MANNER}}$ *mówiąc*, przepchnęli.]$_{\text{MESSAGE}}$
'We didn't win this match; we, *colloquially speaking*, pushed it through.' (*Życie Podkarpackie*)

Similarly, *pospolicie* 'commonly', ranking 78th and occurring twice in the corpus, also conveys a comparable meaning. It indicates that the subsequent statement is easily understandable and clearly presented, often using colloquial language.

## 4.13. Other ACs

Within the ranking list, notable lexemes include individual ACs, namely *poważnie* 'seriously', *po ludzku* 'humanly', *żartobliwie* 'jokingly', *nieskromnie* 'immodestly', *praktycznie* 'practically', *patetycznie* 'pathetically', *paradoksalnie* 'paradoxically', and *obiektywnie* 'objectively', occupying ranks 21, 37, 42, 44, 55, 56, 57, and 59, respectively. The meanings of these ACs can be attributed to various distinct semantic frames.

The first AC evokes the SERIOUSNESS frame, as in (24). In this sentence, the speaker approaches the topic without frivolity or humor, indicating that the forthcoming information should be taken seriously, with no intention of jest or lightheartedness. The meaning of *po ludzku* 'humanly' can be interpreted relative to the RELATABILITY frame, in which a speaker states an opinion about a certain entity or state of affairs by highlighting that the ensuing statement is expressed in a manner that is relatable, understandable, and pertinent to human experiences, as demonstrated in (25).

(24) [Interesują nas od wielu lat tylko, [*poważnie*]$_{\text{MANNER}}$ *mówiąc*, sprawy społeczne.]$_{\text{MESSAGE}}$
'We have been interested in social matters for many years, *speaking seriously*.' (*Trybuna Śląska*)

(25) [*Po ludzku*]$_{\text{MANNER}}$ *mówiąc*, [im człowiek starszy, tym chętniej poprawia i uzupełnia, co raz napisał.]$_{\text{MESSAGE}}$
'*Speaking in human terms*, the older a person gets, the more willingly they revise and supplement what they once wrote.' (*Polityka*)

The adverb *żartobliwie* 'jokingly' evokes the JEST frame, as reflected in (26). In this frame, the speaker adopts a humorous or jesting tone while formu-

lating a playful or amusing comment, prompting the audience to interpret the statement with a touch of lightheartedness rather than as a strictly literal assertion. *Nieskromnie* 'immodestly' can be described in terms of the semantic frame of SELF-PROMOTION or BOASTING. In this frame, the speaker conveys a message highlighting their achievements or positive qualities, even if they lack humility. For example, as in (27), the addition of *nieskromnie* suggests that the speaker is about to make a statement that might be perceived as immodest or self-aggrandizing. *Praktycznie* 'practically' invokes the PRACTICALITY frame. In this frame, a speaker presents a complicated topic or situation in a manner that emphasizes its practicality, usefulness, or applicability in real-life scenarios, as in (28). *Patetycznie* represents the frame of ELEVATION, in which the speaker uses elevated or grandiloquent language to convey a sense of seriousness, significance, or solemnity, as presented in (29).

(26) [Ci uczniowie to, [*żartobliwie*]$_{\text{MANNER}}$ *mówiąc*, tacy półwariaci, którym nieustannie buzują hormony]$_{\text{MESSAGE}}$ (…)

'These students, *jokingly speaking*, are a bit eccentric, with hormones constantly in a frenzy (…)' (*Polityka*)

(27) [I to było takie zadanie, które wymagało, no, [*nieskromnie*]$_{\text{MANNER}}$ *mówiąc*, gruntownej wiedzy prawniczej.]$_{\text{MESSAGE}}$

'It was such a task that required, well, *immodestly speaking*, a thorough legal knowledge.'

(28) [W debacie nad informacją o źródłach (…), a [*praktycznie*]$_{\text{MANNER}}$ *mówiąc*, o funkcjonowaniu funduszów pomocowych w Polsce, przedstawiam]$_{\text{MESSAGE}}$ (…)

'In the debate on the information regarding the sources (…), and *practically speaking*, about the functioning of aid funds in Poland, I propose that (…)' (*Sprawozdanie stenograficzne z obrad Sejmu RP*)

(29) [*Patetycznie*] $_{\text{MANNER}}$ *mówiąc*, [historia daje Kaczyńskiemu i Tuskowi szansę, jakiej nie było bodaj od 1989 roku.] $_{\text{MESSAGE}}$

'*Pathetically speaking*, one might say that history is granting Kaczyński and Tusk an opportunity that arguably has not been present since 1989.' (*Ozon*)

The meaning of *paradoksalnie* 'paradoxically' can be described in relation to a semantic frame associated with CONTRADICTION or IRONY. Within this frame, a speaker introduces a statement that may initially seem contradictory or counterintuitive given the context; however, it is presented to highlight an unexpected or ironic aspect. For example, *paradoksalnie mówiąc* in (30) illustrates the paradox between the teaching of love and the necessity of dying on

the cross, adding an element of irony or surprise to the statement. Additionally, *obiektywnie* 'objectively' is used in this construction to convey that the statement following it aims to be presented from an objective, unbiased standpoint, free from personal emotions or subjective opinions, as in (31). Thus, the semantic frame it evokes is that of OBJECTIVITY. In this context, the speaker seeks to present information or a situation without personal biases or emotions, concentrating on the facts and maintaining a neutral tone.

(30) [*Paradoksalnie*] $_{\text{MANNER}}$ *mówiąc*, [musiał umierać na krzyżu, bo nauczał, że Bóg jest Miłością i trzeba kochać swoich nieprzyjaciół.] $_{\text{MESSAGE}}$

'*Paradoxically speaking*, he had to die on the cross because he taught that God is Love and one must love their enemies.'

