JAZYKOVEDNÝ ČASOPIS

JAZYKOVEDNÝ ÚSTAV ĽUDOVÍTA ŠTÚRA

SLOVENSKEJ AKADÉMIE VIED

3

ROČNÍK 75, 2024

sciendo

SAP
SLOVAK ACADEMIC PRESS

JAZYKOVEDNÝ ČASOPIS
VEDECKÝ ČASOPIS PRE OTÁZKY TEÓRIE JAZYKA

JOURNAL OF LINGUISTICS
SCIENTIFIC JOURNAL FOR THE THEORY OF LANGUAGE

Vydáva/Published by: Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied, v. v. i.
- v tlačenej podobe vo vydavateľstve SAP – Slovak Academic Press, s. r. o.
- elektronicky vo vydavateľstve Sciendo – De Gruyter
  https://content.sciendo.com/view/journals/jazcas/jazcas-overview.xml

Adresa redakcie/Editorial address: Jazykovedný ústav Ľ. Štúra SAV, Panská 26, 811 01 Bratislava
Kontakt: jazykovedny.casopis@juls.savba.sk
Elektronická verzia časopisu je dostupná na internetovej adrese/The electronic version of the journal is available at: https://www.juls.savba.sk/ediela/jc/, https://www.sav.sk/?lang=sk&doc=journal-list&journal_no=26

Vychádza trikrát ročne/Published triannually

Dátum vydania aktuálneho čísla (2024/75/3) – február 2025

Quartile ranking 2023: Q2
CiteScore 2023: 0,5
SCImago Journal Rank (SJR) 2023: 0,258
Source Normalized Impact per Paper (SNIP) 2023: 0,583

JAZYKOVEDNÝ ČASOPIS je evidovaný v databázach/JOURNAL OF LINGUISTICS is covered by the following services: Baidu Scholar; Cabell's Journalytics; CEEAS (Central & Eastern European Academic Source); CEJSH (The Central European Journal of Social Sciences and Humanities); CEEOL (Central and Eastern European Online Library); CNKI Scholar (China National Knowledge Infrastructure); Dimensions; EBSCO; ERIH PLUS (European Reference Index for the Humanities and Social Sciences); ExLibris; Google Scholar; IBR (International Bibliography of Reviews of Scholarly Literature in the Humanities and Social Sciences); IBZ (International Bibliography of Periodical Literature in the Humanities and Social Sciences); International Medieval Bibliography; J-Gate; JournalGuide; JournalTOCs; KESLI-NDSL (Korean National Discovery for Science Leaders); Linguistic Bibliography; Linguistics Abstracts Online; MLA International Bibliography; MyScienceWork; Naver Academic; Naviga (Softweco); ProQuest; ReadCube; ScienceON/AccessON; SCILIT; SCImago (SJR); Scite_; SCOPUS; Semantic Scholar; Sherpa/RoMEO; TDNet; Ulrich's Periodicals Directory/ulrichsweb; WorldCat (OCLC); X-MOL.

ISSN 0021-5597 (tlačená verzia/print) **ISSN 1338-4287 (verzia online)** **MIČ 49263**

JAZYKOVEDNÝ ČASOPIS

JAZYKOVEDNÝ ÚSTAV ĽUDOVÍTA ŠTÚRA
SLOVENSKEJ AKADÉMIE VIED

3

ROČNÍK 75, 2024

sciendo

SAP
SLOVAK ACADEMIC PRESS

OBSAH
CONTENTS

265 **Gabriela MÚCSKOVÁ: Slavic Lingustics Society a jej 18. konferencia v Bratislave**
Gabriela MÚCSKOVÁ: Slavic Linguistics Society and its 18th conference in Bratislava

Štúdie
Studies

269 **Laura A. JANDA: What goes around, comes around: Cases that keep me going**
Laura A. JANDA: Všetko sa raz vráti: pády, ktoré ma poháňajú

291 **Martina IVANOVÁ: Clitics in Slovak with(out) constraints. Corpus frequency versus acceptability ratings**
Martina IVANOVÁ: Klitiky v slovenčine s obmedzeniami aj bez nich. Korpusová frekvencia verzus hodnotenie prijateľnosti

309 **Stefan Michael NEWERKLA: Multilingualism in the Habsburg monarchy and Slavic linguistic heritage in German in Austria**
Stefan Michael NEWERKLA: Viacjazyčnosť v habsburskej monarchii a slovanské jazykové dedičstvo v rakúskej nemčine

330 **Mark Richard LAUERSDORF: In search of patterns of historical language variation and user interaction (or: Who used what linguistic features with whom, when, where, why, and how?)**
Mark Richard LAUERSDORF: Hľadanie zákonitostí historickej jazykovej variantnosti a používateľskej interakcie (alebo: Kto používal aké jazykové prvky s kým, kedy, kde, prečo a ako?)

sciendo

SLAVIC LINGUSTICS SOCIETY A JEJ 18. KONFERENCIA V BRATISLAVE

Linguista sum; linguistici nihil a me alienum puto
(Roman Jakobson)

Výrok Romana Jakobsona parafrázujúci známy citát z komédie Publia Terentia Afera vystihuje ducha dnes už dvadsaťročnej medzinárodnej vedeckej spoločnosti *Slavic Linguistics Society* (SLS), ktorá ho ako svoje motto uvádza aj na svojej webovej stránke (https://slaviclinguistics.org/about). Spoločnosť podporuje medzinárodnú komunitu vedcov a študentov venujúcich sa slovanským jazykom a vo všetkých svojich aktivitách sa usiluje byť čo najotvorenejšia k rôznym školám, výskumným oblastiam, prístupom a teóriám – a teda neodmieta nič, čo je lingvistické.

Spoločnosť vznikla v roku 2004 na podnet účastníkov výročného stretnutia American Association of Teachers of Slavic and East European Languages (AATSEEL) vyplývajúci z už dlhšie prebiehajúcich diskusií o potrebe širšie zameranej vedeckej platformy so zameraním na slovanskú lingvistiku. Tieto diskusie sa spájajú aj s konferenciou *Formal Approaches to Slavic Linguistics* (FASL), ktorá sa od roku 1992 pravidelne koná na rôznych univerzitách v USA a v Kanade. Konferencia FASL sa zameriava najmä na otázky formálneho opisu slovanských jazykov a bola preto vnímaná ako tematicky vyhranená. Z potreby vytvoriť platformu so širším tematickým záberom, ktorá by zahŕňala viaceré aspekty výskumu slovanských jazykov a zároveň ich prepojila s oblasťami všeobecnej lingvistiky, vznikol už v roku 1993 časopis *Journal of Slavic Linguistics* (vychádza vo vydavateľstve Slavica Publishers v Bloomingtone), ktorý sa od roku 2006 stal a dodnes je oficiálnym vedeckým časopisom spoločnosti SLS. Jeho cieľom je „slúžiť celej populácii slovanských jazykovedcov bez ohľadu na teoretickú orientáciu alebo tému výskumu“ (Franks 2015, s. 189).[1]

Za vznikom časopisu, ale aj spoločnosti SLS nie sú, samozrejme, asociácie a konferencie, ale konkrétni ľudia. Za mnohých je potrebné spomenúť meno profesora Stevena Franksa z Univerzity v Bloomingtone, ktorý bol autorom myšlienky vytvorenia časopisu (a od roku 1997 aj jeho hlavným redaktorom), ako aj založenia SLS ako spoločnosti lingvistov z rôznych krajín a vedeckých slavistických inštitúcií.

Dnes je *Slavic Linguistics Society* širokým medzinárodným spoločenstvom združujúcim takmer 350 slavistov z rôznych krajín Európy, Ameriky a Ázie a jej

[1] Franks, Steven (2015): The Slavic Linguistics Society Comes of Age. In: *Journal of Slavic Linguistics*, roč. 23, č. 2, s. 189 – 196.

aktivity sú zamerané na široký lingvistický výskum a propagáciu slovanských jazykov. Spája akademické pracoviská v Európe, Severnej Amerike a Ázii, a to najmä prostredníctvom svojich každoročných konferencií, ktoré sa konajú od roku 2006 v rôznych krajinách na pôde vybraných univerzitných alebo akademických pracovísk, ktoré sa na príprave konferencie organizačne podieľajú. V minulosti sa konferencie konali v Severnej Amerike (Bloomington, Columbus, Chicago, Lawrence, Seattle, Toronto, Eugene, Champaign, Provo), v Európe (Berlín, Zadar, Aix-en-Provence, Szczecin, Heidelberg, Ľubľana, Postupim, Bratislava) a raz aj v Ázii, v japonskom Sappore. V minulom roku spoločnosť oslávila 20 rokov svojej existencie a v roku 2025 sa bude v talianskej Verone konať už jej 20. výročná konferencia (https://sites.google.com/view/sls20verona/home).

Týmto špeciálnym číslom *Jazykovedného časopisu* si však chceme pripomenúť 18. ročník konferencie tejto spoločnosti, ktorý sa konal v dňoch 24. – 26. augusta 2023 v Bratislave na pôde Filozofickej fakulty UK v Bratislave a ktorej hlavným organizátorom bol Jazykovedný ústav Ľ. Štúra SAV, v. v. i., a spoluorganizátorom Studia Academica Slovaca – centrum pre slovenčinu ako cudzí jazyk Filozofickej fakulty Univerzity Komenského v Bratislave.

Na tejto bratislavskej konferencii odzneli štyri plenárne prednášky významných pozvaných hostí: profesorka **Laura Alexis Janda** z University v Tromsø, profesorka **Martina Ivanová** z Prešovskej univerzity, profesor **Stefan Michael Newerkla** z Viedenskej univerzity a profesor **Mark Richard Lauersdorf** z Univerzity v Kentucky. Práve ich plenárne príspevky prinášame v tomto špeciálnom čísle *Jazykovedného časopisu* v rozšírenej podobe.

Okrem nich však na konferencii v priebehu troch dní a v troch sekciách odznelo viac ako 80 príspevkov z rôznych lingvistických oblastí a metodologických zameraní, medzi ktorými dominovali oblasti gramatiky (so zameraním najmä na aspekt a kategóriu vidu alebo modálnosť, ale aj iné kategórie) a syntaxe (v prepojení s otázkami slovosledu a pozície enklitík), svoje zastúpenie mali aj oblasti lexiky a jej slovníkového spracovania, ako aj fonetiky a fonológie. Takmer všetky štruktúrne orientované príspevky prinášali nové poznatky a výsledky výskumov jazykovej dynamiky a variability reflektovanej v komunikácii a reálnych rečových interakciách. Okrem toho boli bohato zastúpené rôzne témy z oblasti sociolingvistiky, sociálnej lingvistiky a pragmalingvistiky, ďalej jazykovej politiky a jazykového plánovania, ako aj jazykových ideológií. Nechýbalo ani zastúpenie diachrónnych pohľadov do minulosti a vývinu slovanských jazykov a praslovančiny, a širšie koncipované jazykovokulturologické a etnolingvistické témy. V sumarizácii bohatého tematického spektra nemožno obísť predstavenie nových technológií a ich využitie vo výskume a vyučovaní slovanských jazykov. V samostatných sekciách sa konali dve panelové diskusie tematicky zamerané na 1) ukrajinský jazyk a jazykovú situáciu na Ukrajine (s názvom *The Ukrainian language today: Issues of history, sociolinguistics, and*

areal typology), ktorej predsedal Andriy Danylenko (Pace University v New Yourku, USA) a 2) binominálne konštrukcie substantív v slovanských jazykoch (*Actional features of nouns: The case of binominal constructions with classifierlike nouns in Slavic languages and beyond*) pod vedením Marca Biasia (Univerzita v Modene, Taliansko) a Alessie Lacroce (Tre Univerzita a Univerzita Sapienza v Ríme, Taliansko). Program konferencie a zborník abstraktov[2] sú stále prístupné v archíve Jazykovedného ústavu Ľ. Štúra SAV (https://www.juls.savba.sk/attachments/sls_2023/PROGRAM_2-1.pdf) a na webovej stránke spoločnosti SLS.

Podujatie navštívilo viac ako 150 domácich i zahraničných lingvistov z 25 krajín vrátane USA, Číny, Japonska, Južnej Kórey, Izraela a Nového Zélandu. Za jeho úspešným priebehom stojí úsilie organizačného tímu Jazykovedného ústavu Ľ. Štúra SAV, v. v. i., spoluorganizátora podujatia Studia Academica Slovaca – centra pre slovenčinu ako cudzí jazyk FF UK, ako aj výboru spoločnosti Slavic Linguistics Society. Špeciálne číslo časopisu je tak vyjadrením vďaky všetkým organizátorom za organizačnú prípravu a finančnú podporu, ale aj všetkým účastníkom konferencie, ktorí sa prišli podeliť o výsledky svojich výskumov. Osobitne sa chceme poďakovať plenárnym prednášateľom a prednášateľkám za ochotu poskytnúť ich referáty v podobe vedeckých štúdií pre náš časopis. Dúfame tak, že toto číslo bude nielen pripomienkou 18. konferencie SLS, ale aj motivujúcim podnetom k účasti na jej budúcich konferenciách.

Gabriela Múcsková

[2] *Book of Abstracts: 18th Annual Meeting of the Slavic Linguistics Society (SLS-18)*. Ed. M. Zumrík. Bratislava: SAP – Slovak Academic Press, 2023.

sciendo
DOI 10.2478/jazcas-2024-0044

WHAT GOES AROUND, COMES AROUND: CASES THAT KEEP ME GOING

LAURA A. JANDA
UiT The Arctic University of Norway, Tromsø, Norway

JANDA, Laura A.: What goes around, comes around: Cases that keep me going. Jazykovedný časopis (Journal of Linguistics), 2024, Vol. 75, No. 3, pp. 269 – 290.

Abstract: I am fascinated by the treasure trove of meanings tucked away in the grammatical morphemes that many people think of as mere functional fillers. As a student, the Slavic case endings baffled me, then later delighted me with their complex stories about trajectories, time, benefit and harm, labels, and so much more. Some twenty years ago I was satisfied that I had cracked that code, and after writing some articles and a couple of textbooks on the topic, I moved on. But the cases came back to me again and again. In this article, I tell the story of how my work on case semantics later helped to inspire three further projects: two major online resources, the Strategic Mastery of Russian Tool and the Russian Constructicon, and an analysis of president Putin's portrayal of Russia, Ukraine, and NATO. At first glance it might seem that this line of research is rather shallow and merely descriptive, however digging into case semantics reveals some deep philosophical issues concerning the relationship of meaning to grammar, the assumptions inherent in linguistic reference works, the representation of paradigms in the minds of speakers, and the ways in which we can measure grammatical norms and deviation.

Keywords: Slavic languages, case, semantics, cognitive linguistics, construction grammar, political discourse.

1. INTRODUCTION

Theoretical linguistics, applied linguistics, and language pedagogy are sometimes kept separate from each other, on the assumption that they are very different pursuits. Here I present case semantics as a red thread that has led through a series of projects I have undertaken that link these three disciplines to each other. Section 2 presents the meanings of grammatical cases, using Cognitive Linguistics as the theoretical framework and the Russian case system as the material basis. In Section 3 I turn to the distribution of grammatical case in corpus data and a pedagogical resource created to use this data to provide strategic input for language learners. Case never occurs in a vacuum, always hosted by words and embedded in constructions, but the majority of grammatical constructions cannot be deduced from traditional reference works, a fact that motivated the building of the Russian Constructicon, which is the topic of Section 4. In Section 5 I venture into analysis of political discourse through the distribution and meanings of grammatical case. Conclusions are offered in Section 6.

2. CASE SEMANTICS: LINGUISTIC THEORY AND DESCRIPTION

After studying Russian for three years in the 1970s, I realized I had a problem. I knew a lot of words, and I could parse just about any sentence, but I was still often stumped about what a given sentence meant. A big part of my problem was the meanings of the Russian (and mutatis mutandis Slavic) cases. As a student I was perplexed by the seemingly random long lists of prepositions and verbs I was assigned to memorize for each case. It was clear to me already then that the grammars I was reading couldn't be telling the whole story. Later, when I had a steady job, I tackled what I assumed were the hardest cases, the Dative and Instrumental (Janda 1993). Little did I suspect that the Genitive (Janda 1999) and Accusative (Janda 2000) cases would offer plenty of challenges as well. Even the Nominative and Locative were not trivial, and they rounded out the set for two textbooks that I co-authored (Janda – Clancy 2002, 2006).

2.1 Theoretical issues

On one level I was doggedly picking apart the nitty-gritty details of grammatical case, considered by some linguists to be a syntactic phenomenon devoid of meaning. On another level I was confronting some basic philosophical tenets of linguistics, namely the role of meaning in grammar, and my appreciation for the form-meaning relationship continued to grow. It is common to think of a language as consisting of a lexicon – a set of words that contain the meanings, and a grammar that shows how the words are combined. From the perspective of Cognitive Linguistics, the lexicon and the grammar are not separate entities, but parts of a single system, or as Langacker (2008, p. 15) describes it: "lexicon, morphology, and syntax form a continuum". In this system, all units have both form and meaning, although the meanings of syntactic expressions tend to be relatively more schematic and polysemous than those of lexemes. More specifically with regard to my research agenda, Langacker (2008, p. 95) states that the "basic grammatical classes are semantically definable".

While on the face of it the claim that grammatical categories invoke meanings might seem surprising to some, there are several types of evidence that support grammatical meaning: a) typological variation in how functions are expressed, and b) the internal structure of cognitive categories shared across lexicon and grammatical categories.

Many functions are expressed grammatically in some languages, but lexically in others, and often the very same function can be expressed both ways even in the same language and even simultaneously in a single utterance. Here are just a few examples of how synthetic grammar and analytic use of lexemes compete in the same semantic domains. The functions of the grammatical cases we find in Slavic languages can be expressed by means of adpositions in languages without

grammatical case. For example, many uses of the Slavic Genitive case can be rendered with the English preposition *of*, as in Russian *načalo fil'ma* 'the beginning of the movie', and many uses of the Slavic Dative case can be rendered as English *to*, as in *dat' graždanam nadeždu* 'give hope to the citizens'. In English *we travel by car*, but in Czech *jedeme autem* using the Instrumental case, sometimes redundantly augmented by a preposition: *jedeme s autem*. And throughout the Slavic languages the meaning of the Locative case is supplemented by prepositions. Some might object that adpositions "don't count" as lexemes because they are merely "function words", but it is not hard to find examples where we need undisputed lexemes to translate the meaning of a grammatical case, as for example Russian *u nego kepka blinom* 'he has a hat that looks like a pancake', where the Instrumental case points to what the hat looks like. This blurring of how functions are expressed synthetically vs. analytically is by no means limited to the meanings of cases. Definiteness expressed by the English article *the* can be expressed by suffixes in Bulgarian, as in *kniga-ta* 'the book', and by either an article or a suffix or even both at once in Norwegian *denne bok-en* 'the/that book'. Verbal categories of tense, aspect, and mood likewise admit both grammatical and lexical expression. It seems that about half of the languages of the world lack an inflectional future (cf. Dryer – Haspelmath 2013 – WALS Feature 67A), and about 40% lack an inflectional past tense (cf. ibid. – WALS Feature 66A). While in some languages these roles are taken on by auxiliary verbs and one could debate whether auxiliary verbs are mere "function words" or full lexemes, in some languages you need an adverb to express tense, as in North Sámi, where *ihttin* 'tomorrow' or some other temporal expression is needed to specify future. Languages like Slavic that express aspect grammatically are in the minority in the world (cf. ibid. – WALS Feature 65A); most languages resort to adding in lexemes or whole phrases with meanings like 'finished', 'completely', 'continually', 'was in the habit of' when there is a need to make aspectual meaning clear. Even adjectives are not exempt from such variation, for we find that comparative and superlative meanings can be produced both by affixes and by lexemes; compare synthetic Persian *zibâ-tar* [beautiful-COMPARATIVE], *zibâ-tar-in* [beautiful-COMPARATIVE-SUPERLATIVE] with analytic English equivalents 'more beautiful', 'most beautiful'. Virtually every grammatical category reveals similar examples where the same function can be expressed either synthetically with grammatical morphemes or analytically with lexemes. In other words, there seems to be no clear boundary separating grammatical from lexical meaning in terms of form.

There is likewise no clear boundary between grammar and lexicon in terms of the internal structure of meaning categories. If the main purpose of language is to convey meaning, perhaps it is not surprising that grammar and lexicon jointly participate in this task. And if we cannot definitively distinguish grammatical meaning from lexical meaning, then perhaps the next question, is: how does meaning work? Here I lean upon scholarship reaching back to Eleanor Rosch (1973a, 1973b).

Meaning is not "out there" in the world, but is rather a cognitive construct created by human beings based on their perception of realia. Through her research on categorization, Rosch discovered that human beings do not operate in terms of Aristotelian categories defined by sets and boundaries, but by what she termed "radial categories". Radial categories are structured around a central prototype (or cluster of prototypes) with extensions radiating from that prototype. Rosch famously showed that English speakers have a radial category for 'bird': prototypical birds are small, feathered and fly, like robins and sparrows, whereas chickens (with limited flying ability and used as food) are les prototypical, while ostriches and penguins are peripheral. Likewise, apples are a prototypical fruit, while lemons are less so, and avocadoes are quite peripheral. While grammatical meanings are typically more schematic, they can also involve a polysemous radial category structured around a prototype. Janda et al. (2013) present the meanings of the prefixes that signal Perfective aspect in Russian, many of which display an internal radial structure. For example, the prefix *raz-* has a prototypical meaning APART manifest especially when used with motion verbs, as in *razojtis'* 'disperse, walk away in different directions'. This meaning can be extended to apply specifically to the edges of a two-dimensional item, with SPREAD as the meaning in *raskatat'* 'roll out (dough)', or a three-dimensional item, with SWELL as the meaning in *razdut'* 'inflate'. Further metaphorical extension yields the meaning EXCITEMENT, as in *razgorjačit'sja* 'heat up, irritate'. The examples presented in Section 2.2 aim to reveal the structures of the case meanings of the Russian cases.

Close examination of case meanings confirms the tenet of Cognitive Linguistics that grammar and lexicon are not distinctly separate, but constitute a continuum, all parts of which contribute to the mission of conveying meaning. Although grammatical meaning may be more abstract and schematic than lexical meaning, meaning at all points along the continuum is a cognitive construct in which prototypical meanings motivate extensions to more peripheral ones.

2.2 Overview of the case meanings, with Russian as an example

Because the details are important to support the theoretical points made above and to motivate the projects described below in Sections 3–5, I will walk through the meanings of all six of the Russian grammatical cases. In the heading introducing each case, I will identify a schematic meaning that summarizes the abstract overall idea expressed by the case and then briefly present a network of between one and four meaning nodes, each cited in small caps, that form the core of the case's meaning. I will point out how the meanings are linked to each other in a relationship of structured polysemy, and I will also give some indications of the further metaphorical and metonymic extensions of these meanings. This is a very condensed version of the contents of this line of research; for a fuller exposition of these meanings, see Janda 1993, 1999, 2000, and Janda – Clancy 2002.

2.2.1 Nominative: Identification

The Nominative case has two central meanings. NOMINATIVE: A NAME can point at an item, be used to call someone, or serve as the grammatical subject. NOMINATIVE: AN IDENTITY is associated with verbs meaning 'be' in formulations meaning 'X is Y' (as in *Ivan* ***xorošij student*** 'Ivan is **a good student**').

2.2.2 Genitive: Backgrounding with respect to a proximate item

The Genitive case establishes the relationship of a focused entity (a trajector) to something that is backgrounded (a landmark marked with the Genitive). GENITIVE: A SOURCE references a point of departure further specified by prepositions meaning 'from' (*iz*, *s*, *ot*, plus *iz-za* 'from beyond', metaphorically extended to mean 'because of' and *iz-pod* 'from beneath') as well as verbs expressing withdrawal (like *izbegat'* 'avoid', *bojat'sja* 'be afraid of'). This meaning is extended metaphorically to other domains such as time (*s* ***detstva*** 'since **childhood'**), cause (*smert' ot* ***razryva serdca*** 'death due to **heart attack**'), and human relationships (*iz* ***milosti*** 'out of **charity**'). GENITIVE: A GOAL references the opposite maneuver, further specified by prepositions (like *do* 'up to, until', *dlja* and *radi* 'for', *protiv* 'against') and verbs and adjectives expressing (mostly metaphorical) approach (like *ždat'* 'wait for', *želat'* 'wish'). This meaning is extended metaphorically to other domains such as time (*do* ***svidanija*** 'until **we meet again**') and purpose (*dlja* ***rešenija*** 'in order **to solve**'). GENITIVE: A WHOLE references the existence of something as a part of a larger unit or collection. This meaning motivates uses of the Genitive case that translate as 'of' and expressions of possession (*ošibka* ***prezidenta*** '**the president's** mistake') or color (*galstuk* ***belogo cveta*** 'a **white** tie') in English. This meaning is also associated with complex prepositional phrases (*v kačestve* ***polnopravnyx učastnikov*** 'in the capacity of **full-fledged participants**') as well as quantification by numerals (*sto* ***studentov*** 'one hundred **students**'), and in partitive expressions (*vypit'* ***čaju*** 'drink **some tea**'). GENITIVE: A REFERENCE locates an item with respect to a landmark in domains of space (like *u* 'by, at'), time (like calendar dates, as in **četvertogo** ijulja '**the fourth** of July'), comparison (*god budet lučše* ***predyduščego*** 'this year will be better than **the previous one**'), and absence (*bez* 'without').