(*Czy Jezus to powiedział? Zagadki ewangeliczne*)

(31) [*Obiektywnie*] $_{\text{MANNER}}$ *mówiąc*, [służyło to dezintegracji komunistycznego państwa i przybliżało jego totalny krach.] $_{\text{MESSAGE}}$

'*Objectively speaking*, this served the disintegration of the communist state and brought its complete collapse closer.'

(*Edward Gierek: życie i narodziny legendy*)

## 5. Concluding Remarks

The primary objective of this research was to explore the inherent qualities of the AC-*mówiąc* construction, an adverbial participial structure that has not been examined using frame semantics, usage-based construction grammar, and quantitative corpus-based linguistics. Diverging from prior publications that primarily focused on the most prevalent instances of its usage, this empirical study offers fresh perspectives on the construction itself and its preference for specific categories of ACs. Consequently, this study significantly advances our understanding of the AC-*mówiąc* construction across diverse linguistic fields.

Firstly, the findings reveal that the AC-*mówiąc* construction is a partially schematic pattern, pairing form and meaning/function in its own right. This allows hearers and readers to identify combinations such as *krótko mówiąc* 'briefly speaking' or *nawiasem mówiąc* 'in parentheses speaking' as conventional expressions. These combinations can be treated as constructs, representing the concrete instances or tokens that the construction instantiates. The pattern can be viewed as a cognitive schema, generalization, abstract type, or mental representation that captures the construction's general features. This schema includes one fixed component (the participle *mówiąc*) and a more flexible slot that can accommodate certain ACs but not others (e.g., the instrumentals *słowami* 'in one's words', *żargonem* 'in jargon', or *slangiem* 'in slang'). The

construction displays substantial variation in the lexical items that can appear in the adverbial slot and in the grammatical units that instantiate it. This constructional schema is very productive, allowing speakers to generate creative utterances and process new complex expressions based on that construction with a high degree of cognitive ease. The productivity of the construction is manifested through its high type frequency and the occurrence of low-frequency instantiations. The corpus evidence shows that the construction is represented by 130 types, with a significant portion of these types (68) appearing fewer than four times in the construction.

Secondly, this study offers a fresh perspective on both the discourse function of the construction and the semantics of ACs appearing in the adverbial slot of this specific construction. It provides a more detailed classification of adverbial expressions based on semantic frames and their specific functions within these frames. The AC-*mówiąc* construction effectively conveys a structured implication, wherein a speaker communicates a specific message on a particular topic to an intended recipient in a defined manner or style. This construction serves various roles in discourse. Broadly speaking, it functions as a pragmatic marker within discourse, allowing speakers to comment on the manner or style used to convey the message and, in turn, express their perspectives, attitudes, positions, intentions, or viewpoints regarding the statement. Furthermore, it performs diverse functions within discourse, depending on the adverbs or adverbial phrases associated with the construction.

For example, the AC-*mówiąc* construction is frequently used in discourse to introduce summaries (e.g. *krótko mówiąc* 'briefly speaking', *najkrócej mówiąc* 'most briefly speaking', or *w skrócie mówiąc* 'in short speaking'), candid statements (*prawdę mówiąc* 'truth speaking', *szczerze mówiąc* 'sincerely/honestly speaking', or *uczciwie mówiąc* 'honestly speaking'), asides (*na marginesie mówiąc* 'by the way speaking', *między nami mówiac* 'between us speaking', and *nawiasem mówiąc* 'parenthetically speaking'), restatements (*inaczej mówiąc* 'differently speaking' and *innymi słowy mówiąc* 'in other words speaking'), understatements (*delikatnie mówiąc* 'delicately/mildly speaking', *oględnie mówiąc* 'cautiously speaking', or *łagodnie mówiąc* 'mildly speaking'), accurate statements (*ściślej mówiąc* 'more strictly speaking', *dokładniej mówiąc* 'more accurately speaking', or *precyzyjnie mówiąc* 'precisely speaking'), general statements (*ogólnie mówiąc* 'generally speaking', *najogólniej mówiąc* 'most generally speaking', or *generalnie mówiąc* 'generally speaking'), simpler explanations (*najprościej mówiąc* 'most simply speaking', *prościej mówiąc* 'more simply speaking', or *w uproszczeniu mówiąc* 'in simplification speaking'), positive or negative evaluations (*brzydko mówiąc* 'crudely/bluntly speaking', *niełądnie mówiąc* 'improperly speaking', *ordynarnie mówiąc* 'rudely speaking', *eufemistycznie mówiąc* 'euphemistically speaking', *elegancko mówiąc* 'elegantly speaking', or *ładnie mówiąc* 'nicely speaking'), figurative or literal explanations (*obrazowo mówiąc* 'figuratively speaking', *metaforycznie mówiąc* 'metaphorically speaking',

*symbolicznie mówiąc* 'symbolically speaking', or *dosłownie mówiąc* 'literally speaking'), conventional usage (*umownie mówiąc* 'conventionally speaking', *tradycyjnie mówiąc* 'traditionally speaking', or *zwyczajnie mówiąc* 'ordinarily speaking'), and colloquial usage (*kolokwialnie mówiąc* 'colloquially speaking' or *potocznie mówiąc* 'colloquially speaking'). In addition, it is occasionally used to convey seriousness (*poważnie mówiąc* 'seriously speaking'), introduce a human perspective (*po ludzku mówiąc* 'humanly speaking'), add humor or playfulness (*żartobliwie mówiąc* 'jokingly speaking'), express immodesty (*nieskromnie mówiąc* 'immodestly speaking'), suggest practicality (*praktycznie mówiąc* 'practically speaking'), employ elevated language for seriousness (*patetycznie mówiąc* 'pathetically speaking'), signal paradox (*paradoksalnie mówiąc* 'paradoxically speaking'), and present objective information (*obiektywnie mówiąc* 'objectively speaking'). Interestingly, unlike its English counterpart, this construction is not generally used to specify a particular field of reference (e.g., politically, geographically, or historically speaking). The only exceptions to this norm are *teologicznie* 'theologically', *psychicznie* 'psychologically', *nowomodnie* 'fashionably', and *sądownie* 'judicially', all of which appear once in the construction.