2.2.3 Dative: Interaction

The Dative case encodes the capacity of an entity to interact with its surroundings, by receiving objects, absorbing experiences, or exerting equal or superior strength. DATIVE: A RECEIVER is used primarily to mark the indirect object (*učitel' podaril* ***studentu*** *knigu* 'the teacher gave **the student** a book'), including with verbs of communication (*otvetit'* ***komu-to*** 'answer **someone**') and payment (*zaplatit'* **komu-to** 'pay **someone**'). DATIVE: AN EXPERIENCER is associated with words denoting harm (*mešat'* 'hinder, annoy'), benefit (*služit'* 'serve'), belonging to (*prinadležat'* 'belong to'), and needing (*trebovat'sj*a 'be necessary to'). DATIVE:

A COMPETITOR expresses the capacity of the Dative entity as compared to another entity that is either equal (*protivostojat'* 'withstand') or lesser in strength or influence (*poddavat'sja* 'give in to'), and is associated with the prepositions *k* 'toward' and *po* 'along'.

2.2.4 Accusative: Direction

The Accusative case signals a path toward a destination, or merely the endpoint of that path. ACCUSATIVE: A DESTINATION marks a direct object (*učitel' kupil* ***knigu*** 'the teacher bought **a book**'), which is a metaphorical version of the destination meaning, and is associated with metaphorical extensions to domains such as time (*v* ***ponedel'nik*** 'on **Monday**'), purpose (*otvet na* ***ego vopros*** 'the answer to **his question**'), change of state (*inogda ljubov' perexodit v* ***nenavist'*** 'sometimes love turns into **hatred**'), and mathematics (*v* ***četyre*** *raza* '**quadr**upled'). In the spatial domain, the path referenced by the Accusative case is further specified by prepositions such as *v* 'into', *na* 'onto', *za* 'beyond', *pod* 'under'. ACCUSATIVE: A DIMENSION measures a distance or size in the domain of space (*rasstojanie v* ***dva*** *kilometra* 'a distance of **two** kilometers'), or a duration in the domain of time (*interval v* ***dve*** *nedeli* 'an interval of **two** weeks'). ACCUSATIVE: AN ENDPOINT is primarily associated with the domains of space and time as specified by both prepositions (such as *v* and *za*, both indicating the end of a distance or duration, as in *za* ***odnu nedelju*** 'in/by the end of **a week**') and postpositions (such as *nazad* 'ago').

2.2.5 Instrumental: Accessory

The Instrumental case expresses "an accessory for something else" (Janda – Clancy 2002, p. 19). INSTRUMENTAL: A MEANS expresses a conduit for an action, such as a path that facilitates motion (as in *idti* ***lesom*** 'go **through/by means of the forest**') or an instrument that makes an action possible (as in *rezat' xleb* ***nožom*** 'slice bread **with a knife**'). This meaning is metonymically extended to include use with verbs signifying control (*zavedovat'* 'manage') and evaluation (*vostorgat'sja* 'be delighted with'), and to the agent in a passive construction (*kniga pročitana* ***studentom*** 'the book read **by the student**'). INSTRUMENTAL: A LABEL is used with verbs denoting being, becoming, and seeming, as in *koška javljaetsja* ***mlekopitajuščim*** 'a cat is **a mammal**'. INSTRUMENTAL: AN ADJUNCT occurs with the preposition *s* 'with' and expresses companionship. INSTRUMENTAL: A LANDMARK signifies peripheral locations without contact with the prepositions *nad* 'above', *pod* 'under', *pered* 'in front of', *za* 'behind', and *meždu* 'between'.

2.2.6 Locative: Location

The Locative case has only one meaning, LOCATIVE: A PLACE, which identifies locations in space or other domains, such as time (*v* ***ètom godu*** '**this year**', *pri* ***kommunizme*** 'during **the time of communism**') and states of being (*v* ***vostorge*** 'in

ecstasy'). The meaning of the Locative case is always further specified by prepositions *v* 'in', *na* 'on', *pri* 'at', *o* 'about', *po* 'after'.

2.3 A coherent account of case semantics

While many of the details of case meanings listed above may seem trivial, their consolidation into a coherent system serves both theoretical and pedagogical purposes. This analysis brings a mass of disparate details together in a clear and elegant model. The model furthermore neatly predicts the use of case with novel vocabulary. For example, the borrowed verb *dirižirovat'* 'conduct (a musical group)' governs the Instrumental case, following the model of a group of native Russian verbs meaning 'manage, govern, lead' such as *rukovodit'* that govern the Instrumental. Similarly, the borrowed adjective *izomorfnyj* 'isomorphic' governs the Dative case, following the model of native Russian adjectives like *ravnyj* 'equal' that govern the Dative. And this model can be directly implemented in the classroom, for it is much more tractable than a long and scattered list of seemingly unmotivated contexts for one case or another that must be memorized. The model gives students a meaningful scaffold on which to build their understanding of grammar.

Just knowing the meanings of the cases, however, is not enough for a student to gain a secure grasp of Russian grammatical case. One also has to connect the cases to the morphemes that express them, as well as to the specific words and contexts in which the cases typically appear. Together with able teams of colleagues I have had the opportunity to build two resources to address these needs: The Strategic Mastery of Russian Tool (SMARTool) and the Russian Constructicon.

3. THE STRATEGIC MASTERY OF RUSSIAN TOOL (SMARTOOL)

For decades I made beginning Russian students rehearse inflectional paradigms. I would write out the paradigm for a word on the board and have the students call out one form after another, then I would erase a couple of the forms, and make the students call out the forms again, and I would repeat this until the students were calling out the entire paradigm from memory in front of a blank chalkboard. I assumed that memorization of paradigms was necessary to equip students with inflectional forms in a way that mimicked the capacity of native speakers. Surely, I reasoned, all native speakers have somewhere in their internal grammars the entire paradigms of all words. But once large digital corpora started becoming available in the early 2000s, I began to suspect that I might not be right. Later, an experiment (Janda – Tyers 2018) proved me wrong. The results of this experiment inspired the creation of the Strategic Mastery of Russian Tool, called the "SMARTool" for short.

3.1 The Distribution and Learnability of Inflected Forms

A striking characteristic of all corpus data is the skewed distribution of items. The frequencies of words follow Zipf's Law (Zipf 1949), a power law according to

which the second most frequent word is only one-half as frequent as the most frequent word, the third most frequent word is one-third as frequent, and so on, with a long tail of words that appear only once. The latter are known as hapaxes, which constitute one half of the total unique lexemes in a corpus. The very existence of so many hapaxes undermines the notion of paradigms: these words by definition cannot be represented in all their forms. And it is not just hapaxes that call the existence of paradigms into question: Zipf's Law applies also to inflected forms, meaning that even high frequency words have skewed distributions of forms within their paradigms. This fact has important implications for the understanding of paradigms, and indeed for the question of whether paradigms are a cognitive reality or just a convenience constructed by linguists and language pedagogues.

In an inflected language like Russian, nouns, verbs, and adjectives all have large numbers of inflected forms. Even a small vocabulary of a few thousand words represents over 100,000 potential forms. But the vast majority of those forms are rarely, if ever used, so one wonders whether we can assume that they are all in the heads of native speakers either.

The largest available corpora of Russian already exceed the volume of the lifetime exposure of a native speaker to their language. If we use a corpus as a proxy for such exposure, we can measure the skew in the distribution of inflected forms. In other words, we can estimate the frequency of various paradigm forms in the input that a native speaker would encounter. However, we don't need to measure from the largest corpora because Zipf's Law scales up: the proportions are stable even as corpus size grows. And this is fortunate because it means that we can use smaller "gold standard" corpora annotated for disambiguation of syncretic forms that yield reliable data on inflection.

When we examine corpus data, we find that even among high frequency words only about 10% of inflected forms are encountered frequently; the remainder are absent or rare. The percentage of lexemes in a word class that are attested in all paradigm forms depends upon the size of the paradigm, and this number decreases dramatically as the size of the paradigm increases. For the small paradigm of English nouns with only two forms – Singular and Plural – only 24% of nouns are found in both forms in a corpus. Norwegian marks both number and definiteness on nouns, meaning that there are four forms in the paradigm, but we find only 3% of nouns in all paradigm forms in a corpus. With a bigger paradigm like that of Estonian nouns with 28 forms, the number of nouns attested in all forms in a corpus is vanishingly small, approaching zero. Russian has a moderate-sized noun paradigm of twelve forms if we combine the second Locative (as in *v snegu* 'in the snow') with the Locative, the second Genitive (as in *čaju* 'some tea') with the Genitive, and the second Accusative (*pojti v soldaty* 'join the ranks of soldiers') with the Accusative and leave aside the "new" vocative (*Svet!* 'Sveta!'). Only 0.06% of nouns appear in the full set of paradigm forms in a Russian corpus (see more on this research in Janda – Tyers 2018).

In light of this distribution, it is reasonable to ask: how can Russian inflection be learned? Francis Tyers and I ran a machine-learning experiment that tested two possible answers to this question: learning by exposure to full paradigms vs. learning by exposure to only the lemma and the single most frequent inflected form of each word. Our experiment is explained in full detail in Janda and Tyers (2018), so I offer only an abbreviated description here. We ranked nouns, verbs, and adjectives according to their frequency in a corpus, and took the 5400 most frequent lexemes (this was the ceiling set by a threshold for frequency and available data), dividing them into groups of 100, starting from the highest frequency items. Aside from the fact that the full paradigms model got to see the whole paradigm of each word, whereas the single forms model saw only the most frequent form, the experiment was the same for both tests. First the two models were trained on the top 100 words, then each model was given just the lemmas of the second 100 (unseen) words as a test. The test was to produce a specific inflected form (actually the most frequent form for that lemma) given only the lemma for each of the 100 previously unseen words. The machine's guesses were recorded and scored for accuracy. Then the second 100 words were added to the training data and the third 100 words were used to test both models. And then the third 100 words were added to the training data and the fourth 100 words were used to test both models. This procedure was iterated until we ran out of data at the 54th trial. The results were remarkable. Whereas both models performed poorly in the first few iterations, by the time they reached the sixteenth iteration, the single forms model surpassed the full paradigms model, which it consistently outperformed both in terms of overall accuracy and in terms of the egregiousness of errors (measured as Levenshtein distance between an error and the correct form).

In sum, the machine found it easier to master Russian inflection when learning only the most frequent word forms than when learning entire paradigms. The single forms model made fewer errors and the errors it did make were not as bad. This finding is consistent with a usage-based cognitively plausible model of morphological inflection. Given this outcome, it was clear to me that I needed to make a radical change in the way I taught inflection. If learning inflection by means of entire paradigms was too hard for a computer and entire paradigms are not reflected in corpus data, I shouldn't be asking my students to learn that way. Corpus data would play a major role in creating a new learning resource, making it possible to discover exactly what forms are most frequent for each lexeme.

3.2 Building and Using the SMARTool

Inspired by our experiment and funded by a grant from the Norwegian Directorate for Higher Education and Skills, I set about creating the Strategic Mastery of Russian Tool (SMARTool) together with a team of colleagues and students at UiT The Arctic University of Norway (UiT), the Higher School of

Economics in Moscow (HSE), and the University of Helsinki (UH): Radovan Bast (UiT), Tore Nesset (UiT), Francis Tyers (HSE), Mikhail Kopotev (UH), Valentina Zhukova (HSE), Elizaveta Kibisova (HSE), Svetlana Sokolova (UiT), Evgeniia Sudarikova (HSE), Ekaterina Rakhilina (HSE), Olga Lyashevskaya (HSE), and James McDonald (UiT). The SMARTool is freely available to the public without any password or login at: https://smartool.github.io/smartool-rus-eng/ and all data and code is stored open-source on github. A subset of the SMARTool, called SMARTool for Min russiske reise (https://smartool.github.io/min-russiske-reise/) serves just the A1 vocabulary broken down according to the lessons in our introductory online course materials (a free MOOC available at https://open.uit.no/courses/course-v1:UiT+mrr+2023/about). The building process and functions of the SMARTool are summarized here (for more details see Janda 2019).

The guiding principles for this project were that: 1) machine learning indicates that focus on the most frequent word forms is the best path to full mastery of inflectional morphology, and 2) language technology resources make it possible to identify the most frequent word forms and the grammatical constructions and collocations that motivate their use. In other words, our aim was to make learning of inflection maximally strategic by focusing on authentic usage. Of course, language teachers have always focused on certain forms and contexts that are commonly encountered, but this has been based on intuition. For the first time we would do this in a scientific way, designing a resource based on empirical evidence.

We aggregated from textbooks a vocabulary of over 3000 inflected words, consisting of nouns, verbs, and adjectives and representing the CEFR (Common European Frame of Reference) language proficiency levels A1, A2, B1, and B2. In the research for Janda – Tyers 2018, we had learned that even high frequency words tend to appear commonly in only three or fewer inflected forms, so our goal was to discover which forms were the most strategic for each of the 3000 vocabulary items. For this task we turned to the SynTagRus corpus, a "gold standard" corpus which offers 100% manually corrected disambiguation of forms. For most words we collected the three most frequent forms, but if fewer than three forms accounted for over 90% of the attestations of a word, then we collected only those forms. For example, over 90% of the attestations of the noun *sentjabr'* 'September' are either the Genitive Singular *sentjabrja* or the Locative Singular *sentjabre*, so we collected only those two forms. Once we had collected the most strategic inflected forms, we needed to identify their typical contexts in order to show how they are used. We consulted a variety of corpora (primarily the RNC and the Collocations Colligations Corpora at http://cococo.cosyco.ru/) to find representative example sentences that we then edited as necessary for the various levels. Finally we designed a user-friendly website.

In the SMARTool, a user first chooses the appropriate proficiency level (A1 through B2, or "all levels") and then selects the vocabulary to focus on through one

of three filters: topic, analysis, and dictionary. All searches return words represented by their three or fewer most frequent inflected forms presented in example sentences. The user can click a button to show English translations of the sentences and can click another button for audio of each sentence. The "Search by dictionary" button returns a list of words at the given level. The "Search by topic" button offers a menu of topics, such as *vremja (time)*, *eda (food)*, and *životnye/rastenija (animals/plants)*, and users can toggle through all the items in the given category. For example, under *eda (food)*, one finds the word *sous* 'sauce' and these three sentences with the top three most common inflected forms of the word:

Vasja prigotovil kuricu v slivočnom souse. (Loc.Sing)
'Vasya cooked a chicken in a creamy sauce.'
Ljuboe mjaso on ljubit est' s soevym sousom. (Ins.Sing)
'He likes to eat all kinds of meat with soy sauce.'
Lučše vsego on gotovil tomatnyj sous. (Acc.Sing)
'Best of all he could cook tomato sauce.'

The "Search by analysis" button is handy for finding words and contexts for specific combinations of grammatical categories. For example, if one wants to find the most strategic words for learning the Dative Plural at the A2 level, the SMARTool returns these items in corpus-inspired example sentences: *pričinam* 'reasons', *sapogam* 'boots', *sportsmenam* 'athletes', *stroiteljam* 'builders', *šaxmatam* 'chess'. If at the B2 level one searches for Perfective Gerunds, one gets a longer list of items including *ogloxnuv* 'deafened', *ogljanuvšis'* '(after) taking a look around', *posočuvstvovav* 'feeling sorry for'.

While the SMARTool provides information on the most likely combinations of all grammatical categories for each word, case is perhaps the most prominent category, since it relates to two of the three parts of speech in the SMARTool – nouns and adjectives – and one of those, nouns, is by far the most common part of speech, both in corpora of Russian and proportionately also in the SMARTool. Therefore, a major strength of the SMARTool is the way it represents case usage.

Another resource inspired by the research in Janda and Tyers 2018 has been created for Czech: GramatiKat (Kováříková et al. 2023; https://korpus.cz/gramatikat/). The GramatiKat interface allows users to view the distribution of morphological case both as a baseline (i.e., for all lexemes of a given part of speech) and for individual lexemes. GramatiKat opens the way for researchers to gauge differences in grammatical distributions between a reference corpus and target texts both overall and at the level of specific lexemes.

Of course, it is one thing to build a resource and quite another thing to get students to actually use it. To this end we have devised a secondary resource with exercises to engage students with the SMARTool: https://smartool.github.io/

exercises/. The SMARTool exercises are of two types, Treasure Hunt and Story Time, designs that emerged from work with a student focus group. Both types of exercise can be part of self-study, assigned as homework, or used in group work in a classroom.

The Treasure Hunt prompts the learner to use a SMARTool search function to gather data to help them to find the answer to a question. The questions range across levels of proficiency and probe various topics relating to patterns that students might not otherwise notice on their own, such as:

- Most Russian words beginning in *a-* or *è-* are foreign borrowings.
- The word *rossijskij* 'Russian' is used to describe items connected to Russia as a state (like *pasport* 'passport' and *Federacija* 'Federation') but *russkij* 'Russian' is used to describe items connected to the Russian language, culture, and ethnic identity (like *alfavit* 'alphabet', *literatura* 'literature').
- The prepositions *na* 'on(to)' and *s* 'from' are used with large open spaces or events, while other places use the prepositions *v* 'in(to)' and *iz* 'from'.

Story Time trains learners to compose texts on various topics, using vocabulary, grammatical constructions, and collocations modeled in the SMARTool. For example, a B1 learner is asked to write 2-3 connected sentences on the topic of *zdorov'e (health)* using a given set of SMARTool vocabulary items, and in the SMARTool the student also finds examples of how these words are used in sentences with specific collocations and grammatical contexts:

- *prinimat'* 'take': + *lekarstvo* 'medicine'
- *operacija* 'operation': + *na* 'on' + Locative; + *provoditsja pod občšim narkozom* 'is conducted under general anesthesia'
- *želudok* 'stomach': *u* 'by' + Genitive + *bolit* 'hurts' + ('X has a stomach ache'); *bol'* 'pain' + *v* 'in' + Locative; *rasstrojstvo* 'upset' + Genitive
- *analiz* 'analysis, test': + *krovi* 'blood'; *rezul'taty* 'results' + Genitive

An enduring theme of our work with the SMARTool has been that inflectional morphology doesn't happen in a vacuum; it is part of a bigger ecosystem of context involving word-specific preferences for both collocations and grammatical constructions. The lack of adequate resources to address this ecosystem motivated us to undertake another project, namely the building of the Russian Constructicon.

3.2 The Russian Constructicon

Like the work on case semantics, this project grew out of a frustration with existing resources. According to Construction Grammar (Goldberg 2006), an entire language can be described in terms of the form-meaning pairings that constitute grammatical constructions, but the vast majority of constructions are not represented in reference works. An example of the multiword constructions that are

underrepresented is NP-Dat Copula *daleko do* NP-Gen, as in *Tebe daleko do lučšego rabotnika* 'You are by far not the best worker (lit. To you it is far to the best worker)'. The Russian Constructicon (Janda et al. 2018 – https://constructicon.github.io/russian/) is an attempt to fill this gap, and our online resource currently provides semantic and syntactic descriptions, examples, and much more for over 4000 Russian constructions on a website that is free, open to the public, and searchable according to a large number of parameters. Case semantics play a role in a large portion of Russian multiword grammatical constructions, and the initial inventory of the Russian Constructicon was based on my earlier work on case semantics, and then later expanded through various methods (Janda et al. 2021). The Russian Constructicon is a multipurpose resource, designed to serve linguists as well as learners and teachers of Russian, and has spawned further publications (Endresen – Janda 2020; Janda et al. 2023a; Janda – Endresen – Zhukova 2024; Zhukova – Janda 2024; Rakhilina et al. 2022).

3.2.1 Theoretical and practical arguments for a constructicon

Linguists traditionally describe languages in terms of a lexicon and the rules of basic grammar that operate on lexemes. The theoretical framework for the Russian Constructicon project, Construction Grammar (Fillmore – Kay 1999, Croft 2001, Tomasello 2003, Fried – Östman 2004, Goldberg 2006), however, takes a very different approach to language description by taking the construction as the basic (but not elementary) unit of language and claiming that an entire language can be described in terms of an interconnected system of constructions. Goldberg (2013, p. 17) defines constructions as "conventional, learned form-function pairings at varying levels of complexity and abstraction". This definition is intentionally very broad; it recognizes all language structures as constructions. At the extremes of the two dimensions of complexity and abstraction are items that are readily recognized by traditional linguistics. Examples of highly complex constructions are entire discourse structures such as an interview or a short story. The minimum of complexity is a simplex item with only one unit, and these can be found at both ends of a continuum from concrete to abstract. The concrete simplex items of language are individual words and morphemes, like the Russian adverb *daleko* 'far' and the preposition *do* 'to', and these are represented in dictionaries. In constructions, we term such items "anchors". The abstract simplex items of language are bits of core grammar such as the subject of a sentence or the object of a preposition and are defined by the grammatical categories they express, such as case, tense, etc. In constructions we call such items "slots" and refer to the lexemes that fill slots as "fillers". These abstract slots belong to the core syntax typically described in a grammar. In our construction above there are two NP slots, each with a case value (Dative and Genitive), as well as a copula for which the tense is not specified. Between these extremes there are thousands of essential multi-word expressions

comprised of one or more anchors and/or fillers, the vast majority of which are not represented in traditional reference works. While all of these items, both the extremes and the multi-word expressions, are constructions, Construction Grammar tends to focus primarily on the multi-word expressions in an attempt to fill this gap.

A constructicon is a collection of the constructions of a language. While it is perhaps not feasible to create a resource that would contain all of the constructions of a language, the Russian Constructicon project takes seriously the tenet that this is in principle possible, resulting in the largest existing constructicon for any language, currently with over 4000 constructions. The patterns that emerge from this large-scale constructicon make it possible to trace the relationships that hold among constructions and the contexts in which various phenomena exist. For example, rather than investigating reduplication in isolation, it is now possible to extract the subset of grammatical constructions that have repeated elements and reveal their relationships to the rest of the Russian Constructicon (Janda – Endresen – Zhukova 2024). The Russian Constructicon is not a list. We find that "no construction is an island"; the Russian Constructicon is an interconnected system of thousands of constructions in which lexicon and grammar are fully integrated. Particularly striking are the arrays of semantic connections that join constructions into groupings across all levels, from the most local families of (nearly) synonymous constructions to the most abstract high-level semantic classes. We additionally find a variety of syntactic affinities across constructions, as well as links based on morphology, and the use of specific anchor and filler lexemes (for more on the systematic relationships among constructions, see Zhukova – Janda 2024).

3.2.2 Building and using the Russian Constructicon

The Russian Constructicon is an ongoing team effort that has involved collaboration between faculty and students at both UiT The Arctic University of Norway and the Higher School of Economics in Moscow. Some of my most prominent collaborators are: Radovan Bast (UiT), Anna Endresen (UiT), Daria Mordashova (HSE, MGU), Ekaterina Rakhilina (HSE), Valentina Zhukova (UiT), and at least forty students over a period of nearly a decade have contributed. The Russian Constructicon project has received financing from the Norwegian Directorate for Higher Education and Skills, the Ministry of Science and Higher Education of the Russian Federation, and the National Research Foundation of Korea. The Russian Constructicon is a free open-source resource available without registration or password.

Case semantics has played a major role in the Russian Constructicon from the very beginning, when our first collection of constructions was derived from the pages of the Case Book for Russian (Janda – Clancy 2002). Since then we have employed a variety of methods, including manual collection from reading texts and scripts, semiautomatic collection of frequent multiword collocations, and intuitive

probing of native speakers' competence to fill out families of (nearly) synonymous constructions (for details on this process, see Endresen et al. to appear). And since nearly every grammatical construction contains a noun phrase or an adjective or a participle (i.e., something that can be inflected for case), grammatical case figures prominently in the entire Russian Constructicon.

When a user opens the Russian Constructicon page, they find a window where they can browse over 4000 constructions. From this homepage is possible to filter constructions by typing in specific anchor words or slot tags, as well as selecting a proficiency level (from A1 to C2). For example, if we type in (using Cyrillic) the word *daleko* 'far', we find eleven constructions with that anchor word, among them the construction mentioned above. When we click on that construction, we get this information (here additionally annotated with information in square brackets, and with all Cyrillic rendered in Latin transcription, and translations of Russian text):

> 473 [an ID number used internally by developers]
> **NAME NP-Dat Cop daleko do NP-Gen** [the name of the construction]
> *Tebe daleko do lučšego rabotnika.* [a short recognizable illustration of the construction, here: You are far from being the best worker]

DEFINITION (Russian) [most constructions come with a definition in Russian, some also have Norwegian and English definitions, this is still under development]
Konstrukcija oboznačaet, čto [učastnik situacii]$_{\text{Participant}}$ ili [ob"ekt]$_{\text{Theme}}$ ne obladaet dostatočnymi kačestvami i nedostatočno xoroš, čtoby byt' kak [ètalon]$_{\text{Standard}}$. Konstrukcija osnovana na sravnenii i soderžit ocenočnuju xarakteristiku vozmožnostej ili kačestv [učastnika]$_{\text{Participant}}$ ili [ob"ekta]$_{\text{Theme}}$ kak značitel'no ustupajuščix vozmožnostjam ili kačestvam [togo ètalona, s kotorym oni sravnivajutsja]$_{\text{Standard}}$. Kak esli by govorjaščij sčital, čto rasstojanie ot učastnika ili ob"ekta do ètalona očen' veliko.
[The construction indicates that the [participant in the situation]$_{\text{Participant}}$ or [object]$_{\text{Theme}}$ does not possess sufficient qualities and is not good enough to be like the [standard]$_{\text{Standard}}$. The construction is based on comparison and contains an evaluative characteristic of the abilities or qualities of the [participant]$_{\text{Participant}}$ or [object]$_{\text{Theme}}$ as significantly inferior to the abilities or qualities of the [standard with which they are compared]$_{\text{Standard}}$. It is as if the speaker believes that the distance from the participant or object to the standard is very great.]

> **EXAMPLES** [Five corpus examples are given, but here we show only one]
> 1. Vidite li, delo v tom, čto [gubernatoru oblasti]$_{\text{Participant}}$ eščë daleko do [prezidenta]$_{\text{Standard}}$.
> [You see, the point is that the regional governor is far from being the president.]

CEFR LEVEL: A2

When the user clicks to get additional information, they can find: equivalent constructions in Norwegian and English; common fillers; the semantic and syntactic types of the construction; the syntactic function, structure, and part of speech of the anchor; the dependency structure of the name of the construction and its illustration; the communicative type of the construction (e.g., Declarative); a usage label (e.g., Colloquial); a comment (often citing closely-related constructions); and references to relevant scholarly works.

The Advanced search page of the Russian Constructicon allows the user to filter constructions according to all parameters for which constructions are tagged: semantic types, semantic roles, morphology, syntactic type of construction, syntactic function of anchor, syntactic structure of anchor, part of speech of anchor, and CEFR level. Among other things, this makes it possible to search for constructions that involve each of the grammatical cases.

On the Daily dose page a user can choose a proficiency level and receive five randomly selected grammatical constructions to train on.

The Statistics page shows graphs and raw numbers for the distribution of syntactic types of constructions, syntactic functions of anchors, and semantic types of constructions. The graph for semantic type can be modified to show only the distribution for a selected syntactic type. This page also lists the ten most frequent anchor words for each of three parts of speech: verbs (starting with *znat'* 'know', *govorit'* 'say', *xotet'* 'want'), nouns (starting with *vremja* 'time', *delo* 'thing, case', *raz* 'time'), and adjectives (starting with *ravnyj* 'equal', *xorošij* 'good', *polnyj* 'full').

The site has an Instructions page to guide the user through all the terms and conventions and an About page that describes the project. There is also a YouTube channel with instructional videos about the project: https://www.youtube.com/channel/UC8q-_F8c8bx9gI7fYET1-dQ.

Several spinoff projects are under development, including constructicons for Ukrainian, Persian, and Hill Mari. Since the code is open-source and publicly available it is possible for researchers who wish to create constructicons for other languages to reuse and adapt our model.

As with the SMARTool, a further challenge is to make the Russian Constructicon more accessible to language learners. To this purpose we have created a related resource, called Construxercise!: https://constructicon.github.io/construxercise-rus/. This resource facilitates hands-on learning of Russian constructions through exercises aimed at a strategic group of 57 Russian highly frequent discourse constructions that students can use to structure a discourse by doing things like introducing a topic, clarifying a point, giving an example, adding information, expressing an opinion, asking someone for their opinion, hedging, drawing a conclusion, etc. The constructions that support these skills are presented in twelve

lessons on topics like Introducing yourself, Getting a job, Getting around, Holiday celebrations. Construxercise! is conceived of as a multifunctional resource that serves the needs of different types of users and offers educational materials that can be used as either a central or complementary teaching component, either in class or for self-guided study.

4. ANALYSIS OF POLITICAL DISCOURSE: PUTIN MAKES HIS CASE

Of course, it is not only linguists and language learners who use case. Grammatical case is a feature of over two-thirds of the world's languages, used by all speakers of those languages, and it makes sense to ask what role case is playing particularly in the most powerful of those speakers. This brings us to Putin and a question that has bothered me for a long time: Why is Putin so popular? Why do Russians find him convincing? Public opinion polls (see https://www.levada.ru/en/ratings/, https://media.fom.ru/fom-bd/d46pi2024.pdf) have consistently shown Putin's approval rating at between 60% and 90% over the past quarter century (at the time of this writing it stands between 82% and 87%). Although opinion polls carried out in Russia are not entirely reliable, certainly this means that there are a lot of Russians who stand behind their leader. Putin is not a brilliant orator, as anyone who has watched his hours-long speeches can attest, but maybe there is something in the way he delivers his messages that makes them compelling to his listeners. In Janda et al. 2023b we looked just at how Putin uses grammatical case, and found consistent deviations from Russian norms. This research was carried out in collaboration with Masako Fidler (Brown University), Václav Cvrček (Charles University), and Anna Obukhova (UiT) and funded by a grant from the Norwegian Research Council (https://threat-defuser.org/).

Our research is based on four assumptions. The first is that a corpus of a language of a sufficient size can serve as a proxy for the linguistic experience and expectations of native speakers. A corpus is perhaps an imperfect representation but it is the closest thing we have to a model of the input that gives a native speaker their special competence in a language, their conscious and unconscious knowledge of the norms of their language. Second: speakers are known to be sensitive to deviations from these norms. Third: while words can be consciously chosen, grammar is less under conscious control and more systematic. Fourth: grammar and meaning are joined in a semantic continuum; grammar is not just empty scaffolding.

In a nutshell, our idea was to compare Putin's use of grammatical case with what we find in a corpus of Russian and analyze the deviations for how they support his political messages. To this end, we performed the first extension of Keyword Analysis to a new methodology we call "Keymorph Analysis". Keyword Analysis (cf. Scott 1996) is a well-established method widely used in corpus-

assisted discourse analysis. Keyword Analysis focuses on the distribution of words, identifying as "keywords" those that are unusually frequent in a target text as compared to a reference corpus. In this way keywords reveal the "aboutness" of a text. However, Keyword Analysis has mostly been performed on English, which has little morphology and no grammatical case. We created the first proof-of-concept for Keymorph Analysis using as our target text Putin's speeches during a three-week period leading up to and following the full-scale invasion of Ukraine in February 2022 (34,720 tokens), and as our reference corpus the Russian InterCorp portion of the Czech National Corpus (www.korpus.cz, 20.1 million tokens).

We examined Putin's use of case with three words: *Rossija* 'Russia', *Ukraina* 'Ukraine', and *NATO* 'NATO'. These three words occur a total of 395 times in the Putin target text and 7801 times in the reference corpus. All attestations of these words in the target text were manually annotated for the precise case meaning expressed. The relevant case meanings that appeared most often with these words in both the target text and the reference corpus are the following:

- Nominative: agent (subject); label
- Genitive: agent or patient; possession
- Dative: potential agent (usually human)
- Accusative: patient (direct object); destination
- Instrumental (with preposition *s*): collaborator
- Locative: a place

We found that Putin's use of grammatical case with the three nouns deviates significantly from the case usage observed in the reference corpus, and that Putin's usage strongly underpins his political message. *Rossija* 'Russia' is statistically overrepresented in the Genitive and Dative cases, *Ukraina* 'Ukraine' is overrepresented in the Genitive case but underrepresented in all other cases, and *NATO* 'NATO' is overrepresented in the Accusative case and strongly underrepresented in the Dative and Instrumental cases.

Closer examination of the specific case meanings that Putin uses is more revealing. *Rossija* 'Russia' is represented as a dynamic agent (Nominative subject of transitive verbs), a collaborator (Instrumental case), a victim that has been treated unfairly (Accusative), and as a humanized entity that inspires empathy (Dative). *Ukraina* 'Ukraine' by contrast plays a passive role (Nominative subject with stative verbs), is manipulated (Accusative) and dehumanized (Dative severely underrepresented), is not a collaborator (Instrumental severely underrepresented), and is merely a location or region (use of *na* 'on' + Locative and Genitive). *NATO* 'NATO' is similarly dehumanized and not seen as an agent (Nominative) or a collaborator (Instrumental). NATO's signature role in Putin's narrative is as a future destination for Ukraine (Accusative, Locative). In sum, Putin depicts Russia

as a dynamic, agentive, foregrounded actor, a reliable partner for collaboration, but also the victim of unfair geopolitical maneuvers. Ukraine, by contrast, is dehumanized, relatively static, and backgrounded, often merely a territorial location rather than a state. NATO appears primarily as the label for an untrustworthy organization and as a destination for Ukraine.

One year after the full-scale invasion of Ukraine, on February 21, 2023 Putin delivered a speech to the Federal Assembly (10,538 tokens), which news media declared to be "more of the same". However, in a further analysis we found some important shifts in Putin's message conveyed by grammatical case. In this speech Putin emphasized the great potential of Russia's self-sufficient economy and the ways that Russia has been unfairly targeted by the West. Ukraine was mentioned only twelve times in this speech, referred to mainly as Russia's "historical territories" and the West's "Anti-Russia". NATO was no longer depicted as the destination of Ukraine, but instead foregrounded as an aggressor.

We have demonstrated that Keymorph Analysis can complement Keyword Analysis and other traditional methods of discourse analysis. Over- and underrepresentation of grammatical cases can be identified by measuring deviations from corpus norms. This method of analyzing grammatical case reveals the roles of social actors in a discourse, and can be used not only by linguists, but also in the disciplines of the social sciences. While one's choice of words is deliberate and conscious, grammatical case is obligatory and serves as a second channel for signaling the roles notions have in a discourse. We reason that consistent deviation from grammatical norms likely has an impact on hearers, driving home a message like a steady drumbeat. Our results invite further comparisons, for example of Putin with other politicians, and with messages in various types of manipulative texts.

5. CONCLUSION

This story of linguistic theory and its application to language pedagogy is both a professional one and a personal one. The meanings of grammatical case that so frustrated me as a student have inspired an enduring fascination that leads in many directions at once. No matter where I turn, the cases keep coming back to me. This research agenda has supported the core tenet of Cognitive Linguistics that grammar has meaning. I have learned that native speakers probably don't have a full set of paradigms in their heads; instead they most likely triangulate from many smaller partly overlapping subsets of paradigms comprised of the most common forms for individual words. We made a resource to reflect this finding and I changed my pedagogical approach accordingly. We have filled in the some of the gaps between what we find in dictionaries and grammar books with descriptions of thousands of multi-word constructions. And we have used the statistical distribution of grammatical case to probe the ideological messages of Vladimir Putin.

References

CROFT, William (2001): *Radical Construction Grammar*. Oxford: Oxford University Press.

DRYER, Matthew S. – HASPELMATH, Martin (eds.) (2013): *WALS Online (v2020.4)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13950591.

ENDRESEN, Anna – JANDA, Laura A. (2020): Taking Construction Grammar One Step Further: Families, Clusters, and Networks of Evaluative Constructions in Russian. In: *Frontiers in Psychology*. Special issue: M. Putnam – M. Carlson – A. Fábregas – E. Wittenberg (eds.): *Defining Construction: Insights into the Emergence and Generation of Linguistic Representations*). In: Vol. 11, No. 574353, pp. 1–22. Available at: https://doi.org/10.3389/fpsyg.2020.574353.

ENDRESEN, Anna – JANDA, Laura A. – RAKHILINA, Ekaterina – ZHUKOVA, Valentina – MORDASHOVA, Daria (to appear): From a list to a network: Growing the Russian Constructicon via research on families of constructions. In: *Constructing constructicons.* Amsterdam: John Benjamins Publishing Co.

FILLMORE, Charles J. – KAY, Paul (1999): *Construction Grammar.* Berkeley: University of California.

FRIED, Mirjam – ÖSTMAN, Jan-Ole (2004): Construction grammar: A thumbnail sketch. In: M. Fried – J.-O. Östman (eds.): *Construction grammar in a crosslanguage perspective*. Amsterdam: John Benjamins Publishing Co., pp. 11–86.

GOLDBERG, Adele E. (2006): *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.

GOLDBERG, Adele (2013): Constructionist approaches. In: T. Hoffmann – G. Trousdale (eds.): *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, pp. 15–31.

JANDA, Laura A. (1993): *A Geography of Case Semantics: The Czech Dative and the Russian Instrumental* (*Cognitive Linguistics Research, Vol. 4*). Berlin: Mouton de Gruyter.

JANDA, Laura A. (1999): Peircean semiotics and cognitive linguistics: a case study of the Russian genitive. In: M. Shapiro (ed.): *The Peirce Seminar Papers*. New York/Oxford: Berghahn Books, pp. 441–466.

JANDA, Laura A. (2000): A cognitive model of the Russian accusative case. In: R. K. Potapova – V. D. Solov'ev – V. N. Poljakov (eds.): *Trudy meždunarodnoj konferencii Kognitivnoe modelirovanie*, No. 4, part I. Moscow: MISIS, pp. 20–43.

JANDA, Laura A. (2019): Businessmen and Ballerinas Take Different Forms: A Strategic Resource for Acquiring Russian Vocabulary and Morphology. In: *Russian Language Journal*, Vol. 69, No. 1, pp. 175–193.

JANDA, Laura A. – CLANCY, Steven J. (2002): *The Case Book for Russian*. Bloomington, IN: Slavica Publishers.

JANDA, Laura A. – CLANCY, Steven J. (2006): *The Case Book for Czech*. Bloomington, IN: Slavica Publishers.

JANDA, Laura A. – ENDRESEN, Anna – KUZNETSOVA, Julia – LYASHEVSKAYA, Olga – MAKAROVA, Anastasia – NESSET, Tore – SOKOLOVA, Svetlana (2013): *Why Russian aspectual prefixes aren't empty: prefixes as verb classifiers*. Bloomington, IN: Slavica Publishers.

JANDA, Laura A. – TYERS, Francis M. (2018): Less is More: Why All Paradigms are Defective, and Why that is a Good Thing. In: *Corpus Linguistics and Linguistic Theory.* Vol. 17, No. 1., pp. 109 – 141. Available at: https://doi.org/10.1515/cllt-2018-0031.

JANDA, Laura A. – LYASHEVSKAYA, Olga – NESSET, Tore – RAKHILINA, Ekaterina – TYERS, Francis M. (2018): Chapter 6. A Constructicon for Russian: Filling in the Gaps. In: B. Lyngfelt – L. Borin – K. Ohara – T. Timponi Torrent (eds.): *Constructicography: Constructicon development across languages* [*Constructional Approaches to Language* 22]. Amsterdam: John Benjamins Publishing Co., pp. 165–181. DOI 10.1075/cal.22.06jan. Available at: https://constructicon.github.io/russian/.

JANDA, Laura A. – ENDRESEN, Anna – ZHUKOVA, Valentina – MORDASHOVA, Daria – RAKHILINA, Ekaterina (2021): How to build a constructicon in five years: The Russian Example. In: F. Brisard – T. Colleman – A. De Wit – R. Enghels – N. Koutsoukos – T. Mortelmans – M. Sol Sansiñena (eds.): *The Wealth and Breadth of Construction-Based Research. Belgian Journal of Linguistics 34*, pp. 161–173.

JANDA, Laura A. – ENDRESEN, Anna – RAKHILINA, Ekaterina – ZHUKOVA, Valentina – MORDASHOVA, Daria (2023a): From data to theory: an emergent semantic classification based on a large-scale constructicon. In: *Constructions and Frames*, Vol. 15, No. 1, pp. 1–58. Available at: https://doi.org/10.1075/cf.00066.jan.

JANDA, Laura A. – FIDLER, Masako – CVRČEK, Václav – OBUKHOVA, Anna (2023b): The case for case in Putin's speeches. In: *Russian Linguistics*, Vol. 47, No. 1, pp. 15–40. Available at: https://doi.org/10.1007/s11185-022-09269-2.

JANDA, Laura A. – ENDRESEN, Anna – ZHUKOVA, Valentina (2024): Typology of reduplication in Russian: constructions within and beyond a single clause. In: M. Kopotev – K. Kwon (eds.): *Constructions with lexical repetitions in East Slavic languages. Trends in Linguistics. Studies and Monographs [TiLSM]. Vol. 384*. De Gruyter Mouton, pp. 71–96. Available at: https://doi.org/10.1515/9783111165806-003.

KOVÁŘÍKOVÁ, Dominika – BENEŠ, Martin – SMEJKALOVÁ, Kamila – KOVÁŘÍK, Oleg (2023): The Structuralist tradition meets empirical data: Corpus data enhancing the Czech Internet Language Reference Book. In: *Word Structure*, Vol. 16, No. 2–3, pp. 233–257. Avaolable at: https://www.euppublishing.com/doi/10.3366/word.2023.0230.

LANGACKER, Ronald W. (2008): *Cognitive Grammar.* Oxford: Oxford University Press.

RAKHILINA, Ekaterina – ZHUKOVA, Valentina – DEMIDOVA, Daria – KUDRYAVTSEVA, Polina – ROZOVSKAYA, Gloria – ENDRESEN, Anna – JANDA, Laura A. (2022): Фразеология в ракурсе «Русского конструктикона».In: *Trudy Instituta Russkogo Iazyka imeni V. V. Vinogradova*, Vol. 2, No. 32, pp. 13–44. Available at: https://doi.org/10.31912/pvrli-2022.2.2.

ROSCH, Eleanor (1973a): Natural categories. In: *Cognitive Psychology* Vol. 4, No. 3, pp. 328–350.

ROSCH, Eleanor (1973b): On the internal structure of perceptual and semantic categories. In: T. E. Moore (ed.): *Cognitive Development and the Acquisition of Language*. New York: Academic Press, pp. 111–144.

SCOTT, Mike (1996): *Wordsmith tools*. London: Oxford University Press – Lexical Analysis Software. Available at: https://lexically.net/wordsmith/.

TOMASELLO, Michael (2003): *Constructing a Language: A Usage-Based Approach.* Cambridge: Harvard University.

ZHUKOVA, Valentina – JANDA, Laura A. (2024): Russian grammar as a constructicon: Beyond a list. In: *Russian Grammar Updated*, a special issue of *Russian Linguistics*, Vol. 48, No. 15. Available at: https://doi.org/10.1007/s11185-024-09298-z.

ZIPF, George K. (1949): *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison-Wesley.

sciendo
DOI 10.2478/jazcas-2024-0045

CLITICS IN SLOVAK WITH(OUT) CONSTRAINTS. CORPUS FREQUENCY VERSUS ACCEPTABILITY RATINGS

MARTINA IVANOVÁ
Institute of Slovak and Media Studies, Faculty of Arts, University of Prešov, Slovakia
Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences, Bratislava, Slovakia

IVANOVÁ, Martina: Clitics in Slovak with(out) constraints. Corpus frequency versus acceptability ratings. Jazykovedný časopis (Journal of Linguistics), 2024, Vol. 75, No. 3, pp. 291 – 308.

Abstract: Using data from a representative corpus of Slovak and an acceptability survey, the preferential placement of clitic components in object clause constructions in Slovak have been investigated. Slovak clitics are usually described as elements following Wackernagel's Law and belonging to the category of second-position clitics. However, usage-based investigations show that their placement varies within a clause, depending on various pragmatic and syntactic factors and a set of constraints which limit their movement within the clause structure. By comparing data from corpus analysis and acceptability ratings by native speakers, it is shown how judgments and actual usage of clitics may converge or diverge in particular cases.

Keywords: clitics, word order, Wackernagel's Law, corpus, acceptability judgements, barriers, clitic climbing, object clause constructions.

1. INTRODUCTION

In most languages, sentence constituents may be linearized in two or more different ways, at least in some well-formed sentences. Nevertheless, all natural languages are restrictive in one way or another: no language allows for all possible linear orders of sentence categories in 100% of sentences, linearization constraints are salient for all word order systems. In Slavic languages like Slovak, most combinations of scrambling types are available for sentence categories represented by non-clitic words, while the number of scrambling types available for clitics is more reduced (cf. Zimmerling 2011, p. 754). The goal of the paper is to investigate possible patterns of clitic placement in object control clauses on the basis of the corpus data and to answer the question how these word order variants are evaluated by speakers in an acceptability rating experiment.

Clitics represent one of the most specific and intricate word order phenomena of many languages. Slovak belongs to those languages which follow Wackernagel's Law, and its clitic elements belong to the category of second-position clitics (2P) (cf. Franks – Holloway King 2000). However, Wackernagel's Law is not always

applicable without exception and the pattern of clitic placement undergoes systematic variations under certain conditions. The cues from the grammar system and information structure can interfere in the linearization patterns of clitic components and remove them from the second position.

The paper is structured as follows: The introductory remarks are presented in Section 1. Section 2 considers phonological and non-phonological definitions of clitics as usually presented in various linguistic approaches. Section 3 provides a concise overview of the methodological framework adopted in the paper as well as the design of the dataset including both corpus and experimental data. In Section 4, results of the corpus investigation and survey of clitics placement in object control clauses are presented. Section 5 summarizes the findings.

2. CLITICS IN SLOVAK

2.1. Definition of clitics

Clitics can be compared with full words and affixes.

As opposed to full words, clitics are typical of:

(i) prosodic deficiency: they are elements "that lack independent stress" (Pescarini 2021, §1.1), they are words in the morpho-syntactic sense, but not in the phonological sense (Booij 2012, p. 290), therefore they are unable to appear sentence-initially;

(ii) bondedness: they cannot occur in isolation, they are "defective in their phonological representation and therefore have to prosodically combine with an adjacent non-clitic word" (Ionova 2019, p. 22), usually termed as prosodic host.

In their paper on the English negative *-n't*, Zwicky and Pullum (1983, pp. 502–504) contrast clitics with affixes. There are two characteristics which set clitics apart from affixes:

(i) non-selectivity or promiscuity: they are typical of the lack of word-class selectivity, i.e. they are not selective with regard to their host;

(ii) morphological stability: affixed words tend to display morphophonological and semantic idiosyncrasies; clitic groups do not (cf. Zwicky – Pullum 1983, p. 504).

Clitics can be described as elements with "triple" citizenship. Phonologically, they lean on their prosodic hosts; positionally, they precede or follow their structural host or anchors; and functionally, they form morphological, lexical, or syntactic units with their matrix item. Depending on their contextual environment, the roles of prosodic host, anchor, and matrix item can overlap and be expressed by the same sentence component or, alternatively, different sentence components can fulfil the roles of prosodic host, anchor, and matrix item.

According to Haspelmath (2023) anchor is the word preceding an enclitic, and the word following a proclitic, whereas host is the element with which a clitic forms a prosodic word.

In example (1), *by* is an enclitic as it cannot occur at the beginning of a free form (**by bolo*), the infinitive *prehovárať* can be described as an anchor as it precedes the enclitic *by* and the component *by* forms a prosodic word with verb form *bolo* so that *bolo* is its prosodic host. At the same time, *bolo by* is an analytical grammatical form in Slovak, so that *bolo* is a matrix expression for the conditional component *by.*

(1) *Prehovárať ju* ***by*** *bolo zbytočné.*
persuade-INF her-ACC COND be-PAST-NEUTR.SG useless
'It would be useless to persuade her.'

However, in word order variant (1a), the verb *bolo* fulfils both the role of anchor and host.

(1a) *Bolo* ***by*** *zbytočné prehovárať ju.*
be-PAST-NEUTR.SG COND useless persuade-INF her-ACC

It confirms the claim of Franks and Holloway King (2000) who argue that the direction of prosodic attachment of clitics is underspecified, i.e. that it can attach to a host to their right as well as to a host to their left.

2.2. The sets of investigated clitics

The following set of criteria can be applied in classification of clitic components:

(i) Tenacity criterion: refers to the fact whether clitic items keep their clitic status in different contextual environments:

- clitics tantum (or constant 'clitics' – cf. Rosen 2001, Hana 2007; pure clitics – cf. Avgustinova – Oliva 1997) are elements which always appear in the second position;

- volatile clitics (or semi-clitics – cf. Avgustinova – Oliva 1997) can have phonological autonomy under certain contextual conditions and thus appear in the second position only optionally.

The clitic status of semi-clitics can be proved by their realization within the clitic cluster. On the basis of the rule described in Hana (2007, p. 76), element X between 1P and clitic component is a clitic:

(2) *Oni* ***nám*** ***ho*** *vzali.*
They-NOM us-DAT it-ACC take-PAST-PL
'They took it from us.'

As the semi-clitic component *nám* is interposed between 1P (*Oni*) and the permanent clitic (*ho*), it can be considered a clitic;

(ii) Functional criterion: refers to the functional status of the clitic component:

(a) auxiliary verbal clitics: assist main verbs in conveying person and number grammatical meanings in the past participle forms: *(písal) som* '(wrote)-PRES.SG.1', *si* '(wrote)-PRES.SG.2', *sme* '(wrote)-PRES.PL.1', *ste* '(wrote)-PRES.PL.2';

(b) non-reflexive argument clitics: clitics that refer to arguments of the verb:

- short forms of personal pronouns with existing long counterparts: *ma* 'me-ACC' (as opposed to *mňa*), *ťa* 'you-SG.ACC' (as opposed to *teba*), *ho* 'him-ACC' (as opposed to *jeho*, *neho*), *mi* 'me-DAT' (as opposed to *mne*), *ti* 'you-SG.DAT' (as opposed to *tebe*), *mu* 'him-DAT' (as opposed to *jemu*, *nemu*),

- short forms of personal pronouns missing long counterparts: *nás* 'us-ACC', *vás* 'you-PL.ACC', *ich* 'them-ACC', *ju* 'her-ACC', *jej* 'her-DAT', *nám* 'us-DAT', *vám* 'you-PL.DAT', *im* 'them-DAT';

(c) reflexive clitics *sa*, *si*: are notoriously ambiguous when it comes to their functional status, usually described as verbal components or pronominal expressions;

(d) particle clitics: a special form of the originally auxiliary *be*-form *by* used to build periphrastic form of the conditional mood.

3. METHODOLOGY AND DESIGN OF DATASETS

3.1. Aim and framework

The present paper combines synchronic corpus analyses with an experimental method, namely acceptability judgements to assess frequency distributions and speakers' acceptance of different word-order variants concerning clitic placement in Slovak. This approach stems from those works which emphasize that grammaticality can be operationalized by acceptability, e.g. Riemer (2009).

Acceptability judgments are found by many researchers to be a useful source of data, although, as with any source, they must be used carefully (see e.g. Sprouse 2007, 2008, 2009) for a discussion of the limits of acceptability judgments.