Thirdly, the findings of this study reveal that the construction exhibits a notable tendency to co-occur with distinct categories of ACs, invoking the following semantic frames: STATEMENT, CANDIDNESS, SUMMARIZATION, ASIDE, CONFIDENTIALITY, REPHRASING, UNDERSTATEMENT, ACCURACY, GENERALITY, SIMPLIFICATION, NEGATIVE EVALUATION, POSITIVE EVALUATION, BLUNTNESS, FIGURATIVITY, LITERAL DESCRIPTION, CONVENTIONALITY, COLLOQUIAL STATEMENT, SERIOUSNESS, RELATABILITY, SELF-PROMOTION, PRACTICALITY, ELEVATION, CONTRADICTION, OBJECTIVITY, JEST, BOASTING, or IRONY. Many of these ACs have not been previously recognized in discussions regarding the use of adverbial expressions, while others (such as *oględnie* 'cautiously', *nieskromnie* 'immodestly', *żartobliwie* 'jokingly', *w uproszczeniu* 'in simplification', *popularnie* 'popularly', *eufemistycznie* 'euphemistically', *elegancko* 'elegantly', or *patetycznie* 'pathetically') have only been briefly mentioned as associated elements of the participle *mówiąc*.

Fourthly, the findings of the current investigation have significantly enhanced our understanding of the distributional characteristics of the AC-*mówiąc* construction across various linguistic registers. To illustrate, the analysis of its prevalence across the six genres within the National Corpus of Polish (NKJP) has revealed that this construction primarily appears in journalism. It maintains a relatively frequent presence in quasi-spoken texts and demonstrates a tendency for fairly frequent occurrences in literature as well. On the other hand, it shows a comparatively lower frequency in Internet sources and in texts of a scientific-didactic nature. Notably, it is infrequent in conversational contexts.

Lastly, the current usage-based study has significantly enhanced our understanding of the pattern's status while providing a plausible rationale for

its recurrent pairing with specific adverbial types. Regarding its status, the AC-*mówiąc* pattern satisfies all the criteria outlined by Goldberg (2006) and Hilpert (2019) for identifying a linguistic pattern as a construction: It possesses sufficient frequency, deviates from canonical participle forms, exhibits idiosyncratic features, displays partial non-compositionality of meaning, and shows a collocational preference for specific categories of adverbs and adverbial expressions. The main explanation for the frequent co-occurrence of the identified ACs with the participle *mówiąc* may stem from the construction's semantics and its discourse functions. Nevertheless, the repeated use of these ACs within the pattern could potentially reinforce and influence its semantic and discourse-functional characteristics, ultimately leading to its entrenchment as a distinct symbolic unit—specifically, a combination of form and meaning/function. In other words, the frequent use of various AC types in this construction and the high token frequency of prototypical constructions (e.g., *krótko mówiąc* 'briefly speaking' with 1,826 occurrences) appear to contribute to the solidification of the AC-*mówiąc* schema, thereby establishing the constructional status of this pattern.[3]

In conclusion, it is essential to consider several caveats when interpreting these findings. Firstly, this investigation was based solely on corpus data sourced from the NKJP. Secondly, the quantitative evidence for all ACs that co-occur with *mówiąc* could not be thoroughly analyzed due to spatial limitations. Thirdly, a detailed comparative analysis of the distribution patterns between this construction and the *mówiąc*-AC construction was not feasible in this study. Consequently, future research endeavors may involve comparing this construction with alternative structures to detect subtle differences in distributional usage. Furthermore, it would be beneficial to explore the occurrence of this construction in different languages, specifically focusing on uncovering subtle variations in its use across various forms of written and spoken registers.

## Sources

[NKJP] National Corpus of Polish. Available at: http://www.nkjp.uni.lodz.pl/index_adv.jsp.

The FrameNet project. Available at: https://framenet.icsi.berkeley.edu/framenet_data.

[3] For a similar perspective, refer to Casenhiser and Goldberg (2005), Arnon and Snider (2010), and Goldberg (2019).

## References

Arnon, Inbal, and Neal Snider. (2010) "More than words: Frequency effects for multi-word phrases". *Journal of memory and language* 62: 67–87.

Barlow, Michael, and Suzanne Kemmer, eds. (2000) *Usage-based models of language*. Chicago: The University of Chicago Press.

Bednarek, Adam. (1989) *Wykładniki leksykalne ekwiwalencji. Analiza semantyczna wyrażeń typu czyli*. Toruń, Poland: TNT.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. (1999) *Longman grammar of spoken and written English*. Harlow: Longman.

Birzer, Sandra. (2012) "From subject to subjectivity: Russian discourse structuring elements based on the adverbial participle *govorja* 'speaking'". *Russian linguistics* 36(3): 221–49.

————. (2015) "Generally speaking, connectivity and conversation management combined: The functions of Russian *voobšče govorja* and Polish *ogólnie mówiąc*". *Russian linguistics* 39(1): 81–115.

————. (2017) "Assessing the role of pattern and matter replication in the development of Polish discourse structuring elements based on non-finite 'verba dicendi'". *Zeitschrift für slavische Philologie* 73(1): 159–85.

Bogusławski, Andrzej, and Jan Wawrzyńczyk. (1993) *Polszczyzna, jaką znamy. Nowa sonda słownikowa*. Warsaw: Uniwersytet Warszawski. Katedra Lingwistyki Formalnej.

Boyd, Jeremy K., and Adele E. Goldberg. (2011) "Learning what not to say: The role of statistical preemption and categorization in 'a'—adjective production". *Language* 87(1): 1–29.