Research experiments on possible correlations between corpus data and acceptability ratings have revealed so far that there is a correlation between corpus data and acceptability, but it is not proportional or symmetric, i.e. we cannot count on the proportions to correspond precisely to value judgments, nor is it always possible to abstract predictions about acceptability from corpus data (Bermel – Knittl 2012, p. 246). Kempen and Harbusch (2005) as well as Bader and Häussler (2010) find considerable support for the thesis that corpus frequency is not a fully reliable predictor of acceptability. This phenomenon, known as the "frequency/acceptability mismatch", also called the "grammaticality/frequency gap", refers to the observation that there is no reliable correlation between the frequency of a grammatical unit and its acceptability.

In the following parts, the present paper brings results of a corpus and survey analysis which were conducted to understand the factors determining the word order variability of clitic components.

3.2. Dataset design

The data for the current research are of twofold sources: corpus data and experiment based on an acceptability judgment task.

Corpus data for this research were retrieved from corpus Omnia Slovaca IV Maior Beta which has 6 596 573 997 tokens, and it is compiled by the Slovak National Corpus and web corpora. Occurrences of word order pattern are retrieved from the corpus using CQL queries in which morphological tags with word- and lemma-based attribute searches are combined. To get information on variation in the word order patterns, first individual word order variants were searched for, and then sifted manually to remove erroneous results.

Experimental data were summoned using a survey which took place in spring 2023 over the web in the form of an online questionnaire. All in all, 153 respondents (the majority of them were students at different universities or teachers at schools across the country who were recruited online) were asked to evaluate different word order variants of the same structures with respect to their acceptability. Table 1 brings a closer look at the characteristics of the respondents' sample.

Table 1. Age, profession and gender of respondents

Age	n =	Profession	n =	Gender	n =
0 – 18	2	Linguists	36	M	33
19 – 35	23	Non-linguists	117	F	120
26 – 35	40				
36 – 45	35				
46 – 55	27				
56 – 65	10				
66 – 75	6				

To assess the acceptability of certain word order patterns, gradient Likert's scale with numerical values from 1 to 5, i.e. from fully acceptable vs. fully unacceptable was used. The scale had descriptors at all points from 1 to 5. For the individual stimuli, examples were taken from the corpus wherever possible, sometimes simplifying and toning them down to avoid having respondents react to irrelevant elements in the sentence. The order of word order patterns was randomized.

Responses on the Likert scale are regarded as ordinal rather than interval data, suggesting that non-parametric tests should be our first resort. However, properly designed and implemented Likert-scale linguistic surveys are often subjected to parametric analysis, which can, in many instances, be more accurate and revealing. Correlation tests, which are commonly used on experimental data to show relationships between the variables, were used for the analysis. Out of possible correlation tests, a two tailed t-test for independent samples was used and two dependent variables were tested in the experiment: Age and Profession.

Age as a dependent variable is said to have some effect on the choice of word order pattern, as was shown in analytical works on clitic placement in Slovak (e.g. Ivanová, to be published). Namely, the older generation shows more acceptability of

the patterns which are viewed as stylistically marked by younger generation (the support for these claims can be found in the work by Ivanová (to be published)).

In Spencer (1973, p. 87), one can find the view that it is possible that the behaviour of producing linguistically relevant intuitions has developed into a specialized skill, no longer directly related to the language behaviour of the speech community. As a consequence, the judgements of linguists may be an artifactual system which reflects the accretion of conceptual organization by linguists. This is why the data from the survey are calculated for the groups of linguists (L) and non-linguists (NL) individually.

In the case of Age as a dependent variable, the following Null hypothesis and Alternative hypothesis were formulated:

Null hypothesis	Alternative hypothesis
There is no difference between the *46 – 75 years old* and the *0 – 45 years old* groups with respect to the dependent variable Value	There is a difference between the *46 – 75 years old* and *0 – 45 years old* groups with respect to the dependent variable Value

In case of Profession as a dependent variable, the following Null hypothesis and Alternative hypothesis were formulated:

Null hypothesis	Alternative hypothesis
There is no difference between the *L* and *NL* groups with respect to the dependent variable Value	There is a difference between the *L* and *NL* groups with respect to the dependent variable Value

4. RESULTS

4.1. Clitic placement in complex clauses

In the given subsection, more attention will be paid to cases when clitic or clitic cluster is licensed not by the predicate in the matrix clause, but by a predicate in a superordinate non-finite clause. In such cases, the clitic which is associated with a verb complex in a subordinate clause can actually be pronounced in a construction with a higher predicate even though it may have no obvious semantic or syntactic connection to that verb (Spencer – Luís 2012, p. 162). Such a phenomenon is referred to as clitic climbing and it is defined as a realization of clitics in a syntactic constituent higher than the licensing predicate (cf. Kulik 2023, p. 211).

The aim in the present subsection is to classify the configurations in which climbing is possible or barred in Slovak. Two types of complex clauses have to be distinguished: if the subject of a matrix verb controls the reference of the PRO subject of its infinitival complement, the verb is called a subject control verb; if the

object of a matrix verb controls the reference of the PRO subject of its infinitival complement, the verb is called an object control verb. The phenomenon of clitic climbing in object control clauses has been widely discussed in many studies on Czech clitics. They have observed that, in the case of infinitive complements, Czech pronominal and reflexive clitics behave in a different way: while clitics can climb out of infinitives which are governed by raising and subject control matrix verbs, some additional restrictions occur in the case of object control matrix verbs (e.g. Dotlačil 2004, Rezac 2005, Hana 2007). On the other hand, there are authors who completely reject possibility of climbing in object control clauses (Junghanns 2002). One such study on clitic climbing proved that some additional restrictions also occur in the case of subject control clauses, (cf. Ivanová (to be published)).

In structures with multiple predicates, clitic climbing can be:

- obligatory: clitic climbing out of infinitival complements of modal verbs is necessary (according to Veselovská 1995, p. 305) and the same applies to complex clauses with phasic matrix verbs like *začať* 'to begin' (Adam 2024, p. 49),
- optional: clitic climbing out of infinitival complements of verbs with subject control is possible and it competes with local placement of clitics within infinitival phrases in these configurations,
- blocked: clitic climbing is blocked in case of some object control clauses (cf. Dotlačil 2004; Rezac 2005; Hana 2007).

The aforementioned rules are not in effect without any exceptions. Even though the local placement of clitics is rather limited with modal verbs, it is possible, for example, in tentative remoteness constructions (cases where a rather vague element of tentativeness, diffidence, extra politeness comes into play):

(3) *To by chcelo pustiť* ***sa*** *do nejakej*
it-NOM COND want-PAST-NEUTR.SG start-INF REFL in some
ucelenej koncepcie.
coherent concept-GEN
'It would be desirable to form some coherent concept.'

(3´) **To by* ***sa*** *chcelo pustiť do nejakej ucelenej koncepcie.*

On the other hand, corpus data bring the evidence on clitic climbing out of object control clauses leading to formation of clitic clusters in which infinitive clitics precede clitics licensed by a matrix verb, as in example (4):

(4) *Pomôž* ***sa*** ***mi*** *obuť.*
help-IMP.SG.2 REFL me-DAT put on the shoes-INF
'Help me put on the shoes.'

One may ask whether the configurations as in example (4) are frequent in Slovak. An initial corpus search for the structure of object control clause with the verb *pomôcť* and climbed reflexive clitic yielded a total of 322 occurrences (5%),

while the search for structure with local placement yielded a total of 5289 occurrences (94%). Corpus data thus show that the given type of structure is relatively infrequent, yet not completely marginal.

Clitic clusters are defined as contact strings of clitics excluding permutation of elements and insertion of non-clitic words. Clitic climbing in subject or object control clauses may give rise to mixed clitic clusters (cf. Kolaković et al. 2022) in which clitics licensed by different matrix VPs occur in adjacent position. However, according to Zimmerling and Kosta (2013), clitic cluster can be formed only by clitics with identical heads which is crucial for distinguishing clitic clusters from occasional word orders like X°— $CL1^X$ | $CL2^Y$ — Y° where two adjacent clitics $CL1^X$ and $CL2^Y$ belong to different syntactic heads X° and Y°.

Slavic languages including Slovak impose grammaticalized constraints on the placement of clitic elements within a clitic cluster. Clitics in clitic clusters are arranged in a rigid order according to language-specific rules called "Clitic Templates" or "Ranking Rules" (Zimmerling – Kosta 2013, p. 179). The internal organization of clitic clusters in Slovak, based on the grammaticalized constraints, can be described as follows.

Table 2. Clitic template of clitic clusters in Slovak

A		B	C						D
Particles		Auxiliary	Pronouns						Connectives
Affirm.	Opt.	Present tense indicative BE-auxiliary	Refl.	Non-argument Dative	Argument Dative	Accusative	Demon.	PPP	Advers.
Že	*By*	*som, si, sme, ste*	*sa, si*	*mi, ti, nám, vám*	*mi, ti, mu, jej, nám, vám, im*	*ma, ťa, ho, ju, nás, vás, ich*	*to, tak, tu, tam*	*s ním, s ňou, k vám*	*však, ale*

Affirm. – Affirmative, Opt. – Optative, Refl. – Reflexive, Demon. – Demonstrative, PPP – prepositional pronoun phrase, Advers. – Adversative

If A°, B° and C° are clusterizing clitics and the fixed order of clitics is [Clitic Phrase A°, B°, C°], no other order like *[Clitic Phrase B°, A°, C°], *[Clitic Phrase C°, A°, B°] should be possible in the canonical position of clusterisation. In accordance with the proposed ordering rules, in constructions with object control verbs reflexive clitics of infinitive Y° can precede the dative clitic of matrix verb X°. This word order pattern X°— $CL1^Y$ | $CL2^X$ — Y° can be occasionally found in the corpus data, not frequently, yet not marginally, e.g.

(5) *Pomohol* ***sa*** ***mu*** *postaviť* *na* *nohy.*
help-PAST-MASC.SG REFL him-DAT stand up-INF on feet-ACC
'I helped him to stand up on his feet.'

The reflexive infinitive clitic *sa* (*obliecť sa*) can even penetrate into the clitic cluster *som jej* licensed by the matrix verb *pomohol*, forming the pattern X°— $CL1^X$ | $CL2^Y$ | $CL3^X$ — Y°:

(6) *Pomohol* ***som*** ***sa*** ***jej*** *obliecť.*
help-PAST-MASC.SG be-PRES.SG.1 REFL her-DAT dress-INF
'I helped her to dress.'

Such linear orderings fully adhere to ordering rules proposed for clitic clusters. On the grounds of the given examples, the following patterns for complex clauses in which both matrix verb and infinitive are cliticized can be sketched:

(i) Non-adjacent placement of clitics conditioned by local placement of infinitive clitic:

(7) *Simona* ***mu*** *pomohla* *vyzliecť* ***sa.***
Simona-NOM him-DAT help- PAST-FEM.SG undress-INF REFL
'Simona helped him to undress.'

(ii) Adjacent placement of clitics which results in mixed clitic clusters X°— $CL1^Y$ $CL2^X$ — Y°; given mixed clusters are either interposed between matrix verb and infinitive (8) or moved in front of matrix verb and infinitive (9):

(8) *Pomohla* ***sa*** ***mu*** *vyzliecť.*
help-PAST-FEM.SG REFL him-DAT undress-INF
'She helped him to undress.'

(9) *Simona* ***sa*** ***mu*** *pomohla* *vyzliecť.*
Simona-NOM REFL him-DAT help- PAST-FEM.SG undress-INF
'Simona helped him to undress.'

(iii) Stacked clitics X°— $CL1^X$ | $CL2^Y$ — Y° which do not form clitic clusters (therefore, ordering rules are not broken in that case); in these cases, the infinitive clitic undergoes partial climbing and its position within the higher clause is disputable – it can be either described as being moved to the third position of the higher clause or as occupying the first position of the infinitive clause as procliticized component:

(10) *Pomôž* ***mi*** ***sa*** *ovládať.*
help-IMP.SG.2 me-DAT REFL control-INF
'Help me to control myself.'

To see whether these constructions are accepted by native speakers, an acceptability judgement survey was conducted. Two types of object control clauses were investigated: with clitic climbing (CC), as in example (11) and with local placement of clitics (LP), as in example (12):

(11) *Kázal* ***sa*** *mi* *vyzliecť.*
order-PAST-MASC.SG REFL me-DAT undress-INF
(12) *Kázal* *mi* *vyzliecť* ***sa.***
order-PAST-MASC.SG me-DAT undress-INF REFL
'He ordered me to undress.'

To prove the frequency of a given structure in the corpus data, a random sample of 100 tokens was drawn from Omnia Slovaca IV Beta and manually annotated. The investigation shows that 70% of annotated examples represent local placement of clitics whereas in 30% of occurrences, the clitics undergo either full (example 13) or partial climbing (example 14) out of infinitive clause:

(13) *Vraj* *si* *mu* ***ich*** *kázala*
apparently be-PRES.SG.2 him-DAT them-ACC order-PAST-FEM.SG
vyhodiť.
throw out-INF
'They say that you ordered him to throw them out.'
(14) *Hneď* *mi* *kázal* ***sa*** *vyzliecť.*
immediately me-DAT order-PAST-MASC.SG REFL undress-INF
'He immediately ordered me to undress.'

The analysis of acceptability ratings brings the following results:

Table 3. Correlations between Age Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC	*0 – 45*: 2.69 *46 – 75*: 2.59 2.64	p = .682	0.08 very small effect	not rejected
With LP	*0 – 45*: 1.37 *46 – 75*: 1.15 1.26	p = .031	0.4 small effect	rejected

Table 4. Correlations between Profession Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC	*L*: 2.48 *NL*: 2.71	p = .404	0.16 very small effect	not rejected
With LP	*L*: 1.18 *NL*: 1.34	p = .127	0.3 small effect	not rejected

Corpus findings can be confronted with survey data on object control clauses. A highly frequent pattern with local placement of clitics is evaluated as highly acceptable by native speakers. On the other hand, the form that is represented only sporadically in the corpus data has a middling rating (2.69 in Group 1 and 2.59 in

Group 2 in the case of depending variable Age, cf. Table 3, and 2.48 in Group 1 and 2.71 in Group 2 in the case of depending variable Profession, cf. Table 4). The results show that high corpus frequency of the pattern correlates with high acceptability rankings, however, rather scarce occurrence of pattern in the corpus data does not automatically lead to low acceptability rankings.

At the same time, the results of statistical analysis show that in the case of pattern with local placement of clitic, the dependent variable Age proves relevant in the case of structure with local placement of clitic component. While the given word order pattern is highly accepted, a sufficient number of respondents from Group 1 consider it acceptable, yet not fully neutral.

There is no statistical difference between Group 1 and Group 2 with respect to Profession as a dependent variable. However, the behaviour of linguists and "naïve" users differs with respect to acceptability span: while the patterns with local placement achieve only the ratings 1 and 2 by linguists, in the non-linguists' group they are rated by full scale from 1 to 5, i.e. the linguists showed significantly greater within-subject consistency than the non-linguists in the given experiment. This proves that fact that linguists may tend to judge strings differently from non-linguists. One possible explanation is that linguists look for reasons behind their acceptance or rejection of a sentence, which takes away spontaneity and makes their judgment processes different from those of naive subjects, who presumably have neither the inclination nor the knowledge necessary to perform this analysis (cf. Schütze 2019, p. 114).

Corpus data on local placement and clitic climbing in object control clauses bring several interesting observations. Clitic climbing can be blocked due to various reasons which are usually described as constraints imposed on clitic components.

Our data confirm the relevance of a constraint labelled as 'Same case, different governors constraint' (cf. Kolaković et al. 2022) which says that clitic climbing might be blocked if two clitics depending on two different matrix predicates have the same case (e.g. in Dative):

(15)	*Kázal*	***mi***	*volať*	***vám***.
	Order-PAST-MASC.SG	me-DAT	call-INF	you-DAT
	'He ordered me to call you.'			
(15a)	**Kázal*	***mi***	***vám***	*volať*.
	Order-PAST-MASC.SG	me-DAT	you-DAT	call-INF

On the other hand, the corpus data do not confirm 'Reflexivity Constraint'. Reflexivity Constraint has been described for cases of multiply embedded infinitive complements. It has been proved that reflexivity of the infinitive that embeds further infinitives plays a crucial role in preventing clitic climbing (cf. Jurkiewicz-Rohrbacher et al. 2017). Apart from stacked infinitives, climbing of reflexive clitics is blocked when the matrix verb has a reflexive counterpart. For example, the

impossibility to move the reflexive clitic into the second position of the matrix verb *nútiť* 'force' is caused by the existence of reflexive verb *nútiť sa* 'force oneself'. The only interpretation of examples like (16) is that the action is understood as being self-oriented (with haplology of reflexive clitics licensed by matrix verb and infinitive):

(16) *Nútila* ***sa*** *zapájať* *do* *rozhovoru.*
force-PAST-FEM.SG REFL join-INF into conversation-GEN
'She forced herself to join into conversation.'

To express extroverted meaning, the only solution is to apply local placement of infinitive clitic:

(16a) *Nútila* *zapájať* ***sa*** *do* *rozhovoru.*
force-PAST-FEM.SG join-INF REFL into conversation-GEN
'She forced (somebody) to join into conversation.'

In structures with the matrix verb *kázať* 'order', reflexive infinitive clitics can climb to a higher clause and occupy the second position. It can be explained on the basis of non-existence of a reflexive counterpart **kázať sa.* Reflexive infinitive clitics can be moved to a higher position as it cannot be confused with reflexive clitics of the matrix verb:

(17) *Na* *recepcii* *som* ***sa*** *kázal*
on reception-LOC be-PRES.SG.1 REFL order-PAST-MASC.SG
zobudiť *o* *šiestej* *večer.*
wake up-INF at six in the evening
'I ordered to wake me up at six o´clock at the reception.'

The data also show the relevance of the so-called 'Person Case Constraint' (Bonet 1991), a universal constraint blocking accusative clitics other than the third person when a dative is inserted in the same clitic cluster. First and second person accusative infinitive clitics tend to remain in situ whereas third person accusative clitics can move to a higher position:

(18) *Kto* *vám* *kázal* *tľapkať* ***ma*** *po*
who you-DAT order-PAST-MASC.SG tap-INF me-ACC on
zadku?
bottom-LOC
'Who told you to pat my bottom?'

(19) *Kázal* *nám* ***ich*** *prispôsobiť* *na*
order-PAST-MASC.SG us-DAT them-ACC adjust-INF on
vašu *postavu.*
your figure-ACC
'He ordered us to adjust them on your figure.'

4.2 Clitic placement in structures with barriers

In certain syntactic contexts, Slovak clitics appear lower than in clausal 2P. It typically occurs due to informational-structural configurations within the sentence when the topic element occupies the second position, thus preceding the clitics.

(20) *Veterinár bol zvyknutý, že jeho pacienti*
Vet-NOM be-PAST-MASC.SG used to that his patients-NOM
sa *bránia.*
REFL defend-PRES-3PL
'The vet was accustomed to the fact that his patients are defending themselves.'

A similar type of placement may occur after particular sentence constituents which function as a kind of syntactic barrier. Such barriers force clitics to be placed closer to the end of the clause than Wackernagel's Law would lead us to expect.

Following Zimmerling and Kosta (2013), Kosta and Zimmerling (2014), a Barrier can be defined as a syntactic category (a lexical head or a phrase) that has an effect on the position of clitics, namely it can change orientation of a clitic towards the clitic host or move a clitic in a given direction in steps to the right/left of the clitic host. Barrier rules are described as mechanisms that trigger delayed placement of clitics or splitting of clusters.

Several types of barriers are distinguished, namely Obligatory vs. Optional, Grammaticalized (occurring with particular lexical heads) vs. Communicative (phrases with a particular communicative status), Blind (relevant for all types of clitic components) vs. Selective (relevant for certain types of clitic components), Cumulative (when two or more Barriers count as a single Barrier) vs. Undoing (when the second Barrier blocks the effect of the first one).

The NP preceding a clitic, and forming first position within the clause, can be maximally complex (for example, a relative or appositive clause can be added to NP), as long as it still forms one constituent. Given that NP with relative or appositive clauses form the first sentence constituent, a climbed clitic component in these cases follows an intonational break, since it is the position after the first sentence constituent.

(21) *Dvom ďalším prítomným ženám, Faith*
two another present women-DAT Faith-DAT
a Lavender, ***sa*** *podarilo*
and Lavender-DAT REFL succeed-PAST-NEUTR.SG
pozbierať niekoľko zhúžvaných papierových obrúskov.
pick-INF several crumped paper squares-GEN
'Two another women, Faith and Lavender, managed to pick several crumpled paper squares.'

However, NP with relative or appositive clauses can move a clitic one step to the right of the clitic host, functioning thus as a syntactic barrier, e.g.

(22) *Ten,* *ktorý* *sa* *zachránil,*
That-NOM who-NOM REFL save-PAST-MASC.SG
stal ***sa*** *vojakom.*
become-PAST-MASC.SG REFL soldier-INSTR
'The one who saved himself became a soldier.'

To test the possibility of clitic third placement in the object control clauses, and the efficiency of syntactic barrier, the following structures were investigated in the survey:

(i) structures with clitic climbing and clitic third placement:

(23) *Júda,* *vodca* *Makabejcov,* *kázal*
Judah-NOM leader-NOM Maccabees-GEN order-PAST-MASC.SG
sa *modliť* *za* *mŕtvych.*
REFL pray-INF for dead-ACC

(ii) structures with clitic climbing and clitic second placement:

(24) *Júda,* *vodca* *Makabejcov,* ***sa*** *kázal*
Judah-NOM leader-NOM Maccabees-GEN REFL order-PAST-MASC.SG
modliť *za* *mŕtvych.*
pray-INF for dead-ACC

(iii) structures with local placement of clitics:

(25) *Júda,* *vodca* *Makabejcov,* *kázal*
Judah-NOM leader-NOM Maccabees-GEN order-PAST-MASC.SG
modliť ***sa*** *za* *mŕtvych.*
pray-INF REFL for dead-ACC
'Judah, the leader of Maccabees, ordered to pray for the dead.'

Corpus data show that object control constructions with complex initial NP constituents followed by attribute or appositive clause are extremely rare in the corpus. Out of 47 occurrences with verbs *kázať/prikázať*, 63% of examples instantiates local placement of infinitive clitic, 37% brings evidence of clitic climbing (16% of clitics appear in third position, 21% of clitics in second position). Overall, 80% of all clitic component occurs in second position, only 20% of examples exhibits clitic third placement.

The analysis of acceptability ratings brings the following results.

Table 5. Correlations between Age Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC and CL	0 – 45: 2.85 46 – 75: 2.77 2.81	p = .726	0.06 very small effect	not rejected

With CC and no CL	0 – 45: 2.66 46 – 75: 2.33 2.49	p = .232	0.22 small effect	not rejected
Without CC	0 – 45: 1.56 46 – 75: 1.66 1.61	p = .586	0.1 very small effect	not rejected

Table 6. Correlations between Profession Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC and CL	L: 2.72 NL: 2.85 2.78	p = .574	0.11 very small effect	not rejected
With LP and no B	L: 2.31 NL: 2.65 2.48	p = .251	0.22 small effect	not rejected
Without CC	L: 1.48 NL: 1.68 1.58	p = 0.329	0.19 very small effect	not rejected

Similarly to complex clauses, it is the local placement of clitics that is evaluated as the most acceptable word order pattern. Clitic climbing in object control clauses with complex NP in the first position is evaluated the less acceptable word order variant. At the same time, the lowest score applies to clitic third placement, i.e. the patterns in which infinitive clitic occupies third position after matrix VP.

Statistical analysis also proves no significant difference in acceptability ratings determined by the investigated variables of Age and Profession. Of particular interest are the acceptability ratings by 16 respondents who did not choose any word order pattern as fully acceptable (yet the range of possible patterns was exhaustive and no other patterns could be applied in this cases), the majority of them being non-linguists. It shows that non-linguists often tend to behave more conservatively, are tougher graders (they rated sentences less grammatical overall). At the same time, the ratings of linguists spanned from 1 to 5. Perhaps it shows that linguists are liable to be unconsciously prejudiced by their own theoretical positions, tending to judge in accordance with the predictions of their particular version of grammar (Schütze 2019, p. 113).

5. DISCUSSION

The idea that empirical evidence for theoretical claims should be gathered from multiple sources has become increasingly important for linguistic research of late. The empirical analysis proves that word order variance in clitic placement is larger than expected. Word order variants are both evidenced in the corpus or elicited with relatively high acceptability scores by native speakers. However, the relation between corpus frequency and acceptability ratings is not always straightforward:

1. Higher frequency in the corpus (>70%) entails a high acceptability rating (< 2): this is the case of local placement of infinitive clitics in object control clauses without barriers (70% : 1.26);

2. A high acceptability rating (< 2) does not entail a higher frequency in the corpus (>70%); see e.g. this is the case of local placement of infinitive clitics in object control clauses with barriers (63% : 1.61);

3. Lower frequency (< 30%) in the corpus does not entail a low acceptability rating (>2.5): this is the case of clitic climbing in object control clauses with barriers with no late placement (21% : 2.49);

4. A low acceptability rating (>2.5) entails low frequency (< 30%) in the corpus: this is the case of clitic climbing in object control clauses with barriers and clitics late placement (16% : 2.81).

The investigation also shows that grammaticality is not a dichotomous notion, and grammatical constructions are not simply environments or non-environments for rules; rather they may be environments to a degree and form hierarchies along which different speakers have different acceptability thresholds (cf. Schütze 2019). Grammaticality itself may be thus understood as a gradient phenomenon representing a function of constraint accumulation, i.e. combinations of different grammatical constraints lead to a range of grammaticality levels (cf. Keller 2000, Sorace – Keller 2005, Wasow 2007, etc.).