Bybee, Joan. (2006) "From usage to grammar: The mind's response to repetition". *Language* 84(4): 711–33.

Bybee, Joan, and Clay Beckner. (2010) "Usage-based theory". Bernd Heine and Heiko Narrog, eds. *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press, 827–55.

Bybee, Joan, and Paul J. Hopper, eds. (2001) *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Casenhiser, Devin, and Adele Goldberg. (2005) "Fast mapping between a phrasal form and meaning". *Developmental science* 8(6): 500–508.

Croft, William, and Alan D. Cruse. (2004) *Cognitive linguistics.* Cambridge, UK: Cambridge University Press.

Diessel, Holger. (2017) "Usage-based linguistics". Mark Aronoff, ed. *Oxford research encyclopedia of linguistics.* New York: Oxford University Press.

————. (2019) "Usage-based construction grammar". Ewa Dąbrowska and Dagmar Divjak, eds. *Cognitive linguistics: A survey of linguistic subfields*. Berlin, New York: Mouton de Gruyter, 50–80.

Fillmore, Charles J. (2006) "Frame semantics". Dirk Geeraerts, ed. *Cognitive linguistics: Basic readings.* Berlin, New York: Mouton de Gruyter, 373–400.

Fillmore, Charles J., and Colin Baker. (2010) "A frames approach to semantic analysis". Bernd Heine and Heiko Narrog, eds. *The Oxford handbook of linguistic analysis.* Oxford: Oxford University Press, 313–40.

Fillmore, Charles J., Chris Johnson, and Miriam Petruck. (2003) "Background to FrameNet". *International journal of lexicography* 16: 235–51.

Goldberg, Adele. (2006) *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Goldberg, Adele. (2013) "Constructionist approaches". Thomaas Hoffmann and Graeme Trousdale, eds. *The Oxford handbook of construction grammar.* Oxford: Oxford University Press, 15–31.

————. (2019) *Explain me this: Creativity, competition, and the partial productivity of constructions.* Princeton: Princeton University Press.

Grochowski, Maciej. (2002) "Wielowyrazowe jednostki funkcyjne. Wprowadzenia do problematyki". Andrzej M. Lewicki, ed. *Problemy frazeologii europejskiej V.* Lublin: Norbertinum, 35–50.

————. (2007) "Partykuły właściwe o postaci związków frazeologicznych. Wprowadzenie do opisu". Andrzej M. Lewicki, ed. *Problemy frazeologii europejskiej VIII.* Lublin: Norbertinum, 77–87.

Halliday, Michael A. K. (2004) *An introduction to functional grammar.* 3rd ed., revised by Christian M. I. M. Matthiessen. London: Edward Arnold.

Hilpert, Martin. (2019) *Construction grammar and its application to English.* Edinburgh: Edinburgh University Press.

[ISJP]. (2000) *Inny słownik języka polskiego.* Mirosław Bańko, ed. Warsaw: Wydawnictwo Naukowe PWN.

Kisiel, Anna, and Magdalena Żabowska. (2011) "O zakresie klasy partykuł w *Słowniku gniazdowym partykuł polskich*". *Polonica* 31: 113–32.

Kleszczowa, Krystyna. (2015) *U źródeł polskich partykuł. Derywacja funkcjonalna, przemiany, zaniki.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Kubicka, Emilia. (2017) "Jak mówimy jakoś mówiąc? Formalne i semantyczne właściwości adwerbialnych uzupełnień quasi-imiesłowowego mówiąc". *LingVaria* 12(23): 99–113.

Michaelis, Laura A. (2004) "Type shifting in construction grammar: An integrated approach to aspectual coercion". *Cognitive linguistics* 15(1): 1–67.

Moroz, Andrzej. (2007) "Uwagi o ciągach parentetycznych z segmentem *mówiąc*". Joanna Kamper-Warejko and Iwona Kaproń-Charzyńska, eds. *Studia nad słownictwem dawnym i współczesnym języków słowiańskich.* Toruń: Wydawnictwo Naukowe UMK, 187–94.

————. (2010) *Parenteza ze składnikiem czasownikowym we współczesnym języku polskim*. Toruń: Wydawnictwo Naukowe UMK.

Ożóg, Kazimierz. (1990) *Leksykon metatekstowy współczesnej polszczyzny mówionej. Wybrane zagadnienia*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Perek, Florent. (2023) "Construction Grammar and usage-based theory". Manuel Díaz-Campos and Sonia Balasch, eds. *The handbook of usage-based linguistics*. Oxford: Wiley-Blackwell, 561–79.

Pęzik, Piotr. (2012) "Wyszukiwarka PELCRA dla danych NKJP". Adam Przepiórkowski, Mirosław Bańko, Rafał Górski, and Barbara Lewandowska-Tomaszczyk, eds. *Narodowy Korpus Języka Polskiego.* Warsaw: Wydawnictwo Naukowe PWN, 253–79.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. (1985) *A comprehensive grammar of the English language*. New York: Longman.

Schmid, Hans-Jörg. (2000) *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin: Mouton de Gruyter.

————. (2010) "Does frequency in text instantiate entrenchment in the cognitive system?". Dylan Glynn and Kerstin Fischer, eds. *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Berlin, New York: Mouton de Gruyter, 101–33.

Schmid, Hans-Jörg, and Helmut Küchenhoff. (2013) "Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings". *Cognitive linguistics* 24(3): 531–77.

Stępień, Marzena. (2014) *Wyrażenia parentetyczne w strukturze wypowiedzi—właściwości semantyczne, składniowe, prozodyczne*. Warsaw: Wydział Polonistyki Uniwersytetu Warszawskiego: BEL Studio.

————. (2015) "To samo czy inne? Właściwości prozodyczne tzw. wyrażeń quasi-imiesłowowych". Magdalena Danielewiczowa, Joanna Bilińska, Katarzyna Doboszyńska-Markiewicz, and Joanna Zaucha, eds. *Sens i brzmienie*. Warsaw: Wydawnictwo Uniwersytetu Kardynała Stefana Wyszyńskiego, 145–69. [Prace językoznawcze Instytutu Filologii Polskiej UKSW, 7.]

[SWJP]. (1996) *Słownik współczesnego języka polskiego*. Bogusław Dunaj, ed. Warsaw: Wydawnictwo Wilga.

[USJP]. (2004) *Uniwersalny słownik języka polskiego*. Stanisław Dubisz, ed. Warsaw: Wydawnictwo Naukowe PWN.

Weiss, Daniel. (2005) "Nowe przyimki o pochodzeniu imiesłowowym?" Maciej Grochowski, ed. *Przysłówki i przyimki. Studia ze składni i semantyki języka polskiego*. Toruń: Wydawnictwo Naukowe UMK, 177–207.

Wiliński, Jarosław. (2021) "Nouns in the *Be of N*-construction: A corpus-based investigation". *Studies in logic, grammar and rhetoric* 66(3): 747–68.

————. (2025) "*Speaking*-AC vs. AC-*speaking* in Polish: A distinctive collexeme analysis". *Prace Językoznawcze* 27(1): 95–112.

Wróbel, Henryk. (1975) *Składnia imiesłowów czynnych we współczesnej polszczyźnie*. Katowice: Spółka Wydawnicza OD NOWA.

[WSJP PAN]. (2006–24) *Wielki słownik języka polskiego*. Piotr Żmigrodzki, ed. Available at: https://wsjp.pl/.

Żabowska, Magdalena. (2009) "Wyrażenia metatekstowe w funkcji parentezy". Mirosław Skarżyński and Anna Czelakowska, eds. *Język z różnych stron widziany*. Kraków: Wydział Polonistyki UJ, 157–66.

Żabowska, Magdalena. (2020) "Szyk wyrażeń jako cecha dystynktywna w polu komentarzy metatekstowych: *[Adv] mówiąc*, _ vs. *mówiąc [Adv]*". *Poradnik Językowy* 9: 21–37.

## Appendix: Sources of Examples

(1) Source: *Gazeta Wyborcza*; Title: "Kolczyki Izoldy"; Author: Jaroslaw Krawczyk; Publication date: 2 December 1994; Register: Journalism/Daily newspaper.

(2) Source: *O jutrze i górach*; Register: Spoken/Casual conversation.

(3) Source: *Oni*; Author: Teresa Torańska; Publication date: 1985; Register: Publicism/Book.

(4) Source: *Tkacz iluzji*; Author: Ewa Białołęcka; Publication date: 2004; Register: Literature/Book.

(5) Source: *Gazeta Wyborcza*; Author: Piotr Lipiński; Publication Date: 26 September 1997; Register: Journalism/Daily newspaper.

(6) Source: *Tygodnik Podhalański nr 19*; Author: Lenur Junusow; Publication date: 1999; Register: Journalism/Weekly newspaper.

(7) Source: *Narrenturm*; Author: Andrzej Sapkowski; Publication date: 2002; Register: Literature/Book.

(8) Source: *Tabu*; Author: Kinga Dunin; Publication Date: 1998; Register: Literature/Book.

(9) Source: *Nie tylko biegun*; Authors: Marek Kamiński, Wojciech Moskal, Sławomir Swerpel; Publication Date: 1996; Register: Nonfiction/Book.

(10) Source: *Polityka nr 2337*; Author: Lech Wałęsa; Publication date: 16 February 2002; Register: Journalism/Weekly newspaper.

(11) Source: *Dziennik Zachodni*; Publication date: 24 March 2004; Register: Journalism/Daily newspaper.

(12) Source: *Gazeta Poznańska*; Title: "Wzloty i knoty"; Author: KOT; Publication date: 22 November 2003; Register: Journalism/Daily newspaper.

(13) Source: *Początek*; Author: Andrzej Szczypiorski; Publication date: 1986; Register: Literature/Book.

(14) Source: *Finanse publiczne*; Author: Owsiak Stanisław; Publication date: 1998; Register: Scientific-didactic/Book.

(15) Source: *Trybuna*; Publication date: 2006; Register: Journalism/Daily newspaper.

(16) Source: *Kultura nr 11*; Author: K. A. Jeleński; Publication date: 1976; Register: Journalism/Monthly.

(17) Source: *Tygodnik Tucholski nr 752*; Authors: Kamilla Szablewska, Lucyna Zdanowska; Publication date: 2006; Register: Journalism/Weekly newspaper.

(18) Source: *Gazeta Wrocławsk*a; Author: Palacz P. Rafał; Publication date: 8 November 2003; Register: Journalism/Daily newspaper.

(19) Source: *Gazeta Wrocławska*; Author: Palacz P. Rafał; Publication date: 9 November 2003; Register: Journalism/Daily newspaper.

(20) Source: *CKM nr 05/05*; Author: Jerzy Andrzejczak; Publication date: 2000; Register: Journalism/Monthly.

(21) Source: *Lato nieśmiertelnych*; Author: Mariusz Ziomecki; Publication date: 2002; Register: Literature/Book.

(22) Source: *Gazeta Pooznańska*; Title: "Sport Leszno"; Authors: Arkadiusz Kaczmarek, Andrzej Bartkowiak; Publication date: 4 May 2002; Register: Journalism/Daily newspaper.

(23) Source: *Życie Podkarpackie*; Title: "Krok od kompromitacji"; Author: Mariusz Godos; Publication date: 22 November 2006; Register: Journalism/Weekly newspaper.

(24) Source: *Trybuna Śląska*; Title: "Grają i rozśmieszają"; Author: Maciej Kołodziejczyk; Publication date: 27 December 2002; Register: Journalism/Daily newspaper.