It is evident that certain positions of clitics seem to be preferred in particular constructions. As a consequence, scholars may consider the less frequent position to be unacceptable. However, the acceptability ratings show that even less frequent patterns are not rejected by native speakers as completely unacceptable. Therefore, the role of the corpus is to determine the circumstances under which rarely occurring clitic positions can be realized in actual usage. The research thus proves that each method adds to better understanding of the studied phenomenon, thus overcoming the possible shortcomings of methods if used independently.

References

ADAM, Nina (2024): *Optimal Clitic Positioning in Czech and Beyond. Clitic syntax, morphology, and prosody from a constraint-based perspective. Dissertation.* Göttingen: Philosophischen Fakultät der Georg-August-Universität Göttingen.

AVGUSTINOVA, Tania – OLIVA, Karel (1995): On the nature of the Wackernagel position in Czech. In: U. Junghanns – G. Zybatow (eds.): *Formale Slavistik.* Frankfurt am Main: Vervuert, pp. 25–47.

BADER, Markus – HÄUSSLER, Jana (2010): Toward a Model of Grammaticality Judgments. In: *Journal of Linguistics*, Vol. 46, No. 2, pp. 273–330.

BERMEL, Neil – KNITTL, Luděk (2012): Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. In: *Corpus Linguistics and Linguistic Theory*, Vol. 8, No. 2, pp. 241–275.

BOOIJ, Geert E. (2012): *The grammar of words: An introduction to linguistic morphology*. 3rd edition. Oxford: Oxford University Press.

BONET, Eulalia (1991): *Morphology after syntax: Pronominal clitics in Romance languages. PhD diss*. Cambridge, MA: MIT.

DOTLAČIL, Jakub (2004): *The syntax of infinitives in Czech*. *Master's thesis*. Tromsø: Universitetet i Tromsø.

FRANKS, Steven – HOLLOWAY KING, Tracey (2000): *A Handbook of Slavic Clitics.* Oxford: Oxford University Press.

HANA, Jirka (2007): *Czech Clitics in Higher Order Grammar*. Columbus: The Ohio State University.

HASPELMATH, Martin (2023): Types of clitics in the world's languages. In: *LTC [Internet]*, Vol. 3, No. 2, pp. 1–59. Available from: https://typologyatcrossroads.unibo.it/ article/view/16057. [cited 2024 Nov. 23]

IONOVA, Anastasiia (2019): *The unbearable lightness of clitics.* Leiden: Universiteit Leiden.

IVANOVÁ, Martina (to be published): Variations on positioning of clitics in Slovak (on the basis of corpus and experimental data) (unpublished manuscript).

JUNGHANNS, Uwe (2002): Clitic Climbing im Tschechischen. In: U. Junghanns (ed.): *Untersuchungen zur Syntax und Informationsstruktur slavischer Deklarativsätze*. Leipzig: Institut für Linguistik der Universität Leipzig, pp. 57–90.

JURKIEWICZ-ROHRBACHER, Edyta – KOLAKOVIĆ, Zrinka – HANSEN, Björn (2017): Web corpora – the best possible solution for tracking rare phenomena in underresourced languages: clitics in Bosnian, Croatian and Serbian. In: P. Bański et al. (eds.): *Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP) 2017 including the papers from the Web-as-Corpus (WAC-XI) guest section.* Mannheim: Institut für Deutsche Sprache, pp. 49–55.

KELLER, Frank (2000): *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. Ph.D. thesis*. Edinburgh: University of Edinburgh.

KEMPEN, Gerard – HARBUSCH, Karin (2005): The Relationship between Grammaticality Ratings and Corpus Frequencies: A Case Study into Word Order Variability in the Midfield of German Clauses. In: S. Kepser – M. Reis (eds.): *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*. Berlin, New York: Mouton de Gruyter, pp. 329–350.

KOLAKOVIĆ, Zrinka – JURKIEWICZ-ROHRBACHER, Edyta – HANSEN, Björn – FILIPOVIĆ ĐURĐEVIĆ, Dušica – FRITZ, Nataša (2022): *Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian (Open Slavic Linguistics 7)*. Berlin: Language Science Press.

KOSTA, Peter – ZIMMERLING, Anton (2014): Slavic Clitic Systems in a Typological Perspective. In: L. Schürcks – U. Etxeberria – A. Giannakidou – P. Kosta (eds.): *The Structure of Nominal Expression. Slavic and Beyond*. Berlin: De Gruyter, pp. 441–490.

KULIK, Irenäus (2023): Clitic Climbing without Restructuring in Czech and Polish. In: P. Biskup – M. Börner – O. Mueller-Reichau – I. Shcherbina (eds.): *Advances in formal Slavic linguistics 2021. (Open Slavic Linguistics 8)*. Berlin: Language Science Press, pp. 211–232.

PESCARINI, Diego (2021): *Romance Object Clitics. Microvariation and Linguistic Change*. Oxford: Oxford University Press.

REZAC, Milan (2005): The syntax of clitic climbing in Czech. In: L. Heggie – F. Ordóñez (eds.): *Clitic and affix combinations: Theoretical perspectives*. Amsterdam: John Benjamins, pp. 103–140.

RIEMER, Nick (2009): Grammaticality as evidence and as prediction in a Galilean linguistics. In: *Language Sciences*, Vol. 31, No. 5, pp. 612–633.

ROSEN, Alexandr (2001): *A constraint-based approach to dependency syntax applied to some issues of Czech word order. Ph.D. thesis.* Prague: Charles University.

SCHÜTZE, Carson T. (2019): *The empirical base of linguistics: Grammaticality judgments and linguistic methodology (Classics in Linguistics 2).* Berlin: Language Science Press.

SORACE, Antonella – KELLER, Frank (2005): Gradience in linguistic data. In: *Lingua*, Vol. 115, No. 11, pp. 1497–1524.

SPENCER, Nancy J. (1973): Differences between linguists and nonlinguists in intuitions of grammaticality–acceptability. In: *Journal of Psycholinguistic Research*, Vol. 2, No. 2, pp. 83–98.

SPENCER, Andrew – LUÍS, Ana R. (2012): *Clitics: An Introduction.* Cambridge: Cambridge University Press.

SPROUSE, Jon (2007): Continuous acceptability, categorical grammaticality, and experimental syntax. In: *Biolinguistics*, Vol. 1, pp. 118–129.

SPROUSE, Jon (2008): The differential sensitivity of acceptability judgments to processing effects. In: *Linguistic Inquiry*, Vol. 39, No. 4, pp. 686–694.

SPROUSE, Jon (2009): Revisiting satiation: Evidence for an equalization response strategy. In: *Linguistic Inquiry*, Vol. 40, No. 2, pp. 329–341.

VESELOVSKÁ, Ludmila (1995): *Phrasal Movement and X0-Morphology. Word Order Parallels in Czech and English Nominal and Verbal Projections. PhD thesis*. Olomouc: Univerzita Palackého.

WASOW, Thomas (2007): Gradient data and gradient grammars. In: *Chicago Linguistics Society*, Vol. 43, No. 1, pp. 255–271.

ZIMMERLING, Anton W. (2011): Scrambling types in the Slavic languages. In: *Computational linguistics and intellectual technologies*, Vol. 10, No. 17, pp. 750–766.

ZIMMERLING, Anton W. – KOSTA, Peter (2013): Slavic clitics: a typology. In: *Language Typology and Universals. Sprachtypologie und Universalienforschung*, Vol. 66, No. 2, pp. 178–214.

ZWICKY, Arnold M. – PULLUM, Geoffrey K. (1983): Cliticization vs. Inflection: English N´T. In: *Language*, Vol. 59, No. 3, pp. 502–513.

Sources

Omnia Slovaca IV Maior Beta. Bratislava: Jazykovedný ústav Ľ. Štúra SAV, v. v. i. [cited 2024 Nov. 23].

sciendo
DOI 10.2478/jazcas-2024-0046

MULTILINGUALISM IN THE HABSBURG MONARCHY AND SLAVIC LINGUISTIC HERITAGE IN GERMAN IN AUSTRIA

STEFAN MICHAEL NEWERKLA
Department of Slavonic Studies, University of Vienna, Austria

NEWERKLA, Stefan Michael: Multilingualism in the Habsburg monarchy and Slavic linguistic heritage in German in Austria. Jazykovedný časopis (Journal of Linguistics), 2024, Vol. 75, No. 3, pp. 309 – 329.

Abstract: Multilingualism on the current territory of Austria has existed for a long period of time. Roughly since the 6th and 7th centuries, Slavs have settled in Central Europe, including much of present-day Austria. The subsequent expansion of the Magyars, as well as the Bavarianisation of the area, separated the northern and southern Slavs. For the former Habsburg state, we must reckon with eleven main languages in addition to numerous smaller ones. Moreover, already the main languages represented several widely divergent languages: German; two Romance tongues, Italian and Romanian; a range of Slavic languages from all the three branches of that family — western, eastern, and southern; and Hungarian from the Finno-Ugric group. Regarding the multilingual setting in the Habsburg state and its repercussions to this day, this paper outlines the basic assumptions, the methodological toolkit as well as the main general findings of our research projects on language contact in this area which have been running since 2016 as part of the special research programme (SFB) "German in Austria. Variation – Contact – Perception". We conclude with possible implications for further linguistic research in multilingual historical contexts and some links to contemporary phenomena.

Keywords: historical language contact, Slavic languages, German in Austria, theory, methodology, sociolinguistics, historical sociolinguistics, contact linguistics.

1. INTRODUCTION

This paper[1] presents glimpses into my own and my work group's current research on the linguistic situation in Slovakia's neighbour to the west – Austria – with special consideration of the historical and contemporary influence of Slavic languages.

Austria is a relatively small, but linguistically extremely diverse country in Central Europe. This diversity is characterised by "internal" as well as "external multilingualism" (see Wandruszka 1979). The external multilingualism reflects the multilingual tradition in the Austro-Hungarian Empire as well as the linguistic

[1] Written version of the eponymous keynote lecture at the 18th Annual Meeting of the Slavic Linguistics Society – SLS-18 on August 25th, 2023, in the Moyzes Hall, Faculty of Arts of the Comenius University in Bratislava.

consequences of the migration movements in the late 20th and early 21st centuries. This co-existence of different languages and the resulting language contact concur with a pronounced internal multilingualism of German that is still present today. In the light of this linguistic diversity, Austria offers an ideal research laboratory for studies on language variation, contact and change.

In my paper, I will focus on the interplay between internal and external multilingualism and present selected results from the ongoing Special Research Programme "German in Austria. Variation – Contact – Perception" that are particularly interesting from a Slavic perspective such as contact-related explanations for case variation and preposition choice, or coinciding caused motion constructions. For this purpose, I will first briefly introduce the research framework in which we are currently working and then explain what we mean by "internal and external multilingualism" in Austria. The main part of my paper will be devoted to research results and selected case studies, and subsequently, I will conclude with a few words on possible implications for further linguistic research.

2. THE RESEARCH FRAMEWORK

Our work is part of a greater enterprise, namely the Special Research Programme (SFB) "German in Austria. Variation – Contact – Perception" (FWF 60-G23). The currently approved funding period runs from January 1st, 2016, to June 30th, 2026. In other words, this Special Research Programme is funded for a total duration of ten and a half years by the Austrian Science Foundation (Fonds zur Förderung der wissenschaftlichen Forschung | FWF). It is the first major sociolinguistic and variationist project in Austria and consists of five strongly intertwined task clusters. Three of them are devoted to the subject areas "Variation and Change of German in Austria" (Perspectives of Variationist Linguistics), "German and other Languages in Austria" (Perspectives of Language Contact) and "German in the Minds" (Language Attitudes and Perception). The remaining two task clusters are dedicated to the administration ("Coordination") and the processing of the collected data ("Collaborative Online Research Platform").[2] In my paper, I will present results mainly from task cluster C's project part number 06 on "German and Slavic Languages in Austria: Aspects of language contact".[3]

3. INTERNAL AND EXTERNAL MULTILINGUALISM IN AUSTRIA

Internal and external multilingualism in Austria is a translation of Mario Wandruszka's (1979) concept of "innere und äußere Mehrsprachigkeit". It claims

[2] For a detailed description of the entire special research programme "German in Austria. Variation – Contact – Perception" in English, see Budin et al. (2019), in German, see Budin et al. (2018) and Lenz (2018).

[3] For a summary of preliminary results from a contact linguistic perspective, see Newerkla (2022; 2023).

that multilingualism is not limited to speaking several languages, but also extends to the ability of speakers to master several varieties of these languages.

To give you an idea of what this can sound like in Austria, I refer to the transcript of a short conversation recorded as part of the data collection for the Special Research Programme "German in Austria. Variation – Contact – Perception".[4] It is a conversation of two 17-year-old students from the westernmost province of Austria – Vorarlberg. They perform a map task[5] in which one speaker holds a map with a pre-drawn route – and the other has a similar map, but without the route drawn in, and of course he cannot see the first speaker's map. The first speaker (S1) then must explain the route to the second speaker (S2). The part of the route they are talking about is marked with bold arrows on the illustration below.

S1: *denn gohsch diagonal nach links – zu dem Mã*
['Then (you) go diagonally to the left, to the man.']
S2: *okay. Und – gangi an eam vorbei, oder?*
['Okay. And do I go past him, right?']
S1: *ne, staneš ähm u sredini tamo*
['No, you stop, um, there in the middle.']
S2: *dobro*
['Alright.']
S1: *onda ljevo opet*
['Then left again.']
S2: *dobro*
['Alright.']
S1: *jedan centimetar tak*
['One centimetre or so.']
S2: okay
['Okay.']
S1: *onda na gore malo – äh – preko njega*
['Then a bit upwards, er, towards above him.']

[4] The same example was presented by Stephan Elspaß in the unpublished joint keynote lecture with Stefan Michael Newerkla entitled "Austria as a showcase of internal and external multilingualism. Old and new linguistic frontiers" at the 11th International Conference on Language Variation in Europe (ICLaVE|11) on April 11th, 2022, at the University of Vienna, Department of German Studies (https://iclave11.dioe.at/programme/plenaries/ [cit. 25-08-2024]).

[5] Snatched from the website of IDS | Leibniz-Institut für deutsche Sprache, subpage Korpusstruktur "Deutsch heute", section "Interview und Map Task" (http://prowiki.ids-mannheim.de/bin/view/AADG/KorpusTeile [cit. 25-08-2024]).

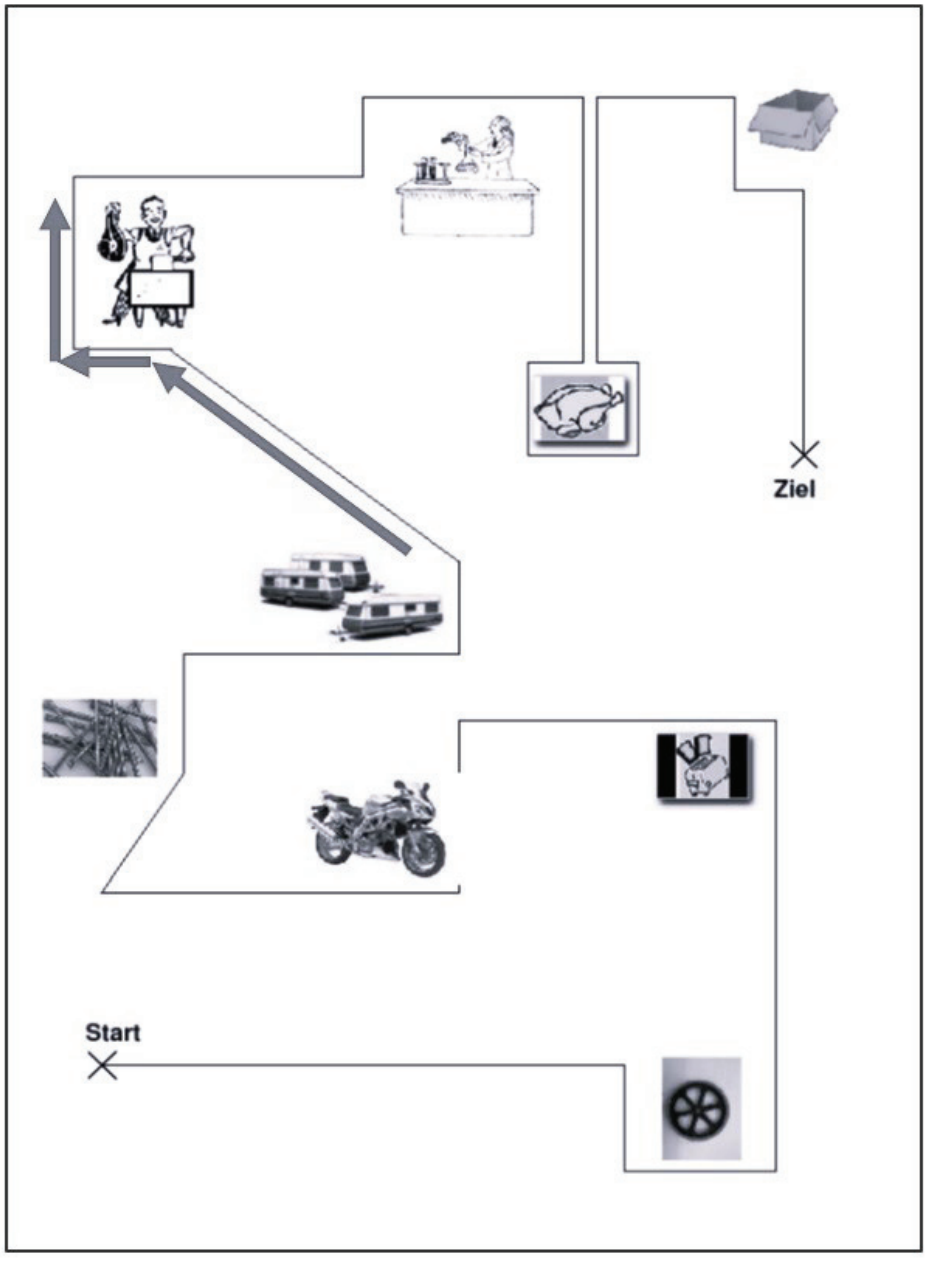


Illustration no. 1: Map task

At first, you may get the impression that both students are speakers of an Alemannic dialect. And they are. But in the third turn, the first speaker switches to an entirely different language, namely Serbian, which appears to be the L1 of the two speakers. So, in addition to a variety of Serbian they speak a dialect of German. And we can assume that they also master the variety of Standard German which is used at their school.

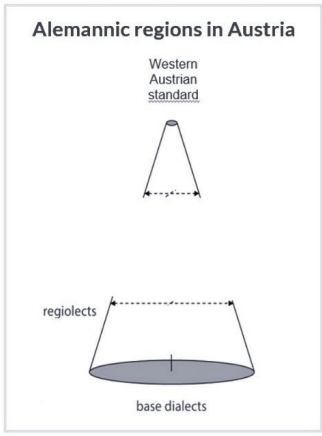
As for internal multilingualism in Austria, most Austrians speak at least one local or regional dialect, and this includes native speakers of such dialects as well as L2 speakers who grew up in Austria. Yet, the dialect landscape in Austria is linguistically extremely diverse. It encompasses Alemannic as well as Bavarian dialects that have existed since the early Middle Ages and, in some regions, have remained virtually unchanged over the centuries. The dialects are often not mutually intelligible. To put things into a bigger perspective, here's a dialect map of High German dialects in Austria.

Illustration no. 2: Dialect map of Austria

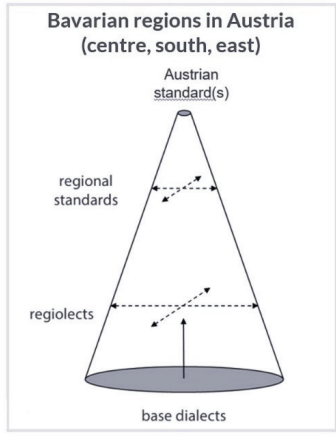
We can clearly see the east-west division, with Alemannic dialect regions in the far west and the much larger Bavarian dialect areas in the centre and the east of the country. The different shadings of darker grey in the west and lighter grey in the rest of the country signify further subdivisions, for instance, into the Central and South Bavarian dialect areas in the eastern part of Austria.

In addition to the geographic variation, we must account for the variation due to specific dialect/standard constellations in Austria. On the one hand, we have a diglossic language situation in the Alemannic regions in the west of Austria (similar to the neighbouring German-speaking parts of Switzerland). On the other hand, the

Bavarian dialect regions are characterised by what can be called "diaglossia" according to Peter Auer's typology of dialect/standard constellations in Europe (see Auer 2005). This typology refers to a situation with "intermediate variants between standard and (base) dialect".



Auer's type B



Auer's type C

Illustration No. 3: Dialect/standard constellations in Austria

However, German in Austria does not only display a high degree of internal variation. We are also confronted with a high intensity of external multilingualism. This has historical reasons. Influences from other languages in Austria have existed for a long period of time. Roughly since the 6th and 7th centuries, Slavs have settled in Central Europe, including much of present-day Austria. The subsequent expansion of the Magyars as well as the Bavarianisation of the research area separated the northern and southern Slavs.

For the former Habsburg state, we must reckon with eleven main languages in addition to numerous smaller ones. Already the main languages represented several widely divergent languages: German; two Romance tongues, Italian and Romanian; a range of Slavic languages from all the three branches of that family — western, eastern, and southern; and Hungarian from the Finno-Ugric group. Moreover, the situation was such that in most of the Habsburg state's Crown lands two, three and more languages were officially in use at the same time.

The multilingual situation has remained intact ever since. In present-day Austria, there are seven historical minority languages, the so-called languages of the six indigenous ethnic groups officially recognised by the Ethnic Groups Act: Burgenland Croatian, Slovene, Czech, Slovak, Hungarian and Romani plus the Austrian Sign Language (ASL). Furthermore, the 20th century brought about significant changes regarding the societal conditions for these and other minority

groups. A shift in the importance of certain groups can be observed: Whereas for example the strong influence of Czech speakers slackened, others – such as speakers of South-Slavic languages, Romanian, but also Hungarian and Slovak, Bulgarian and Italian gained in importance. Other groups, for example the speakers of Turkish and Polish, have remained of relatively stable relevance after a strong increase in the second half of the 20th century. In addition, the consequences of war have recently led to a significant increase in the number of people from Syria, Afghanistan, Ukraine and the Russian Federation.

Citizenship	01.01.2015	01.01.2022	01.04.2022
Foreign nationals together	**1 146 078**	**1 586 709**	**1 642 026**
Germany	170 475	216 731	218 347
Romania	73 374	138 408	140 454
Serbia	114 289	121 613	121 643
Turkey	115 433	117 625	117 944
Bosnia and Herzegovina	92 527	97 347	97 676
Croatia	66 475	95 297	96 759
Hungary	54 939	94 411	94 595
Syria	11 255	68 358	70 901
Poland	54 262	66 090	66 155
Ukraine	8 582	12 673	52 803
Slovakia	32 052	46 707	46 944
Afghanistan	16 779	45 120	45 394
Bulgaria	19 607	35 879	36 335
Italy	22 465	35 696	36 051
Russian Federation	30 032	33 926	34 392

source: STATISTIK AUSTRIA, statistics on population status.

Illustration No. 4: Top 15 foreign nationals in Austria

By far the largest minority groups to date are the migrants from former Yugoslavia (especially Serbia, Bosnia and Hercegovina and Croatia), Germany, Romania, Turkey, followed by migrants from Hungary, Syria and Poland. People from Slovakia are already in ninth place. At present, more than a quarter of the population has a migration background, which means that both parents were born abroad.[6] What also needs to be mentioned in this context is the fact that the largest share of migrants can be found in Vienna. Here, half of all residents have a migration background.[7] However, a similar development has already occurred several times in the history of the Austrian capital, most recently on the threshold from the 19th to the 20th century.

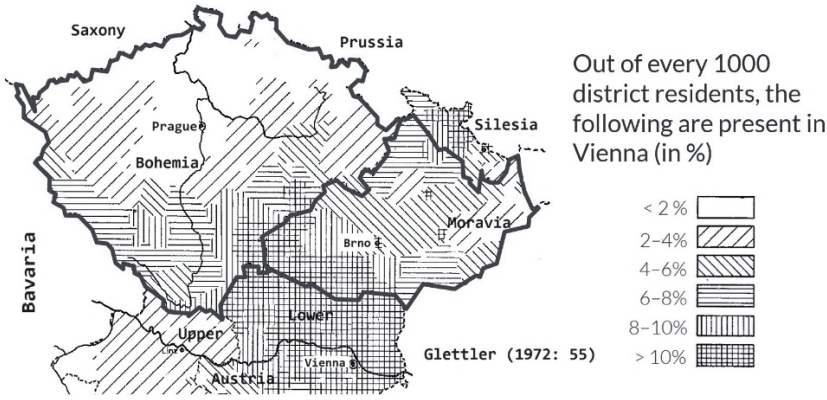


Illustration No. 5: Migration movement to Vienna

After all, the late Habsburg monarchy up to 1918 was shaped by massive socio-demographic changes, particularly in Bohemia and Moravia. This resulted in migration movements to local centres and especially to Vienna, the multilingual capital with then about two million inhabitants. These waves of migration sustainably

[6] In 2022, the overall population in Austria comprised 8,900.800 people, the total population with migration background encompassed 2,351.800 people (26.5%), whereby 1,731.300 belonged to the 1st generation, 620.600 to the 2nd generation. In the meantime, due to migration the Austrian population has grown to 9,179.693 people with reporting date July 1st, 2024 (STATISTICS AUSTRIA, Population statistics. Compiled on 07 August 2024. –Preliminary results for the reporting date 01 July 2024, https://www.statistik.at/en/statistics/population-and-society/population/population-stock/population-at-beginning-of-year/quarter [cit. 25-08-2024]).