(25) Source: *Polityka nr 2514*; Title: "Cień Miłosza"; Author: Jerzy Pilch; Publication date: 30 July 2005; Register: Journalism/Weekly newspaper.

(26) Source: *Polityka nr 249*; Title: "13–16–19"; Publication date: 16 April 2005; Register: Journalism/Weekly newspaper.

(27) Title: "Stenogram z 49. posiedzenia Komisji Śledczej 23 lutego 2005 r."; Author: Kancelaria Sejmu Rzeczypospolitej Polskiej; Publication date: 2005; Register: Quasi-spoken/Stenographic record.

(28) Source: *Sprawozdanie stenograficzne z obrad Sejmu RP*; Title: "Sprawozdanie stenograficzne z obrad Sejmu RP z dnia 07.02.1997, 2 kadencja, 100 posiedzenie, 3 dzień"; Author: Kancelaria Sejmu Rzeczypospolitej Polskiej; Publication date: 1997; Register: Quasi-spoken/Shorthand report.

(29) Source: *Ozon nr 24*; Title: "Szansa na IV RP"; Author: Piotr Semka; Publication date: 2005; Register: Journalism/Weekly newspaper.

(30) Source: *Czy Jezus to powiedział? Zagadki ewangeliczne*; Author: Mieczysław Maliński; Publication date: 2003; Register: Nonfiction/Book.

(31) Source: *Edward Gierek: życie i narodziny legendy*; Author: Janusz Rolicki; Publication date: 2002; Register: Nonfiction/Book.

Jarosław Wiliński
Institute of Linguistics and Literary Studies
University of Siedlce
Siedlce, Poland
jaroslaw.wilinski@uws.edu.pl

# Review

Walter Riggs Thompson. *Epifanii Slavinetskii's Greek-Slavonic-Latin Lexicon between East and West*. Volume 8 of *Empirie und Theorie der Sprachwissenschaft*. Heidelberg: Universitätsverlag, 2024. 412 pp. ISBN 978-38-2539-596-4.

Reviewed by Cynthia M. Vakareliyska

This is a hefty and extremely well-researched doctoral dissertation for the Faculty of Modern Languages at Heidelberg University, published with minimal revision, as is the practice in Germany, and clearly written for the doctoral supervisor and dissertation committee rather than for a reading audience of other specialists in one or more fields. It is no fault of the author, of course, that this *tour-de-force* dissertation looks like a dissertation, but it raises a philosophical question regarding the criteria for writing book reviews of published dissertations. It is hardly the place of a book reviewer to evaluate a dissertation, yet treating a dissertation as a scholarly book in a review requires the pointing out of some features that may be appropriate for a dissertation but that are less than ideal in a peer-reviewed scholarly book.

The goal of the book, as the author states (p. 18), is to examine cultural processes and historical issues concerning the creation, use, and reception of the Ruthenian monk Epifanij Slavineckij's draft copy of his 17th-century *Greek-Slavonic-Latin Lexicon* (*GSL*, in the author's abbreviation). The hypotheses that the book seeks to prove are (a) that Slavineckij (ca. 1600–1675) was not a Graecophile (i.e., a member of the school that stressed the Greek language and Orthodoxy and was opposed to Latin/Roman Catholic influence coming from Poland), as claimed by some scholars, but that he used and translated Latin texts and lexemes, presumably from Western European sources, and that Latin influence is demonstrated by Slavic calques from Latin in the *GSL*, including Greek loanwords transmitted via Latin into Early Modern written European languages ("internationalisms"); (b) that Slavineckij used a variant of Ruthenian Church Slavonic in the draft manuscript, which was later Russified in copies of the finished manuscript produced by Muscovite scribes; and (c) that the dictionary contains features of a developed Church Slavonic "academic," i.e., scholarly, language, particularly in the grammatical terms used.

The book consists of nine chapters, a bibliography, and two appendices. The table of contents is followed by an index of tables and an index of figures (i.e., writing samples) that appear in the text. The individual chapters analyze the *GSL* exhaustively from the perspectives of different disciplines, including history, lexicography, and Slavic linguistics.

Chapter 1 is an introduction focusing on research goals and methods, and the conventions used in the book. Chapter 2 presents a thorough summary of the historical and cultural background of the *GSL* and a detailed review of the scholarly literature, covering lexicography, language attitudes of the early modern East Slavs and of Muscovy during Nikon's reforms, Latin and Greek as languages of communication in Muscovy, and Slavineckij's lexicographic works.

Chapter 3 is a description of the draft *GSL* manuscript at the State Historical Museum in Moscow (Sin. Gr. 383). The chapter provides a detailed codicological and paleographic analysis of the manuscript, comparing it with Titov 67 at the Russian National Library, which is one of the two extant Muscovite clean copies of Slavineckij's now-lost final manuscript and the subject of an article by Olga Strakhov (2006). Chapter 4, "Formal Structural Description," gives an overview of the various types and structures of dictionaries, focusing on issues including the resolution of orthographic and phonetic variation, and the macro-, medio-, and microstructure of dictionaries, based on the model in Wiegand (1984; cited as 1983) and Wiegand et al. (2010). Chapter 5 proposes specific Greek-Latin dictionaries that Slavineckij most likely consulted in the compilation of the *GSL* and looks at his use of primary sources. Thompson compares random alphabetical lemma stretches in the *GSL* to each possible dictionary source proposed by earlier scholars, concluding that Constantin's 1592 Greek-Latin dictionary was the *GSL*'s primary source for the lemmas, with Scapula's 1652 dictionary as a supplemental source rather than as the main source for the *GSL*, as some have argued.