[7] In 2022, the overall population in Vienna comprised 1,915.800 people, the total population with migration background encompassed 951.500 people (49.7%). In the meantime, due to migration the Viennese population has grown to 2,018.653 people with reporting date July 1st, 2024 (STATISTICS AUSTRIA, Population statistics. Compiled on 07 August 2024. –Preliminary results for the reporting date 01 July 2024, https://www.statistik.at/en/statistics/population-and-society/population/population-stock/population-at-beginning-of-year/quarter [cit. 25-08-2024]).

influenced the German varieties used in Vienna: German speaking immigrants brought with them language varieties and linguistic features that were shaped by the prolonged stable bilingualism in their places of origin. And speakers of Czech left their linguistic traces when shifting to German within a few years.[8]

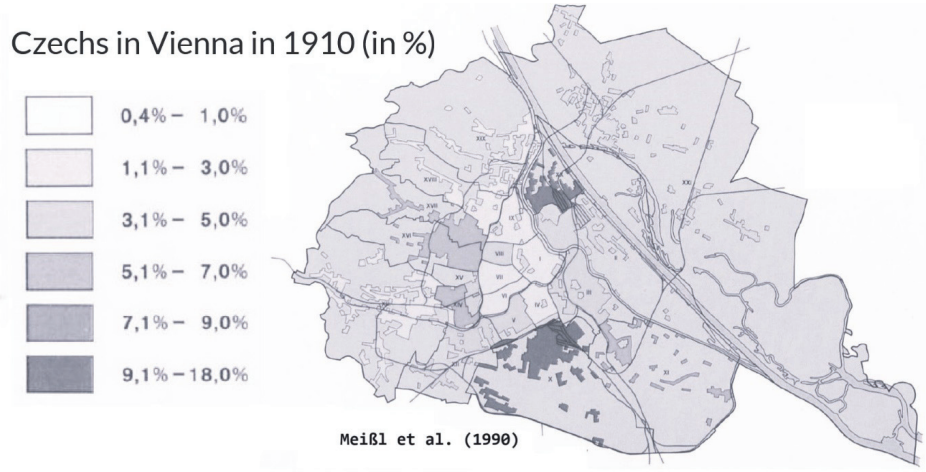


Illustration No. 6: Czechs in Vienna in 1910

It is important to keep in mind that individual and to some extent also societal German-Slavic bilingualism in Austria did not cease to exist after the dissolution of the Austro-Hungarian monarchy in 1918. Nevertheless, the First Austrian Republic declared German its state language – a constitutional law that is still valid today in the same wording. This can be shown quite convincingly by Georg Wenker's questionnaires, which served as the basis for mapping spoken German dialect and involved surveying schoolmasters. These questionnaires not only contained Wenker's well-known sentences to be translated into the respective dialect, but also included sociolinguistic information to be given, such as whether other languages were spoken in the classroom. As can be seen from the data for Lower Austria, even this former Crown land and later supposedly monolingual federal state has always been far from being monolingual.[9]

4. OLD AND NEW LINGUISTIC FRONTIERS – SELECTED CASE STUDIES

Within the framework of our Special Research Programme's task cluster C on language contact, we have analysed several consequences of the contact situation with Slavic languages for the different linguistic levels of German in Austria from

[8] For further details see e.g. Kim (2021) and Newerkla (2013).
[9] Cf. for example Kim (2019).

the last decades of the Habsburg Empire up to the second half of the 20th century. Let us therefore shift now to old and new linguistic frontiers and some selected examples for Slavic influence on German in Austria.

As already mentioned, Vienna and its surroundings have been multilingual not only due to migration movements. A varying degree of German-Slavic bi- and multilingualism has been common in the rural areas east and north-east to Vienna for centuries. Therefore, we can conceive of the whole region as a micro-area of language contact and linguistic convergence within a larger Central European area (Newerkla 2007, 2011, 2020).

To date, a large amount of literature on Slavic-German language contact phenomena in the area has been published. Most studies focus on shared vocabulary. It developed due to the common terminology within the state structure or became manifest as colloquialisms of the Habsburg state's multilingual officials. They shaped the Austrian varieties of German in a typical way, and this led to differences from other varieties of German.

Let us give just a few examples for this phenomenon and at the same time the lexical convergence between the languages of the Habsburg state. On the one hand, there are lexemes still used in the Austrian Standard of German, e.g. *Evidenz* 'public records' (for otherwise "amtliches Register") like in Czech (Cz) *evidence*, Hungarian (Hu) *evidencia*, Slovak (Sk) *evidencia*, Polish (Pl) *ewidencja* or Slovene (Sn) *evidence*; *Malter* 'mortar' (for otherwise "Mörtel") like in Cz *malta*, Hu *malter*, Sk *malta*, Pl *malta*, Sn *malta*; *Matura* 'school-leaving certificate' (for otherwise "Abitur") like in Cz *maturita*, Hu *matura*, Sk *matúra* or *maturita*, Pl *matura*, Sn *matura*; *sekkieren* 'to pester' (for otherwise "belästigen, schikanieren") like in Cz *sekýrovat*, Hu *székal*, Sk *sekírovať*, Pl *sekować*, Sn *sekirati*. On the other hand, there are a lot of shared colloquialisms, e.g. *Fauteuil* 'armchair' (for otherwise "Polstersessel") like in Cz *fotel*, Hu *fotel*, Sk *fotel*, Pl *fotel*, Sn *foltelj*; *fesch* 'dashing, chic' (for otherwise "schick") like in Cz *feš(ný)*, Hu fess, Sk *feš(ný)*, Pl in Silesia *feszny*, Sn *feš*; *Garçonnière* 'one-room flat' (for otherwise "Einzimmerwohnung") like in Cz *garsoniéra*, Hu *garzonlakás*, Sk *garsoniéra*, Pl *garsoniera*, Sn *garsonjera*; *Gat(j)e(hosen)* 'long johns' (for otherwise "lange Unterhose") like in Cz *gatě* or *katě*, Hu *gatya*, Sk *gate*, Pl *gacie*, Sn *gate*. Especially well-known are lexemes related to food, e.g. *Biskotte* 'ladyfinger biscuit' (for otherwise "Löffelbiskuit") like in Cz *piškot(a)*, Hu *piskóta*, Sk *piškóta*, Pl *biszkopt*, Sn *piškot*; *Buchtel* 'yeast roll' (for otherwise "Dampfnudel, Rohrnudel") like in Cz *buchta*, Hu *bukta*, Sk *buchta*, Pl *buchta*, Sn *buhtelj*; *Ribisel* 'currants' (for otherwise "Johannisbeere") like in Cz *rybíz*, Hu *ribiszke*, Sk *ríbezle*, Pl in Silesia *rybi-z/ź-la*, Sn *ribezelj*; *Schnittling* 'chives' (for otherwise "Schnittlauch") like in Cz *šnytlík* or *šnytlink*, Hu *snidling*, Sk *šnitlink* or *šnitling*, Pl in Silesia *sznytloch*, Sn *šnitlink*.[10]

[10] For further examples see Newerkla (2017).

Part of this shared vocabulary is also formed by Slavic loanwords typical of German in Austria. This means that it includes some Austrianisms in the narrow sense. Such lemmas spread from eastern and south-eastern Austria towards the west and comprise both "relic" words and loanwords from the 18th and 19th centuries. Such nationwide Austrianisms are for example *Jause* 'snack' (for otherwise "Brotzeit") from Sn *južina*, or *Feschak* 'dashing guy' (for otherwise "Schönling, gut aussehender Mann, gut gekleideter Mann") as in Cz *ješák*. A lemma confined to the eastern part of Austria is for example *Kukuruz* 'corn' (for otherwise "Mais") as in Serbian *kukuruz*.

As I have said before, much research has already been done on loan words and shared vocabulary. However, comparably minimal systematic and exhaustive linguistic research has been conducted on the grammatical influences and contact phenomena between the Slavic languages (including their varieties) and the Austrian varieties of German. Therefore, within the contact cluster of our SFB German in Austria, we have tried to identify historical and current contact-induced Slavic structural influences on the varieties of German in Austria.

Apart from myself as the project part leader, the current project team consists of post-doc Agnes Kim, financed by the Austrian Science Fund (Fonds zur Förderung der wissenschaftlichen Forschung | FWF), and PhD-student Maria Schinko, financed by own funds from the Rectorate of Vienna University. Another former post-doc was Katharina Prochazka, who was snatched away from basic research by financially more attractive job offers in the private sector. A highlight from her research period were certainly her joint publications with Gero Vogl on language shift in Carinthia in the *Proceedings of the National Academy of Sciences of the United States of America | PNAS* (Prochazka – Vogl 2017)[11] and *Glottotheory* (Prochazka – Vogl 2018). They showed that language shift could be described as a diffusion process in accordance with the physical theory of diffusion: as spread of the dominant language and resulting retreat of the minority language. In their research, they combined a model for language dynamics based on the principles of cellular automata and agent-based modelling with detailed empirical data about language use to describe the dynamics of language shift and thereby identify the driving factors of this specific kind of diffusion. Census data as well as other data about parish and school language were used for testing the model. This made it possible to identify specific factors influencing language shift and to quantify their influence.

In Carinthia, interaction with people who spoke the same language was identified as the most important factor for language maintenance. This includes interaction with people in both the same and surrounding villages. Bilingual schools and parish language did not seem to have a noticeable impact. In fact, bilingual, or

[11] The article also received favourable comments by Anne Kandler and James Steele in the *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* (2017).

so-called "utraquist" schools had even a slightly negative impact on the maintenance of Slovenian.

Another highlight of the whole team was the poster exhibition on historical multilingualism in Austria together with post-doc Katharina Tyran from Vienna University's Department of Slavonic Studies. This exhibition is still available online: https://dioe.at/hist-mehrsprachigkeit [cit. 25-08-2024].

As far as the Special Research Programme's task cluster C is concerned, the main goal of project part 06 on perspectives of German-Slavic language contact has always been to give a comprehensive overview and detailed empirical analysis of contact-induced Slavic influences on the varieties of German in Austria over time. To answer the central research question "To what extent and how has German in Austria been and is still being influenced by contact with Slavic languages?", we facilitate digital corpus data, such as survey data from other projects parts, corpora of present-day German and historical varieties as well as corpora of Czech, Slovak and other Slavic languages, we also utilise qualitative data from linguistic publications such as dictionaries, specialised publications on language contact, linguistic atlases, audio recordings, etc.

So far, we have described a broad range of different possibly contact-induced phenomena of German in Austria on all linguistic levels. For instance, from the field of morphology the borrowing of derivational affixes, or the conjugation with postponed pronouns, from morphosyntax phenomena of congruence and analytical verbal forms, from the field of syntax the use of connectors, the drop of pronouns or verbs, reflexive constructions, or the choice of prepositions, etc.

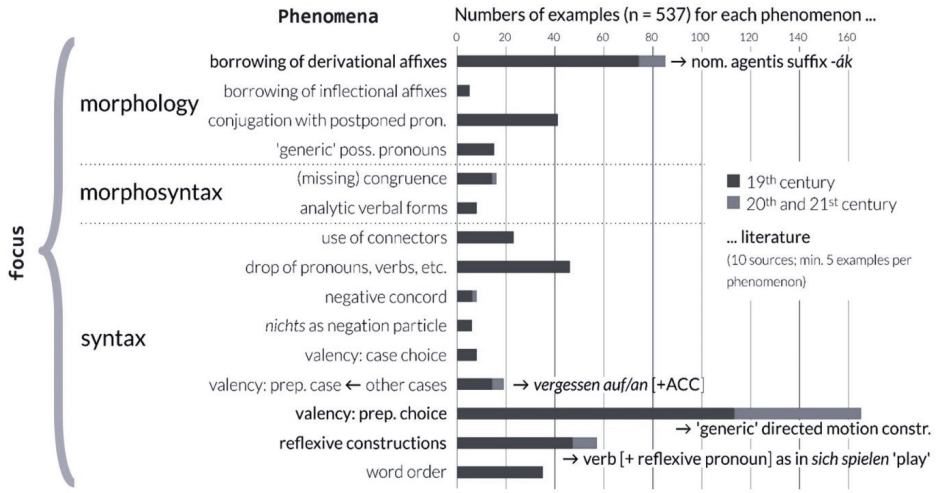


Illustration No. 7: Overview of the phenomena investigated

We analysed each phenomenon according to three groups of criteria: (a) language internal criteria, (b) language external or sociolinguistic criteria, and (c) metalinguistic criteria. Internal criteria (a) took both variationist and contrastive aspects into account. Regarding sociolinguistic criteria (b), we generally reconstructed the relevant historical language contact scenarios. Metalinguistic criteria (c) were applied to adequately reflect the nature of the primary data sources for the corpus.

Our work group assessed the plausibility of each phenomenon's contact explanation and came up with the following threefold results. In some instances, we could confirm the traceability of the contact explanation, as is the case for example in the 'generic' directed motion construction *geben* 'give' as a PUT-verb or the construction *vergessen auf* 'forget' [+acc.]. We will exemplify both these types below.

In other instances, we revealed false beliefs with respect to contact phenomena based on folk etymologies or misinterpretations passed on by generations of experts and non-experts, as is the case with the so-called Viennese e-confusion (in German "Wiener e-Verwirrung"). For this phenomenon, Agnes Kim (2021) could provide evidence for rejecting the claim that the merger of /e/ and /ɛ/ in Viennese into a single phoneme /ɛ/ was induced by contact with Czech. Moreover, the critical discourse analysis of linguistic and popular literature on Slavic influences on German in Austria over time uncovered – not surprisingly – a clear tendency to reproduce language myths.

Ultimately, the detailed results of our research will all be freely accessible in the collaborative online research platform on German in Austria, which is an integral part of the digital research infrastructure of our special research programme.

In the following lines I will present two case studies for investigated contact-related phenomena to provide you with a better impression of our research and its results. Specifically, I will talk about the two afore-mentioned phenomena, for which we could confirm the traceability of the contact explanation.

The first case study features *geben* 'give' as a PUT-verb (cf. Lenz et al. 2020). Normally, in Standard German, we use different verbs to describe situations such as putting the book into a bag, on a shelf, or onto a table, e.g. in German *stecken*, *stellen*, *legen*. Whereas in colloquial Austrian German, the universal PUT-verb is usually *tun* 'do', we can find the use of *geben* 'give' as a PUT-verb in eastern Austrian dialects of German, especially in Vienna and its agglomeration area. We can conclude that – on the one hand – from comprehensive historical data of German varieties such as from Georg Wenker's questionnaires. On the other hand, we can infer this from questionnaires of the project Syntax of contemporary Bavarian | SynBai (cf. Lenz et al. 2015).

Subsequently, we looked for evidence for the central hypothesis that this phenomenon traces back to language contact with Czech as already suggested by various scholars in the 19th century, for example also August Schleicher (1851, p. 41)

who claimed that *geben* ‘give’ for German *setzen*, *legen*, *stellen*, *stecken* ‘put’, etc. behaves as in Czech, e.g. Austrian German *gib es auf den Tisch*, *in die Tasche* = Cz *dej to na stůl*, *do kapsy.*

By making use of corpora of Slavic languages, we were able to confirm this assumption to a reliable degree of certainty. We find it also supported by its high frequency in both formal and informal Czech written texts and by the fact that Czech *dát* ‘give’ in PUT-function has been accounted for since the Old Czech period. Examples from this time illustrate that Old Czech *dáti* ‘give’ was already used in various PUT-contexts at that time, e.g. in the context of ‘cause to sit’; or in the context of ‘cause to lie’. Such constructions cannot be found in Church Slavonic. However, the equivalent of Old Czech *dáti* ‘give’ in Latin *dāre* ‘give’, seems to have had PUT-semantics, too. A diachronic analysis of the development of different semantics and functions of equivalents of ‘give’ in Central European languages and their varieties would therefore also have to consider a possible influence of Medieval Latin. In any case, our data analyses show that *geben* ‘give’ as a PUT-verb has been and is still areally distributed along and spreading from the contact area of Czech and Eastern Austrian varieties of German. And it is also documented for the neighbouring, formerly German-speaking areas in Czechia and Slovakia.

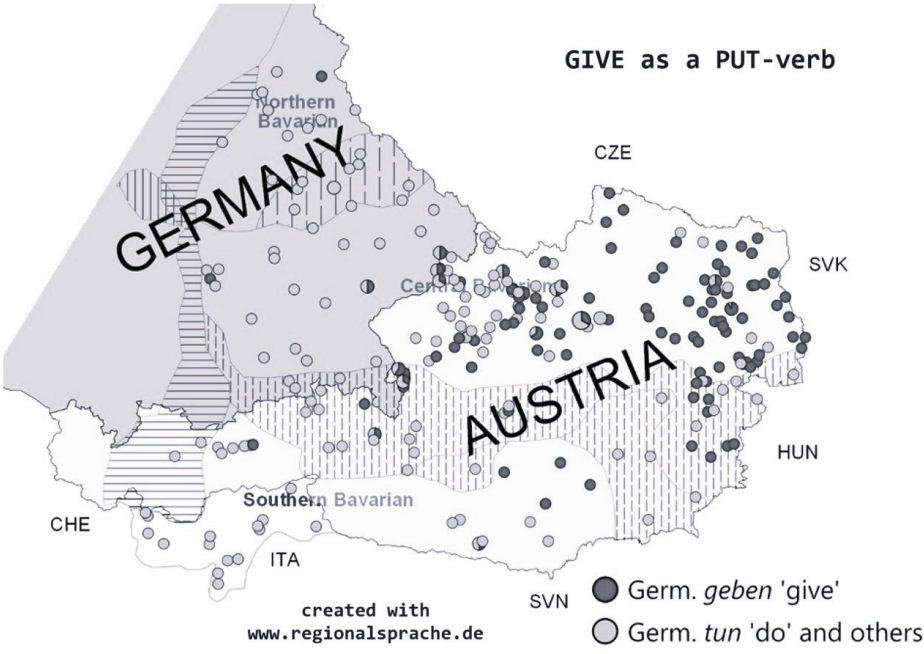


Illustration No. 8: GIVE as a PUT-verb

For details of our research and its results, consult the joint article by Alexandra Lenz, Fabian Fleißner, Agnes Kim, and Stefan Michael Newerkla in the *Journal of Linguistic Geography* (Lenz et al. 2020).

The second case study is on preposition choice (Kim et al. 2020). Prepositional phrases are among the most cited alleged Slavic contact phenomena for German in Austria. However, we also realized the lack of a systematic overview or investigation into these phenomena. Therefore, we have had to assess each case individually. Again, various scholars already from the 19th century suggested language contact phenomena in parallel constructions, for example Hugo Schuchardt (1884, p. 115) who claimed that in hardly any other domain, the foreigner would make more mistakes than in the domain of prepositions, and it would be this domain the native speakers tended to get infected easily, e.g. *auf* 'on' would be the favourite preposition of the German speaking Slavs. And he was right. Several examples of parallel prepositional phrases in colloquial Austrian or Viennese German and Czech most probably result from the language shift from Czech to German or at least have been supported by it, e.g. *auf Urlaub fahren* (for otherwise "in Urlaub fahren") as in Cz *jet na dovolenou* 'go on holiday', *auf zwei Tage* (for otherwise "für zwei Tage") as in Cz *na dva dny* 'for two days', *auf jemanden/etwas denken* (for otherwise "an jemanden/etwas denken") as in Cz *myslet na někoho/něco* 'think of sb./sthg.', *Vorbereitungen auf etwas* (for otherwise "Vorbereitungen für/zu etwas") as in Cz *přípravy na něco* 'preparations for sthg.', *in der Nacht auf* (for otherwise "in der Nacht zu") as in Cz *v noci na* 'in the night to', or *sich auf jemanden/etwas erinnern* (for otherwise "jemanden/etwas erinnern") as in Cz *vzpomenout si na někoho/něco* 'remember sb./sthg.'.

For this paper, we are going to look in detail at the areal variation of the German verb *vergessen* 'to forget' in spoken and written standard registers. While the German standard language exclusively recognises constructions with *vergessen* and a direct argument in accusative, Austrian Standard German accepts constructions with a prepositional argument including the preposition *auf* 'on', too. Already since the 19th century scholars have pointed out a similar grammatical variation in case government for the Czech equivalent *zapomínat/zapomenout* 'to forget', considering the situation in Austrian German to reflect Czech influence.

Our research questions in this context are: Is the construction *vergessen* [*auf* + acc.] restricted to German in Austria? Do the patterns of case variation of the equivalents of 'to forget' correspond in Czech and Austrian varieties of German? And is a contact explanation possible and plausible?

In Austria, we are confronted with a high degree of variation in case government. Apart from examples with a direct argument in accusative, we find a lot of constructions with a prepositional argument including the preposition *auf*. During the second half of the 19th century, we even find constructions with a prepositional argument including the preposition *an*. But is the construction *vergessen* [*auf* + acc.] restricted to German in Austria? As we can see from the Word Atlas of German

Colloquial Languages (Eichhoff 1993, Vol. 3, p. 59), we must answer this question in the affirmative. The construction *vergessen* [*auf* + acc.] is significantly more frequent and almost solely restricted to German varieties in Austria.

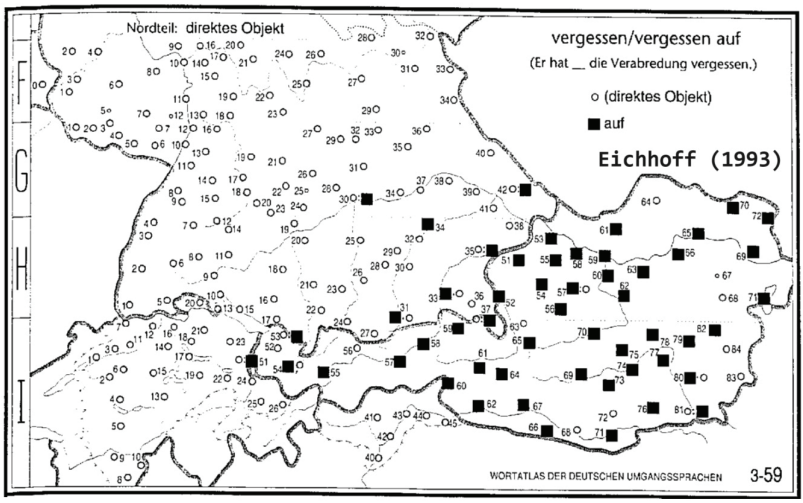


Illustration No. 9: Map 3-59 from the Word Atlas of German Colloquial Languages

This fact is impressively confirmed for the written standard by the Variant Grammar of German (*Variantengrammatik* 2018), as you can see under the keyword *vergessen* 'to forget' (http://mediawiki.ids-mannheim.de/VarGra/index.php/Vergessen [cit. 25-08-2024].

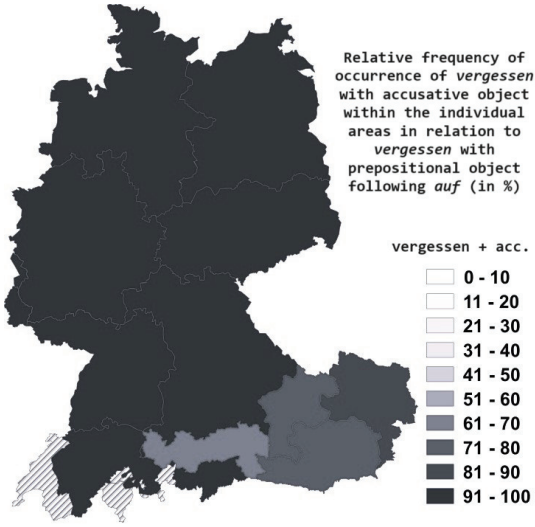


Illustration No. 10: *vergessen* in the Variant Grammar of German

Accordingly, by using contemporary German corpora composed of journalistic texts from Austria, Germany, Switzerland and Liechtenstein, we can also demonstrate that the construction with the preposition *auf* 'on' occurs significantly more frequently in texts from Austria.

Deutsches Referenzkorpus DeReKo – German Reference Corpus | Newspapers[12]
Newspapers and Magazines • 2010–2015 • Germany, Austria, Switzerland, Liechtenstein
size: 715.338.107 tokens • query: *vergessen auf* + acc. • hits: 2164

hits	relative frequency	texts	country
1539	6.323 per million words	1506	Austria
363	1.566 per million words	340	Germany
44	1.536 per million words	42	Liechtenstein
218	1.028 per million words	205	Switzerland

Moreover, we contrastively determined the relations between the two variants of case government and the meaning of the verb in particular sentences in Czech and German in Austria. The analysis of corpora of contemporary journalistic texts from Austria and the Czech Republic shows that the constructions with the prepositional object occur considerably more often with the same meaning of the verb in both languages. Of course, we also investigated into the historical development of the constructions in both languages making use of the DIAKORP corpus of historical Czech, the Old-Czech text bank, the Lexical Database of Humanistic and Baroque Czech, the Oxford GerManC Corpus, the Mannheim Corpus of Historical Newspapers and Periodicals, the Austrian Baroque Corpus AbaC:us, and the Austrian ANNO corpus of Historical Newspapers and Journals.

Having done so, we can determine whether the situation in German in Austria may be attributed to historical language contact. By considering the regional distribution of the constructions, the variation pattern in German and Czech in contrast and its diachronic development, we can eventually conclude that the contact explanation is plausible. For details of this research and its results, consult the joint article by Agnes Kim, Sebastian Scharf and Ivan Šimko in the openly accessible anthology *Areal Convergence in Eastern Central European Languages and Beyond* (Kim et al. 2020).