Chapter 6 reverts from the *GSL* generally back to the features of the draft manuscript, giving a grammatical and lexical description of the manuscript with focus on its Ruthenian features in order to determine the Slavic variant employed in it. This chapter should be of the most interest to Slavic linguists. In it, Thompson argues, with support from secondary sources, that Slavineckij was more likely Belarusian than Ukrainian, and he identifies Ruthenianisms, Polonicisms, German loans via Polish, and Latin words via Polish in the draft *GSL* manuscript, noting that Slavineckij and other translators of the Čudov Circle to which he belonged used Polish texts in translating from Greek and Latin (p. 260). The chapter also looks for sources for Old Testament citations in the *GSL*, concluding that the most likely source was the printed Ostrog Bible (1581), as Pentkovskaja (2017) has suggested. The chapter also discusses

New Testament citations and the abbreviations that Slavineckij employed for references.

Chapter 7 is on the Latin influence on the lexicon and loanwords in the *GSL*, including Latin lexemes of Greek origin, and some lexemes which, Thompson argues, Slavineckij calqued or coined himself. This chapter presents Thompson's main argument: that *GSL* was influenced more by Latin than by Greek, indicating that Epifanij Slavineckij was not a member of the Graecophile school as some have argued. Upon conducting a search of nouns in *-ost-ъ* in the draft *GSL* manuscript, Thompson determines that these words do not calque a Greek suffix, as Strakhov (2006) proposed, but that they must be from Polish instead.

Chapter 8 is on what Thompson calls "academic vocabulary," that is, grammatical and poetic/metrical terminology in the *GSL*, as well as terms for academic disciplines. Chapter 9 presents conclusions. This last chapter is followed by Appendix 1, "Tables for Sources Analysis," which compares *GSL* sample lemma stretches to their equivalents in five relatively contemporaneous Greek-Latin dictionaries. Appendix 2 discusses the methodology for transcribing the 744-folio draft manuscript of the *GSL*, using various handwritten text recognition (HTR) models with various degrees of success.

Turning now to the features of the book that mark it as a dissertation, it should be stressed first that the prose in all the chapters is beautifully written and that the book is exhaustively researched. The chapter structure of the book, however, jumps back and forth among lexicography in general, the text of the *GSL* (as witnessed in Slavineckij's draft and the two Muscovite final copies), and the draft *GSL* manuscript itself, often with no indication of a boundary or a switch between the last two: for example, chapter 2 on lexicography analyzes the *GSL* generally rather than the draft manuscript.

This leads to the issue of the book's intended audience, if one is contemplated (in addition to the dissertation committee). The chapters on the place of the *GSL* within modern dictionary taxonomies appear likely to be of most interest to lexicographers, rather than historical Slavic linguists, for whom the chapters analyzing the draft manuscript should be of primary interest. The question of intended audience also arises on the microlevel because no English translation is provided for long quotations in Greek, Latin, and Church Slavonic, but an English translation follows the quotations in modern Russian, German, and French, even in the chapters that analyze the manuscript from a Slavic linguistic perspective (see, for example, p. 248). Are the translations of the quotations in modern languages for the benefit of the dissertation committee, and if so, did the committee members read Greek, Latin, and Church Slavonic well but not read the modern languages? Or, was this a required convention for dissertations?

The audience question also carries over into the linguistics chapters. Hypothetical morphological forms are given in Old Cyrillic font rather than in Roman transliteration, and when Roman transliteration is used, it is not a standard Slavic linguistics transliteration but the English system often used for history and literature publications, which does not have a letter-to-letter equivalency as linguistic transliteration does. For example, in the discussion of the draft *GSL* manuscript's morphological features, Thompson writes, "forms in *sh-* (comparative) and *-eish-* (superlative) […], even though the *neish-* forms were traditionally treated in (Church) Slavonic as the comparatives of adjectives in *-ьn-*" (p. 272). There is also inconsistency in the use of Roman vs. Cyrillic in the narrative prose: for example, why do Russian sample words appear in Cyrillic but the suffixes above are in Roman transliteration? (Cf. fn 44 on the same page, where the suffix *-im-* is provided in Old Cyrillic this time.) A quotation from a miscellany from 1791 by Nikolja Novikov preserves the *jers* and *jat´* (p. 30), but, as an example, a quotation from 18th-century Ioann Maksimovič appears without *jat´* or *jers* (p. 5).

Another issue is the length of the monograph, which would have benefited greatly from reduction. For instance, in chapter 3, the paleographic description of Slavineckij's draft manuscript, Thompson argues that close analysis of the specific manuscript is important because it contains corrections and insertions, and that the exhaustive examination of the paleographic features of Slavineckij's hand in the chapter may be of use in identifying other manuscripts that he had written. It is unclear, however, why it should be a task of this particular book, in addition to its stated goals, to provide a chapter-long paleographic analysis of the *GSL* draft manuscript as "the first complete, systematic documentation of Epifanii Slavinetskii's handwriting, which may prove useful for working with further manuscripts in his hand" (p. 134). Indeed, the chapter is a grapheme-by-grapheme analysis of the Greek, Latin, and Ruthenian Church Slavonic alphabet letters in the draft manuscript, with a separate illustration for each grapheme, in order to demonstrate that some of the Cyrillic graphemes in the manuscript display Latin influence. All the graphemes, however, could have been presented in a single table, with text commentary on only the more noteworthy of them, rather than on every one of them. The chapter also compares orthographic differences with the Muscovite copy Titov 67, even providing a two-page table of minor orthographic differences in specific words and phrases in the Slavic translation of Greek words, and examining the paleography, phonology, morphology, and lexicon of the latter manuscript, which, as one would expect, was adapted to Muscovite norms (pp. 132–33). The chapter then refers the reader to chapter 6, on morphology and lexicon, for the examination of Ruthenian features in Slavineckij's draft manuscript. The interruption of the linguistic analysis of the draft manuscript by two chapters on lexicographical issues contributes to the disjointed structure of the book.