As already mentioned, task cluster C's project part 06 of our Special Research Programme on German in Austria has so far focused on certain prepositional arguments, e.g. with the verb *vergessen*. Adverbials do not play a big role, because they are hardly found amongst the alleged contact phenomena. There are only a few temporal ones that require closer research. Many phenomena can, however, be attributed to local or directional arguments. By facilitating field data of other project parts, we found evidence for the over-representation of *auf* in Austrian variants of German and hints at a generic motion construction. For this purpose, we analysed

[12] See https://www.ids-mannheim.de/digspra/kl/projekte/korpora/ [cit. 25-08-2024].

conversations among friends from three locations in the Bavarian speaking area of Austria by utilising data from the joint corpus of project part 03 and project part 08. We complemented them with a corpus of sound recordings of adult- and children-directed speech in Vienna. In only 184 out of more than 4,700 prepositional phrases another preposition was chosen than common in the German standard. There are just a few local or directional arguments in which *auf* 'on' was chosen over any other preposition. This pattern occurs in three locations from the Central Bavarian and the transition area. In these varieties, *auf* is slightly over-represented in comparison to Standard German. In South Bavarian dialects, however, another pattern can be observed, namely preposition (and article) drop. Our results show that preposition drop is frequent throughout the South Bavarian area, with its core region in Carinthia. For details of this research and its results, consult the joint article by Agnes Kim and Katharina Korecky-Kröll in the online journal Open Linguistics (Kim – Korecky-Kröll 2021).

5. CONCLUSION

In this paper, I have tried to show that it is precisely the cross-linguistic comparison of historical linguistic research with synchronic linguistic research which represents an added value that has not been fully exploited yet in variationist and contact linguistics, and this despite its potential to provide a deeper understanding of ongoing linguistic processes in complex multilingual societies. By analysing contemporary language use against its historical background, we can for example shed light on how, in the context of the other languages, a certain variety was and is used and valorised as an instrument of social interaction and as a reference point for cultural construction.

In the Slavic-German context, this is especially rewarding in view of the situation in Austria, which is a showcase of internal and external multilingualism – in the past as well as in the present. There are several reasons for that.

First, present-day Austria offers a multitude of varieties of German that co-exist and have co-existed with a multitude of varieties of other – mostly Slavic – languages, including heritage as well as migrant languages.

Second, there is a range of linguistic features that varieties of German in Austria share with non-German – mostly Slavic – languages in Austria and in the neighbouring countries, such as individual lexical items, GIVE as PUT verb constructions and corresponding preposition choice. These shared features have come into being within traditional contiguous dialect areas as the result of centuries-old and intensive contact with non-German languages, especially in the Austro-Hungarian Empire and particularly in the metropolitan centre of Vienna.

And finally, the specific dialect/standard constellations in Austria are the perfect breeding ground for areal variation in German – not only on the linguistic surface, but also regarding the underlying selection criteria for grammatical constructions in language contact situations with Slavic varieties, for example with regard to preposition choice.

We hope that international research on historical multilingualism will not only profit from such findings and results, but will allow for the identification of comparable, distinct and universally applicable aspects of language contact in the areas under investigation. Thus, we can unveil the way different ethnic groups experience the use of a certain variety – mediated through the multiple linguacultural practices – in their everyday life.

References

AUER, Peter (2005): Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In: N. Delbecque – J. van der Auwera – D. Geeraerts (eds.): *Perspectives on Variation: Sociolinguistic, Historical, Comparative*. Berlin – New York: De Gruyter Mouton, pp. 7–42. DOI 10.1515/9783110909579.7. [cit. 25-08-2024].

BUDIN, Gerhard – ELSPAß, Stephan – LENZ, Alexandra N. – NEWERKLA, Stefan Michael – ZIEGLER, Arne (2018): Der Spezialforschungsbereich „Deutsch in Österreich (DiÖ). Variation – Kontakt – Perzeption". The Special Research Programme "German in Austria (DiÖ). Variation – Contact – Perception". In: *Zeitschrift für germanistische Linguistik. Deutsche Sprache in Gegenwart und Geschichte*, Vol. 46, No. 2, pp. 300–308. DOI 10.1515/zgl-2018-0017. [cit. 25-08-2024].

BUDIN, Gerhard – ELSPAß, Stephan – LENZ, Alexandra N. – NEWERKLA, Stefan Michael – ZIEGLER, Arne (2019): The Research Project (SFB) 'German in Austria'. Variation – Contact – Perception. In: L. Bülow – A.-K. Fischer – K. Herbert (eds.): *Dimensionen des sprachlichen Raumes. Variation – Mehrsprachigkeit – Konzeptualisierung*. Berlin: Peter Lang, pp. 7–35.

EICHHOFF, Jürgen (1993): *Wortatlas der deutschen Umgangssprachen. Volume III*. München – New Providence – London – Paris – Bern: K. G. Saur.

GLETTLER, Monika (1972): *Die Wiener Tschechen um 1900. Strukturanalyse einer nationalen Minderheit in der Großstadt* (= Veröffentlichungen des Collegium Carolinum, Vol. XXVIII). München: R. Oldenbourg.

KANDLER, Anne – STEELE, James (2017): Modeling language shift. In: *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, Vol. 114, No. 19, pp. 4851–4853. DOI 10.1073/pnas.1703509114 [cit. 25-08-2024].

KIM, Agnes (2019): Multilingual Lower Austria. Historical sociolinguistic investigation on Wenker's questionnaires. In: L. Bülow – A.-K. Fischer – K. Herbert (eds.): *Dimensionen des sprachlichen Raumes. Variation – Mehrsprachigkeit – Konzeptualisierung.* Berlin: Peter Lang, pp. 187–211.

KIM, Agnes (2021): The Melting Pot Revisited: Historical Sociolinguistic Perspectives on Migration and Language Contact in Vienna. In: A. Auer – J. Thorburn (eds.): *Approaches to Migration, Language and Identity*. Oxford et al.: Peter Lang, pp. 11–40.

KIM, Agnes – KORECKY-KRÖLL, Katharina (2021): Prepositional phrases in German in Austria – identifying patterns of variation. In: *Open Linguistics*, Vol. 7, No. 1, pp. 476–510. DOI 10.1515/opli-2021-0024. [cit. 25-08-2024].

KIM, Agnes – SCHARF, Sebastian – ŠIMKO, Ivan (2020): Variation in Case Government of the Equivalent for the Cognitive Verb to forget in German in Austria and Czech. In:

L. Szucsich – A. Kim – U. Yazhinova (eds.): *Areal Convergence in Eastern Central European Languages and Beyond.* Berlin: Peter Lang, pp. 139–175.

LENZ, Alexandra N. (2018): The Special Research Programme "German in Austria. Variation – Contact – Perception". In: U. Ammon – M. Costa (eds.): *Sprachwahl im Tourismus – mit Schwerpunkt Europa. Language Choice in Tourism – Focus on Europe. Choix de langues dans le tourisme – focus sur l'Europe. Sociolinguistica*, Vol. 32. Berlin – Boston: de Gruyter, pp. 269–277.

LENZ, Alexandra N. – AHLERS, Timo – WERNER, Martina (2015): Zur Dynamik bairischer Dialektsyntax. Eine Pilotstudie. In: *Zeitschrift für Dialektologie und Linguistik*, Vol. 81, No. 1, pp. 1–33.

LENZ, Alexandra N. – FLEIßNER, Fabian – KIM, Agnes – NEWERKLA, Stefan Michael (2020): Give as a put verb in German – A case of German-Czech language contact? In: *Journal of Linguistic Geography*, Vol. 8, No. 2, pp. 67–81. DOI 10.1017/jlg.2020.6 [cit. 25-08-2024].

MEIßL, Gerhard – KOPECKY, Erich – SWOBODA, Manfred (1990): Sprachliche Gliederung der Bevölkerung 1880–1939. Anteil der Personen mit tschechischer bzw. nichtdeutscher Umgangssprache an der Bevölkerung (= map 3.1.5/2). In: F. Czeike (ed.): *Historischer Atlas von Wien. 4. Lieferung*. Wien – München: Jugend und Volk. Online: https://www.wien.gv.at/actaproweb2/benutzung/image.xhtml?id=d2803cf4e35b7ac43647500cce9224e9 [cit. 25-08-2024].

NEWERKLA, Stefan Michael (2007): Areály jazykového kontaktu ve střední Evropě a německočeský mikroareál ve východním Rakousku. In: *Slovo a slovesnost*, Vol. 68, No. 4, pp. 271–286.

NEWERKLA, Stefan Michael (2011): Stopy vzájomného ovplyvňovania a prelínania jazykov a kultúr na východe Rakúska s osobitným zreteľom na slovenčinu. In: O. Orgoňová (ed.): *Jazyk a komunikácia v súvislostiach III*. Bratislava: Univerzita Komenského, pp. 203–214.

NEWERKLA, Stefan Michael (2013): Linguistic Consequences of Slavic Migration to Vienna in the 19th and 20th Centuries. In: M. Moser – M. Polinsky (eds.): *Slavic Languages in Migration*. Berlin – Wien: LIT-Verlag, pp. 247–260.

NEWERKLA, Stefan Michael (2017): Kontaktareale in Mitteleuropa am Beispiel Altösterreich. In: Ch. Mauerer (ed.): *Mehrsprachigkeit in Mittel-, Ost- und Südosteuropa. Gewachsene historische Vielfalt oder belastendes Erbe. Beiträge zur 1. Jahrestagung des Forschungszentrums Deutsch in Mittel-, Ost- und Südosteuropa, Regensburg, 2.–4. Oktober 2014* (Forschungen zur deutschen Sprache in Mittel-, Ost- und Südosteuropa FzDiMOS Band 4). Regensburg: Friedrich Pustet, pp. 17–32.

NEWERKLA, Stefan Michael (2020): Linguistic areas in East-Central Europe as the result of pluridimensional, polycentric convergence phenomena. In: L. Szucsich – A. Kim – U. Yazhinova (eds.): *Areal Convergence in Eastern Central European Languages and Beyond.* Berlin: Peter Lang, pp. 207–228.

NEWERKLA, Stefan Michael (2022): Reconstructing multilingualism in the Habsburg state: lessons learnt and implications for historical sociolinguistics. In: *Sociolinguistica. European Journal of Sociolinguistics*, Vol. 36, No. 1–2, pp. 151–167. DOI 10.1515/soci-2022-0014. [cit. 25-08-2024].

NEWERKLA, Stefan Michael (2023): Reconstructing historical language contact between Slavic languages and Austrian varieties of German: theoretical assumptions, methodological approaches and general results. In: *Jazykovedný časopis*, Vol. 74, No. 2, pp. 645–660. DOI 10.2478/jazcas-2024-0016 [cit. 25-08-2024].

PROCHAZKA, Katharina – VOGL, Gero (2017): Quantifying the driving factors for language shift in a bilingual region. In: *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 114, No. 17, pp. 4365–4369. DOI 10.1073/pnas.1617252114 [cit. 25-08-2024].

PROCHAZKA, Katharina – VOGL, Gero (2018): Are languages like atoms? On modelling language spread as a physicist. In: *Glottotheory*, Vol. 9, No. 1, pp. 77–88. DOI 10.1515/glot-2018-0005 [cit. 25-08-2024].

SCHLEICHER, August (1851): Über die wechselseitige Einwirkung von Böhmisch und Deutsch. In: *Archiv für das Studium der neueren Sprachen und Literaturen*, Vol. 9, pp. 38–42.

SCHUCHARDT, Hugo (1884): *Dem Herrn Franz von Miklosich zum 20. November 1883. Slawo-deutsches und Slawo-italienisches.* Graz: Leuschner und Lubensky.

Variantengrammatik des Standarddeutschen (2018): *Ein Online-Nachschlagewerk.* Verfasst von einem Autorenteam unter der Leitung von Christa Dürscheid, Stephan Elspaß und Arne Ziegler. Openly accessible publication: http://mediawiki.ids-mannheim.de/VarGra [cit. 25-08-2024].

WANDRUSZKA, Mario (1979): *Die Mehrsprachigkeit des Menschen*. München: Piper.

sciendo
DOI 10.2478/jazcas-2024-0047

IN SEARCH OF PATTERNS OF HISTORICAL LANGUAGE VARIATION AND USER INTERACTION (OR: WHO USED WHAT LINGUISTIC FEATURES WITH WHOM, WHEN, WHERE, WHY, AND HOW?)

MARK RICHARD LAUERSDORF
Department of Linguistics, College of Arts and Sciences, University of Kentucky, USA

LAUERSDORF, Mark Richard: In search of patterns of historical language variation and user interaction (or: Who used what linguistic features with whom, when, where, why and how?). Jazykovedný časopis (Journal of Linguistics), 2024, Vol. 75, No. 3, pp. 330 – 347.

Abstract: This essay presents a set of interconnected thoughts, considerations, and conceptualizations that address what it is that we are talking about when we discuss language variation and the interaction of language users in historical contexts, and how we research patterns of historical language variation and user interaction.

Keywords: data-driven, inductive analysis, language and identity, historical language-documentation fieldwork, historical ethnographic fieldwork, tasks of historical sociolinguistics.

1. POSITION STATEMENT 1

The arguments presented in this essay grow out of:

• work in historical situations that involve a high number of (closely-related) language varieties in contact in settings that exhibit a multitude of geophysical, socio-cultural, political, and psychological borders, with ever changing socio-cultural, socio-political, socio-economic contexts, and often no prominent "roofing" standard language or prestige variety, and

• a strong preference for methods that include digital corpora, analytic data visualization, and statistical data analysis, in overarchingly data-driven, inductive investigations (a methodological preference that will figure prominently throughout this discussion).

In other words, the arguments presented here seek to make a case for the necessity of **data-driven corpus-based quantitative investigation of language variation in its complex socio-political and socio-cultural environments, using statistical and visualization methods of data analysis to identify salient patterns.**

One might be tempted to react to this position statement with a dismissive "so what?", wondering whether there is anything new to be said here. After all, wasn't

Barbara Horvath already doing exactly this kind of work in the 1980s in her work in Sydney, Australia, where the rationale behind the methods she employed in that research program was stated as follows (with the parallels to my above-stated position inserted in *[square-bracketed italics]*)?

> Quantitative analyses of large data sets make use of both linguistic and sociological categories in sociolinguistic studies. *[= quantitative; language in its social environment]* [...] The familiar problem of grouping speakers by such sociological characteristics prior to quantitative analysis is addressed *[= data-driven, not category-driven]* and an alternative solution – principal components analysis – is suggested. Principal components analysis is used here as a heuristic for grouping speakers solely on the basis of linguistic behaviour... *[= data-driven; statistical and visualization data analysis]* In addition, by naming the principal components, the major linguistic and social dimensions of the variation in the data can be identified. *[= identify salient linguistic and social patterns]* (Horvath – Sankoff 1987, p. 179)

There are, for our discussion here, two important observations in this regard:
Observation 1: Despite the earlier pioneering work of scholars like Horvath, the data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning has only recently been more widely taken up (often under the label of "coherence"). See, for example, Meyerhoff and Klaere (2017), the collection of studies in Hinskens and Guy (2016), or the more recent studies in Beaman and Guy (2022), where Cerruti and Vietti (2022) note that:

> there is growing interest in the empirical characterization of aggregates of linguistic variables (Guy and Hinskens 2016; Ghyselen and De Vogelaer 2018; Vietti 2019). [In note 3 here Cerruti and Vietti mention that: "Seminal studies include Guy (1980), Horvath and Sankoff (1987), and Trumper and Maddalon (1990)." – MRL] The co-occurrence of linguistic features in relation to social factors has been framed within the concept of coherence. Coherence at the level of empirical observation has been mainly understood as "external" or as co-occurrence between linguistic facts in relation to socio-communicative contexts (Cerruti – Vietti 2022, p. 263).

Thus, the position that I am promoting here, though gaining in application, is not yet a dominant research paradigm in the study of language variation[1], and this

[1] There is, however, steady refinement of the concepts and the methods, including in the direction of studying much larger datasets (see for example Hua et al. 2021), which is an important methodological step especially as regards the further arguments in this essay concerning the need to "use all the data".

essay is intended to lend another voice to the call for the data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning.

Observation 2: Such work has, for a variety of reasons, been undertaken nearly exclusively in modern synchronic situations of language variation, whereas the discussion here focuses on language variation and the interaction of language users *in historical contexts*.

Thus, this essay is expressly intended to be a call for *extending to the field of historical sociolinguistics* the growing application (in present-day synchronic contexts) of data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning.

2. POSITION STATEMENT 2

In promoting this type of data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning specifically *in historical contexts*, another position needs to be propounded regarding the focus of research into historical language variation and user interaction. The traditional research paradigm in the investigation of language history was for a long time:

language history = standard language history = history of national identity

However, with the increasing growth and development of socio-historical emphases in historical linguistic work (i.e., the steady rise of the field of *historical sociolinguistics*), there has been strong movement away from this more narrow focus on the history of the standard language and national (nation-state) identity, and an incorporation of some basic tenets from social history, a field that arose in the mid 20th century bringing new perspectives to the study of the past that were no longer about writing "biographies of great men", but rather about relating the experiences of ordinary people. Indeed, Tilly (1985) identifies the "tasks of social history" as:

1) "documenting large structural changes";
2) "reconstructing the experiences of ordinary people in the course of those changes";
3) "connecting the two" (Tilly 1985, p. 31).

A historical sociolinguist's paraphrase of Tilly's tasks, tailoring them as tasks of the social history *of language* (i.e., "tasks of historical sociolinguistics"), might look something like this:

1) documenting structural language variation and change;

2) reconstructing the experiences and interactions of ordinary people in connection with that variation and change; and
3) connecting the two (a key element for the discussion in this essay).

And this, in turn, suggests a recasting of the traditional research paradigm in the investigation of language history that was listed above, leading to a rendering more like this:

language history =	history of variation and change in all varieties of language in play in user interactions in a given socio-cultural context	=	a multifaceted approach to questions of history of language and identity

This view of the investigation of language history places the focus squarely on describing the complexities of the variable and interactional historical contexts, and the individual actors in those contexts, that produce present-day states (that are themselves filled with variation and interaction among individual actors), and it strengthens the case for the research program proposed at the outset of this essay:

> **the necessity of data-driven corpus-based quantitative *sociolinguistic* investigation of *historical* language variation *and user interaction* in *their* complex socio-political and socio-cultural environments, using statistical and visualization methods of data analysis to identify *and correlate* salient patterns *in the linguistic and socio-historical data***

with the additional elements in italics now sharpening the program's focus on historical investigation not just of language but also user interaction, and on the interconnection between the linguistic and social aspects and the patterning of both. Paraphrasing Joshua Fishman's (1965) classic article title, this research program could be described as the investigation of: Who used what linguistic features with whom, when, where, why, and how?

3. WHO USED WHAT LINGUISTIC FEATURES WITH WHOM, WHEN, WHERE, WHY, AND HOW?

Seeking, in his 1965 article, to formalize the description of language choice in "within-group (or intragroup) multilingualism", Fishman states: "habitual language choice is far from being a random matter of momentary inclination, even under those circumstances when it could very well function as such from a purely probabilistic point of view" (Fishman 1965, p. 67). In examining Fishman's statement beyond the context in which he was working, Fishman's "multilingualism" can be re-stated as

the existence of multiple linguistic codes in a single setting, where "multiple linguistic codes" is understood to include multiple language varieties of any sort ("multi-varietalism"); and the existence of multiple linguistic codes / multi-varietalism in a single setting means that there are multiple linguistic variants that can be accessed by a language user in that setting to fill a given slot in language structure. If, in that setting, the choice of linguistic code by a language user is (in Fishman's terms) "far from being a random matter of momentary inclination", then the selection of any given linguistic variant is (again in Fishman's terms) also not "a random matter of momentary inclination".

This chain of argumentation is meant to draw attention to three important notions: (1) that Fishman's conclusions are as applicable to settings involving closely-related language varieties as they are to settings involving non-closely related language varieties[2], (2) that "language choice" in any of these settings ultimately refers to the language users' selection of features from those available in the language varieties that co-exist in these multi-varietal settings, and (3) that the users' selection of features is "far from being a random matter of momentary inclination". Thus, the discussion that follows can be considered to be broadly applicable to many different contexts where different types of language varieties co-exist in a single societal setting; and the investigation of the use of features in that setting can be considered an exercise in discovering meaningful ("non-random") patterns/clusters of language features as the language users interact with one another in that setting.

This reformulation of Fishman could, once again, be considered a partial description of the type of work that is currently gaining ground among those sociolinguists working on questions of "coherence" in present-day, synchronic contexts, and it is this type of work that I am promoting here for use in historical and diachronic contexts as well.

4. THEORETICAL AND METHODOLOGICAL CONSIDERATIONS

Taking a cue from those colleagues working on coherence, what if, in historical linguistics work, instead of investigating a pre-determined linguistic feature (or several pre-determined features) and pre-determined social characteristics or categories of speakers, we address the question of "Who used what linguistic features with whom, when, where, why, and how?" using a data-driven, inductive approach? Instead of *a priori* deciding specifically which linguistic feature(s) and social characteristics to investigate, what if we simply analyze the entire dataset to discover the linguistic patterns present therein? In other words, what if we approach the study of language history as an exercise in *historical language-documentation fieldwork*,

[2] Encompassing the entire spectrum of what is sometimes referred to as "internal multilingualism" and "external multilingualism" ("innere Mehrsprachigkeit" and "äußere Mehrsprachkigkeit") (see Wandruszka 1975, 1979).

seeking to provide a full description of the patterns in the texts in front of us, treating each text/writer as an individual fieldwork participant who is providing language data and socio-cultural notes?

This approach of studying the full set of linguistic patterns available to us in a given historical context, in a historical language-documentation fieldwork sort of way, finally engages the theoretical stance that Hermann Paul espoused almost 140 years ago:

> The true object of philological study is rather the entire sum of the products of the linguistic activity of the entire sum of individuals in their reciprocal relations. All the groups of sound ever spoken, heard, or represented, with the associated ideas, whose symbols they were; all the numerous relations entered into by the elements of speech in the minds of individuals – all these belong to the history of language, and must, properly speaking, all be thoroughly apprehended to render a full apprehension of its development a possibility[3] (Paul 1891, pp. 2–3).

Of course, Paul himself considered such a task an impossibility[4], but then immediately stated that the impossible nature of the task doesn't mean that we shouldn't aspire to it:

> It is good to state the ideal aim of a science in all its bareness of outline. By so doing we become aware of the gulf between our powers and our possibilities. We learn that we must in many questions content ourselves with an avowal of ignorance; and that super-acuteness, which imagines that it can explain the most complicated historical developments by a few ingenious aperçus, is humbled. But it is for us an inevitable necessity to get a general idea of the play of the forces at work in this huge complex – forces which we must always keep before our eyes, if we would endeavour

[3] In a more recent translation: "The real object of investigation for the linguist consists of the entire body of speech events in all individuals and their influence on one another. Indeed, the history of a language includes all the sound combinations ever spoken, heard, or imagined by an individual and the associated representations of which they were symbols, as well as all the manifold relationships that the elements of a language entered into in the minds of individuals. In theory all these facts would have to be known to us in order to reach a complete understanding of language change" (Auer – Murray 2015, p. 48). The original German text: "Das wahre object für den sprachforscher sind vielmehr sämmtliche äusserungen der sprechtätigkeit an sämmtlichen individuen in ihrer wechselwirkung auf einander. Alle lautcomplexe, die irgend ein einzelner je gesprochen, gehört oder vorgestellt hat mit den damit associierten vorstellungen, deren symbole sie gewesen sind, alle die mannigfachen beziehungen, welche die sprachelemente in den seelen der einzelnen eingegangen sind, fallen in die sprachgeschichte, müssten eigentlich alle bekannt sein, um ein vollständiges verständniss der entwicklung zu ermöglichen" (Paul 1886, p. 22).

[4] "It need hardly be said that to solve such a problem is an impossibility" (Paul 1891, p. 3). "No one need object that there is no point in setting up a task that is so obviously impossible to fulfill" (Auer – Murray 2015, p. 48). "Man halte mir nicht entgegen, dass es unnütz sei eine aufgabe hinzustellen, deren unlösbarkeit auf er hand liegt" (Paul 1886, p. 22).

to arrange correctly the few scanty fragments which we can really claim to possess out of it.[5] (Paul 1891, p. 3)

Hermann Paul is, of course, not the only one who has, in the past, proposed this idea that, in order to even hope to understand language variation and change in historical periods, we must strive to consider the full set of language evidence and language use by individuals. Izidor Kotulič, about 65 years ago, stated the same "ideal aim" as Paul, specifically regarding the social history of Slovak, which he argued:

> ...can only be more comprehensively answered after a thorough and broadly organized investigation of linguistic monuments with the initial goal of a perfect knowledge of the language and documents of the period. The theoretical investigation of the language of the Slovak people must be based not only on a reliable methodological foundation, but also on the relatively complete historical linguistic material that is lying scattered in domestic and foreign archives (Budapest and others). For this it will be necessary to undertake a thorough archival investigation and at the same time publish in greater measure and more systematically than previously these materials on the history of the Slovak language...[6] (Kotulič 1961, pp. 122 – 123; my translation MRL)

[5] "There is a value in imagining the pure, ideal form of a science because: (i) it makes us aware of how remote from the ideal our actual capabilities are, (ii) we learn humility in the face of so many unanswered questions, and (iii) it humbles the know-it-alls who believe they have grasped the most complex historical developments simply by making some witty remarks. It is absolutely necessary to have a general idea of the forces at play in this massive complex, and we need to keep them in mind at all times in order to correctly categorize the scarce fragments that we do have access to" (Auer – Murray 2015, p. 48). "Es ist schon deshalb von wert sich das idealbild einer wissenschaft in seiner ganzen reinheit zu vergegenwärtigen, weil wir uns dadurch des abstandes bewusst werden, in welchem unser können dazu steht, weil wir daraus lernen, dass und warum wir uns in so·vielen fragen bescheiden müssen, weil dadurch die superklugkeit gedemütigt wird, die mit einigen geistreichen gesichtspunkten die compliciertesten historischen entwickelungen begriffen zu haben meint. Eine unvermeidliche notwendigkeit aber ist es für uns, uns eine allgemeine vorstellung von dem spiel der kräfte in diesem ganzen massenhaften getriebe zu machen, die wir beständig vor augen haben müssen, wenn wir die wenigen dürftigen fragmente, die uns daraus wirklich gegeben sind, richtig einzuordnen versuchen wollen" (Paul 1886, pp. 22 – 23).