As noted above, chapter 4, the first of the interrupting chapters, presents a long and detailed taxonomic analysis of the structure of the *GSL* generally and its place within the model of the *Wörterbuch zur Lexikographie und Wörterbuchforschung* (Wiegand et al. 2010); this chapter could have been reduced considerably by removing many or all of the diagrams and lexicographic tree structures in it. Chapter 5, which discusses exhaustively every known Greek-Latin dictionary contemporaneous to the *GSL* in order to speculate which ones likely were the sources that Slavineckij used for the *GSL*, could also have been shortened considerably, and then both shortened chapters could have been collapsed into a single chapter, since all of chapter 4 focuses on the reasons why Thompson proposes certain specific manuscripts as the sources for the *GSL* in chapter 5.

As is common in dissertations, the analysis leans heavily on others' models, including, in chapter 3, Rezec's (2009) four- and six-tier graphematics models for describing the modern German writing system (p. 88), and, over much of chapter 4, the taxonomy of modern dictionaries in Wiegand et al. (2010). The lengthy and very detailed review of the scholarly literature in chapter 2, which may well be required of dissertations in Germany in order to prove that the doctoral candidate has a firm grasp of the literature, also is overdone from the perspective of a scholarly book. The use of paraphrases rather than long quotations from secondary sources would have also reduced the impression of a dissertation.

Appendix 1, "Tables for Sources Analysis," compares *GSL* sample lemmas from a single folio to their equivalents in five relatively contemporaneous Greek-Latin dictionary sources, by Leymarie (1583); Constantin (1592); Scapula (1652); Hesychios (1514); and Favorino (1523), to support the conclusion that Slavineckij compiled the *GSL* from multiple dictionary sources and that Constantin's dictionary was likely his main source for the lemmas in the *GSL*. Appendix 2, which tracks the successes and failures of various HTR models in reproducing parts of the draft *GSL* manuscript, is essentially a book chapter: that is, it is a prose discussion containing tables which will probably be of interest mainly to scholars who work with HTR. This excursus would, however, make a nice article if republished in a Slavic journal that specializes in digital issues such as *Scripta & e-Scripta*.

In conclusion, this book covers Slavineckij's draft *GSL* manuscript, and the *GSL* itself in its extant copies, from nearly every conceivable perspective: linguistic, historical/cultural, lexicographic, and HTR. It is a definitive source on the manuscript that may turn out to be of most interest to lexicographers, particularly those who study Greek-Latin dictionaries. It also provides a meticulous paleographic and linguistic analysis of the draft manuscript. From the perspective of a Slavic linguist, however, one is left wondering whether

readers can follow such an exhaustive close study from so many different angles without feeling overwhelmed, regardless of their specialization.

The monograph has been researched, written, and proofread very meticulously; although I was not actively searching for them, I did not notice any typographical errors in any of the languages, which is remarkable for a 412-page book. I have just one minor quibble: page 232 refers to the "Rylsk Monastery" for the Rila Monastery in Bulgaria.

## Manuscript Sources

Sin. Gr. 383. Slavineckij, Epifanii. "Greek-Slavonic-Latin lexicon". State Historical Museum, Moscow, Synodal collection. Copy, ca. 1660's.

Titov 67 and 68. Slavinetskij, Epifanij. Draft of "Greek-Slavonic-Latin lexicon". Russian National Library, St. Petersburg, Titov Collection, nos. 67 and 68, vols. I and II. Late 17th/early 18th century.

## Printed Sources

Constantin, Robert [Robertus Constantinus]. (1592) *Lexicon sive dictionarium graecolatinum.* 2nd ed., expanded by Franciscus Portus. Paris.

Favorino, Guarino [Guarino da Fevera/Favorino]. (1523) *Magnum ac perutile dictionarium*. Rome.

Hesychios [Hesychiius Alexandrinus]. (1514) Λεξικὸν. Venice.

Leymarie, Guillaume [Guillielmus Leimarius]. (1583) *Lexicon Graecolatinum recens constructum*. Geneva.

Novikov, Nikolja. (1791) *Drevnjaja rossijskaja vivliofika, soderžaščaja v sebe sobranie drevnostej rossijskix, do istorii, geografii i genealogii rossijskija kasajuščixsja.* Part 1. Moscow.

Ostrog Bible. (1581) Printed by Ivan Fedorov. Ostroh.

Pentkovskaja, Tat´jana. (2017) "Tolkovanija na poslanija apostola Pavla v perevode Epifanija Slavineckogo kak istočnik Novogo Zaveta poslednej četverti XVII v". *Vestnik Moskovskogo universiteteta* 9. *Filologija* 2: 27–52.

Rezec, Oliver. (2009) *Zur Struktur des deutschen Schriftsystems: Warum das Graphem nicht drei Funktionen gleichzeitig haben kann, warum ein <a> kein <a> ist und andere Konstruktionsfehler des etablierten Beschreibungsmodells. Ein Verbesserungsvorschlag.* Dissertation, Ludwig Maximilian University of Munich.

Scapula, Johannes. (1652/1580) *Lexicon graeco latinum novum*. Amsterdam.

Strakhov, Olga. (2006) "Jepyfanij Slavynec´kyj's Greek-Slavic-Latin lexicon: The history, contents, and principles underlying the composition of its Greek portion (preliminary remarks)". *Harvard Ukrainian studies* 28(1/4): 269–85.

Wiegand, Herbert Ernst. (1984) "On the structure and contents of a general theory of lexicography". Reinhard R. K. Hartmann, ed. *LEXeter '83: Proceedings*. Tübingen, Germany: Max Niemeyer, 13–30.

Wiegand, Herbert Ernst, et al., eds. (2010) *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of lexicography and dictionary research*. Vol. 1 of *Systematische Einführung/Systematic introduction, A–C*. Berlin/New York: De Gruyter.

Cynthia M. Vakareliyska
Uppsala, Sweden
and
Department of Linguistics
University of Oregon
Eugene, OR, USA
vakarel@uoregon.edu