[6] "...bude možno podrobnejšie riešiť iba po dôkladnom a široko organizovanom výskume jazykových pamiatok, ktorého cieľom v prvej etape bude dokonalé poznanie jazyka a písomností z tohto obdobia. Teoretický výskum jazyka slovenskej národnosti musí sa opierať nielen o spoľahlivú metodologickú základňu, ale aj o relatívne úplný historický jazykový materiál, ktorý leží roztratený v domácich i zahraničných archívoch (Budapešť a i.). Preto bude potrebné vykonať dôkladný archívny výskum a zároveň vo väčšej miere a systematickejšie ako doteraz vydávať materiály k dejinám slovenského jazyka..." (Kotulič 1961, pp. 122 – 123).

though Kotulič's statement is tempered with his recognition that we only have the data that history leaves us. Where Paul says we should strive to examine "...the entire sum of the products of the linguistic activity of the entire sum of individuals in their reciprocal relations...", Kotulič more cautiously encourages us to use "the entire sum of the linguistic products of the entire sum of individuals" that are *available* to us.

Stepping back to our "tasks for historical sociolinguistics" (derived from Tilly's (1985) "tasks for social history"), we have, up to this point in the discussion, considered the theoretical and methodological frame of the first task: "documenting structural language variation and change", bringing us now to a theoretical/methodological consideration of the second task: "reconstructing the experiences and interactions of ordinary people in connection with that variation and change". And in the same way that it can be helpful to think about the first task as "historical language-documentation fieldwork", it might be of assistance to frame the second task as *historical ethnographic fieldwork*. Ethnographic sociolinguists do fieldwork that involves becoming intimately acquainted with the socio-cultural context of the community that they are working with, attempting to identify the many different lines of affiliation/affinity between the members of the community, and the patterns of interaction of the community members, in order to better understand their patterns of language use. What if we attempt to become intimately acquainted with the writers of our historical texts and the personal interactions among them in order to answer the question of "Who *interacted* with whom, when, where, why, and how?" ? Instead of *a priori* deciding which parts of the historical socio-cultural context to focus on as significant social variables, what if we attempt to draw all possible lines of social affiliation/affinity that we can derive from the available socio-historical data and then inductively look for patterns of social interaction?

This approach of studying patterns of interaction among historical language users allows us to more fully engage with theoretical notions and models that have until now perhaps been under-utilized, or under-emphasized, in exploring specifically *historical* language variation and change:

- Mundane mobility – the "mundane movements we engage in in everyday life" are "small-scale, less dramatic in distance, and perhaps in life impact at the level of the individual, [but] their scale, intensity and pervasiveness at the level of the community as a whole mean they cannot be ignored as a source of rather striking dialect contact" (Britain 2013, p. 165, 168).
- Social networks – the contact patterns formed by instances of mundane mobility can be modeled as social networks that examine the interactions of individuals in their communities and in their constellations of contacts as potential determinants in historical language variation and change. (cf. Bergs 2005; Conde-Silvestre 2012)
- Communities of practice – "[t]he community of practice takes us away from the community defined by a location or by a population. Instead, it focuses on

a community defined by social engagement [the social networks] – after all, it is this engagement that language serves, not the place and not the people as a bunch of individuals" (Eckert – McConnell-Ginet 1992, p. 95).

With these theoretical and methodological considerations, we are now working within a framework that encompasses the first two "tasks of historical sociolinguistics" – "documenting structural language variation and change" and "reconstructing the experiences and interactions of ordinary people..." – but, as mentioned earlier, the third task – "connecting the two" – may be considered the key element in the whole theoretical and methodological conceptualization presented here. And this key element is embedded in a principle that I have espoused in other theoretical and methodological contexts and connections, summed up in the phrase: "Use all the data!".

5. USE ALL THE DATA!

This principle of "Use all the data!" has antecedents in the kinds of frames that both Paul (1886) and Kotulič (1961) invoked as "ideal aims" for maximizing the quality of work in historical linguistics, as well as in Janda and Joseph's (2003) "informational maximalism"[7], but the challenge to "Use all the data!" goes beyond the calls for gathering and considering all possible sources of language data (task 1 – "historical language-documentation fieldwork") and maintains that a fuller picture of historical language variation and user interaction will only be visible when we also gather all possible information about the language users' contexts (task 2 – "historical ethnographic fieldwork") and then, critically, bring the two together (task 3) in an examination of "Who used what linguistic features with whom, when, where, why, and how?".

The "Use all the data!" principle states (Lauersdorf 2018a, p. 112; 2018b, pp. 211 – 212; 2021, pp. 215 – 216):

1) Identify all possible sources of language data – data may be "hiding" where you don't expect it, in unexplored physical locations and in unexplored textual locations.
2) Consult the entirety of the language data available to you – avoid selective sampling (inclusion or exclusion) of language data on the basis of *a priori*

[7] "To a great extent, then, what we should really strive for, in diachronic pursuits such as historical linguistics, is what could be called *"informational maximalism"* – that is, the utilization of all reasonable means to extend our knowledge of what might have been going on in the past, even though it is not directly observable. Normally, this will involve a heavy concentration on the immediate present, but it is in fact more realistic just to say that we wish to gain a maximum of information from a maximum of potential sources: different times and different places – and, in the case of language, also different regional and social dialects, different contexts, different styles, different topics, and so on and so forth" (Janda – Joseph 2003, p. 37).

notions of what kind of data you need, how much data you need, where it should come from, etc.

3) Language data isn't the only data – use all the socio-historical data!
   a) Identify and use all possible sources of socio-historical data (again being on the lookout for socio-historical data "hidden" in unexpected places and using the entirety of the socio-historical data available to you).
   b) We only have the language data that history leaves us (what has "survived" through time), so wrap the language data in all possible socio-historical datasets to help complete the picture.

Corollary 1: If you use all the data, view all the data!
   a) If you view all the data, view all the combinations.
   b) If you view all the data, view all the angles.
   c) If you view all the data, use all the techniques.

In addition to being a statement that focuses on the necessity of completing all three "tasks for historical sociolinguistics", this call to "Use all the data!" in historical linguistic investigation also derives, in part, from the fact that historical data is inherently "finite", and therefore it is potentially incomplete, limited, fragmentary, unbalanced, for our investigations in ways that the researcher has no control over (unlike the modern contexts that Fishman or Horvath were working in where, in theory, one can always gather more data). It is therefore imperative to identify and gather as much data as possible for a given historical investigation, from all interrelated sources, both linguistic and socio-historical, and to consider especially non-traditional data sources, if one hopes to be able to assemble a sufficient dataset for data-driven, inductive analysis. If one follows the principle of gathering "all the data" in data collection, the extant historical record can often produce richly layered datasets containing linguistic features of language users in their socio-cultural interactional contexts.

This call to "Use all the data!" and "View all the data!" further derives from "the conviction that using only a selective sample and/or selective methods of analysis of the available data (based on *a priori* assumptions about the features and categories that one should expect to find in the data) limits what one is actually able to find, given that portions of the data are not being considered and that only certain analytical viewpoints are being entertained" (Lauersdorf 2021, p. 216).[8] You will

[8] Feagin (2013) makes a similar statement regarding inadequacies of working non-inductively with preconceived categories: "One danger with selecting informants by pre-selected categories is that results can be self-fulfilling or circular. For a more general community study, Horvath (1985) gathered speech data from a stratified judgment sample in Sydney, Australia, and analyzed it using principal components analysis, a statistical technique which grouped speakers into clusters according to their linguistic similarities, and in that way revealed what the sociolinguistic groupings of Sydney were, based

(almost) always find evidence of things that you are specifically looking for; but you will (almost) never find evidence of things that you are not looking for. And what if the truly salient things are the things that you aren't looking for, or the truly significant patterns and correlations are the ones that you are not considering? Thus, it is imperative to assemble and interrogate the data in a way that facilitates data-driven, inductive examination of all possible combinations of all linguistic and socio-cultural information contained in the rich data layers.[9]

6. USER INTERACTION AND QUESTIONS OF LANGUAGE AND IDENTITY

In placing the emphasis now on the third "task for historical sociolinguistics" – the task of *connecting together* "structural language variation and change" with "the experiences and interactions of ordinary people in connection with that variation and change" – we open up opportunities for consideration of questions of language and identity. Indeed, the question "Who used what linguistic features with whom, when, where, why, and how?", that we have used throughout our discussion to this point, can also be interpreted as an inquiry into language and identity.

In part, the traditional research paradigm in the investigation of language history, with its focus on the development of the standard language variety, was an attempt to explore *historical identity through language*, deriving from the 19th-century rise of the nation-state and the use of culture (including language) to construct and define nationhood and national identity, with a desire to trace the national lineage, through its culture and language, as far back in time as possible.[10] In other words, in the modern nation-state, national identity is, in part, defined by the (national) standard language, so studying the history of the standard language provides information about the history of national identity, hence the emphasis on *standard* language history in the traditional research paradigm. To represent this implementation of the traditional "language history" research paradigm as an attempt to explore "historical identity through language", we can simply reverse the language history equation provided earlier:

---

entirely on speech, rather than on preconceived notions about class membership, sex, or other social groupings" (Feagin 2013, pp. 27 – 28).

[9] It is tempting here to further drive home this point by paraphrasing a quote usually attributed to Canadian hockey star Wayne Gretzky: "You miss 100% of the shots you don't take." – for our purposes we might state this as: "You miss 100% of the data / correlations you don't consider."

[10] As stated by Milroy (2001) in regard to English, "These histories have until quite recently almost always been designed as histories of the internal structure of one variety – the standard language… They are largely *codifications* of the history of the standard language. […] When the language is given an authoritative (almost 'official') history in this way, this assures us that it has not merely sprung up overnight like a mushroom, and it becomes important to trace it as far back as possible" (Milroy 2001, p. 548).

Traditional research paradigm for the investigation of language history:
language history = standard language history = history of national identity

Traditional research paradigm for the exploration of historical identity through language:
history of national identity = standard language history = language history

This traditional interpretation is, in some ways, an example of the aphorism "history is written by the victors", in the sense that the interpretation is driven from a present-day position and view on the past, and as noted by Roach, "Whether we like it or not, we view the past from a modern standpoint, privileging (consciously or otherwise) the interests and ideals of the world we know. As a result, we tend to treat developments towards modernity as natural – and disparage the apparent dead ends that stood in its path" (Roach 2023, p. 8). Applying this to our discussion, this means that we have present-day notions of language and identity based on "how things have turned out", i.e., based on the state of how things are right now – so we take the language and identity frameworks that we have now, and the names that we give them, and we go looking for them in the past to build a lineage for what we have now. But, even if we (probably erroneously) assume that people in the past were always developing their social context in a direction toward the frameworks and names that we have now, they themselves likely didn't frame things in the way we do now and with the names that we use now. Additionally, given the notion that was just mentioned that "we tend to treat developments towards modernity as natural – *and disparage the apparent dead ends that stood in its path* [emphasis added, MRL]", there have likely been many different language and identity frameworks and names in existence historically that have not survived cycles of "disparagement", having been continually disregarded in our ongoing views and perceptions of the "path to modernity". With just those two arguments pointing to the likelihood of many different language and identity frameworks and naming conventions having existed in the past, the question arises: should the starting point of our historical investigations of language and identity really be our present-day frameworks and names?

The issue of "named languages" has become a topic in linguistic research (see, for example, Horner – Weber 2018; Horner – Bradley 2019; Saraceni – Jacob 2019), driven by studies of present-day multilingualism and linguistic "superdiversity", and focusing on the role of the inherited "framing and naming" conventions of the "the victors" who brought us the modern nation-state paradigm: one nation ~ one culture ~ one language. Horner and Bradley (2019) make the point that: "The notion that languages exist as clearly identifiable and bounded objects constitutes a widespread and fundamental language ideological belief. To be sure, the construction of named languages functions similarly to that of other categories, such as ethnicities and nations, all of which can be interpreted and instrumentalized in various ways due to

their potential malleability and their situatedness in given social and political contexts" (Horner – Bradley 2019, p. 298).

"Naming things" is ultimately an act of categorizing/classifying according to a specific framework, thereby reifying our perceptions within and through that framework, in this case involving the categorization/classification of historical language and identity. As an example of this, consider the investigation of language variation in written documents in the Slovak-speaking areas of Central Europe during the historical period before the official codification of a Slovak standard language. Written language in documents from that time and place is generally considered to have been influenced to varying degrees by Czech, Polish, and Slovak language varieties (both spoken and written), with the additional presence, in this socio-historical context, of Hungarian, German, and Latin; and much of the linguistic research into documents from that time and place has traditionally revolved around attempts to identify features that show the position of any individual text on a continuum of "Czech-ness" or "Polish-ness" or "Slovak-ness".[11] As I have argued elsewhere (Lauersdorf 2018b), this kind of research paradigm operates with a specific set of *a priori* assumptions and categories, as well as specific investigative methods and goals that derive from these assumptions and categories, that can be stated something like this:

1) *assumption*: the language of the texts must show some kind of overall genetic alignment / affiliation with a specific "named language" category (with the names and categories often coming from a present-day context);
   goal: identify the specific genetic affiliation of the language of the texts (e.g., are the texts Czech?, Polish?, Slovak?);
2) *assumption*: certain linguistic (phonological, morphological, syntactic, lexical, semantic) features that have developed distinct variants can be considered diagnostic for genetic affiliation with one of the "named language" categories (different variants of a linguistic feature become "genetic markers" of, for example, "Czech-ness", "Polish-ness", or "Slovak-ness");
   method: search the texts for instances of these specific pre-determined linguistic features that have developed distinct variants that will mark genetic affiliation with one "named language" category or another;
3) *assumption*: a preponderance of "genetic markers" of one type or another in the language features of a text is indicative of the overall genetic affiliation of the text with one of the "named language" categories;
   method and goal: track the number of occurrences of each diagnostic variant for each linguistic feature (perhaps determining whether the occurren-

[11] For more on the linguistic and socio-historical context of written documents from the Slovak-speaking area in the 15th, and especially 16th, centuries, and a review of the linguistic research into these documents, see Lauersdorf 1996 and 2010 and the literature cited there.

ces show uniformity or variation in the texts) with the hope that there will be a clearly identifiable profile that points to predominantly one "named language" affiliation or another (and hence one language identity or another) (adapted from Lauersdorf 2018b, pp. 208 – 209).

Not only does this kind of research paradigm involve the use of (likely anachronistic) social and linguistic frames and names from modern times as the position from which it investigates language and identity in earlier periods (a position that this discussion directly argues against), but this practice also limits the scope of the investigation to only selective portions of the overall data using only a selective set of methods and viewpoints, which goes in the opposite direction from the "Use all the data!" principle espoused above.[12]

In contrast with all of this, Horner and Weber's (2018) "social approach" to multilingual contexts emphasizes the role of social interaction and the use of language socially, and the importance of investigating language in its actual interactional context – in our case its actual *historical* interactional context. With this in mind, what if, in our investigations of language variation and user interaction in historical contexts, we do not begin with pre-determined names and categories that channel and constrain the analysis of our data and results? In performing task 3 of the "tasks for historical sociolinguistics" (the task of *connecting* the variation in language features with the social interactions of their users), taking an approach of discovering *inductively* "Who used what linguistic features with whom, when, where, why, and how?", and describing the resulting patterns, allows us to present information about the past and how it evolved into the present state without imposing (likely anachronistic) assumptions that are based on present-day framing and naming.

7. A FINAL POSITION STATEMENT

In the revised research paradigm for the investigation of language history, that I put forth in section 2 above:

[12] This questioning of the use of present-day frames and names in our research paradigms for historical language settings should also prompt questions regarding what the writers of historical documents themselves thought about their written linguistic practice and their identity – i.e., their framing and naming conventions that they used in their time period. Did they give a *name* to the collection of linguistic features that they were writing in any given document (in this example "Czech" or "Polish" or "Slovak"); and did they equate the linguistic features that they were writing with a *named* identity for themselves or with a *named* identity for the intended recipient(s) of any given document? Consideration of these questions goes beyond the scope of this essay, but see Lauersdorf 2018b for some additional discussion of these and related questions of "the writing of historical identity".

language history =	history of variation and change in all varieties of language in play in user interactions in a given socio-cultural context	=	a multifaceted approach to questions of history of language and identity

it is the third portion – "a multifaceted approach to questions of history of language and identity" – that is realized by pursuing the "tasks of historical sociolinguistics" all the way through to the third, highly important task of *connecting language and users*. In sum:

A data-driven, use-all-the-data connection and correlation of:

patterns of language variation derived by data-driven, inductive methods from full-scale "historical language-documentation fieldwork" that collects and analyzes the full scope of available historical data (i.e., the products of task 1)

AND

patterns of social interaction of language users carefully reconstructed from full-scale "historical ethnographic fieldwork" that reconstructs, for example, mundane mobility, social networks, communities of practice (i.e., the products of task 2)

expressly seeks to investigate:

who used what linguistic features with whom, when, where, why, and how? (i.e., task 3).

This set of tasks, carried out in this way as a research paradigm, seeks to specifically avoid *a priori* assumptions that lead to results based on pre-determined selective sampling of data and methods and to interpretations based on likely anachronistic present-day frameworks and names, and thereby holds the potential to paint a more detailed and accurate picture of the social history of language varieties and their users, providing greater richness in our overarching understanding of language history and historical linguistic identity.

References

AUER, Peter – MURRAY, Robert W. (eds.) (2015): *Hermann Paul's* Principles of Language History *Revisited: Translations and Reflections*. Berlin and Boston: Walter de Gruyter.

BEAMAN, Karen V. – GUY, Gregory R. (eds.) (2022): *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. New York: Routledge.

BERGS, Alexander (2005): *Social Networks and Historical Sociolinguistics: Studies in Morphosyntactic Variation in the Paston Letters (1421–1503)*. Berlin and New York: Mouton de Gruyter.

BRITAIN, David (2013): The Role of Mundane Mobility and Contact in Dialect Death and Dialect Birth. In: D. Schreier – M. Hundt (eds.): *English as a Contact Language*. Cambridge: Cambridge University Press, pp. 165–181.

CERRUTI, Massimo – VIETTI, Alessandro (2022): Identifying Language Varieties: Coexisting Standards in Spoken Italian. In: K. V. Beaman – G. R. Guy (eds.): *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. New York: Routledge, pp. 261–280.

CONDE-SILVESTRE, Juan Camilo (2012): The Role of Social Networks and Mobility in Diachronic Sociolinguistics. In: J. M. Hernández-Campoy – J. C. Conde-Silvestre (eds.): *The Handbook of Historical Sociolinguistics*. West Sussex: Wiley-Blackwell, pp. 332–352.

ECKERT, Penelope – MCCONNELL-GINET, Sally (1992): Communities of Practice: Where Language, Gender, and Power All Live. In: K. Hall – M. Bucholtz – B. Moonwomon (eds.): *Locating Power, Proceedings of the 1992 Berkeley Women and Language Conference*. Berkeley: Berkeley Women and Language Group, pp. 89– 99.

FEAGIN, Crawford (2013): Entering the Community: Fieldwork. In: J. K. Chambers – N. Schilling (eds.): *The Handbook of Language Variation and Change*. 2nd ed. West Sussex: Wiley-Blackwell, pp. 19– 37.

FISHMAN, Joshua A. (1965): Who speaks what language to whom and when? In: *La Linguistique,* Vol. 1, No. 2, pp. 67– 88.

GHYSELEN, Anne-Sophie – DE VOGELAER, Gunther (2018): Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the 'Variety' Concept. In: *Frontiers in Psychology*, Vol. 9, Article 385. Available at: https://doi.org/10.3389/fpsyg.2018.00385

GUY, Gregory R. (1980): Variation in the Group and in the Individual: The Case of Final Stop Deletion. In: W. Labov (ed.): *Locating Language in Time and Space*. New York: Academic Press, pp. 1–36.

GUY, Gregory R. – HINSKENS, Frans (2016): Linguistic Coherence: Systems, Repertoires and Speech Communities. In: *Lingua*, Vol. 172–173, pp. 1–9. Available at: https://doi.org/10.1016/j.lingua.2016.01.001

HINSKENS, Frans – GUY, Gregory R. (eds.) (2016): Coherence, Covariation and Bricolage. Various Approaches to the Systematicity of Language Variation. Special issue of *Lingua*, Vol. 172–173. Available at: https://www.sciencedirect.com/journal/lingua/vol/172

HORNER, Kristine – WEBER, Jean-Jacques (2018): *Introducing Multilingualism: A Social Approach*. 2nd ed. London and New York: Routledge.

HORNER, Kristine – BRADLEY, Andrew J. (2019): Language Ideology. In: J. Darquennes – J. C. Salmons – W. Vandenbussche (eds.): *Language Contact: An International Handbook, vol. 1*. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science / Manuels de linguistique et des sciences de communication [HSK], 45/1). Berlin: De Gruyter Mouton, pp. 296–307.

HORVATH, Barbara (1985): *Variation in Australian English: The Sociolects of Sydney*. Cambridge: Cambridge University Press.

HORVATH, Barbara – SANKOFF, David (1987): Delimiting the Sydney speech community. In: *Language in Society*, Vol. 16, No. 2, pp. 179–204.

HUA, Xia – MEAKINS Felicity – ALGY, Cassandra – BROMHAM, Lindell (2022): Language Change in Multidimensional Space: New Methods for Modelling Linguistic Coherence. In: *Language Dynamics and Change*, Vol. 12, No. 1, pp. 78 – 123. Available at: https://doi.org/10.1163/22105832-bja10015

JANDA, Richard D. – JOSEPH, Brian D. (2003): On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Linguistics. In: B. D. Joseph – R. D. Janda (eds.): *The Handbook of Historical Linguistics*. Oxford: Blackwell Publishing, pp. 3–180.

KOTULIČ, Izidor (1961): Z mestských kníh Partizánskej Ľupče zo 16. storočia. In: E. Pauliny (ed.): *Štúdie a pramene k dejinám jazyka slovenskej národnosti. Jazykovedné štúdie VI*, Bratislava: Vydavateľstvo SAV, pp. 119–177.

LAUERSDORF, Mark Richard (1996): *The question of 'cultural language' and inter-dialectal norm in 16th century Slovakia: A phonological analysis of 16th century Slovak administrative-legal texts*. München: Otto Sagner. [Slavistische Beiträge, 335.]

LAUERSDORF, Mark Richard (2010): *The morphology of 16th-century Slovak administrative-legal texts and the question of diglossia in pre-codification Slovakia*. München: Otto Sagner. [Slavistische Beiträge, 473.]

LAUERSDORF, Mark Richard (2018a): Linguistic Visualizations as *objets d'art*? In: N. Bubenhofer – M. Kupietz (eds.): *Visualisierung sprachlicher Daten*. Heidelberg: Heidelberg University Publishing, pp. 91–122. Available at: https://doi.org/10.17885/heiup.345.c4397

LAUERSDORF, Mark Richard (2018b): Historical (Standard) Language Development and the Writing of Historical Identities: A Plaidoyer for a Data-Driven Approach to the Investigation of the Sociolinguistic History of (Not Only) Slovak. In: S. M. Dickey – M. R. Lauersdorf (eds.): *V zeleni drželi zeleni breg: Studies in Honor of Marc L. Greenberg*. Bloomington, Indiana: Slavica Publishers, pp. 199–218.

LAUERSDORF, Mark Richard (2021): Historical Sociolinguistics and the Necessity of Interdisciplinary Collaboration. In: A. Burkette – T. Warhol (eds.): *Crossing Borders, Making Connections: Interdisciplinarity in Linguistics*. Berlin: De Gruyter Mouton, pp. 207–230. Available at: https://doi.org/10.1515/9781501514371-015.

MEYERHOFF, Miriam – KLAERE, Steffen (2017): A case for clustering speakers and linguistic variables: Big issues with smaller samples in language variation. In: I. Buchstaller – B. Siebenhaar (eds.): *Language Variation – European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8), Leipzig, May 2015*. Amsterdam and Philadelphia: John Benjamins, pp. 23–46. Available at: https://doi.org/10.1075/silv.19.02mey

MILROY, James (2001): Language Ideologies and the Consequences of Standardization. In: *Journal of Sociolinguistics*, Vol. 5, No. 4, pp. 530–55.

PAUL, Hermann (1886): *Principien der Sprachgeschichte*. Zweite Auflage. Halle: Max Niemeyer.

PAUL, Hermann (1891): *Principles of the History of Language*. Translated from the second edition of the original by H. A. Strong. New and revised edition (1890). London: Longmans, Green, and Co.

ROACH, Levi (2023): The powerful of one era are not the same as those of the preceding one. Contribution to: Is History Written by the Winners? In: *History Today* 73.6 (June): 8. Available at: https://www.historytoday.com/archive/head-head/history-written-winners

SARACENI, Mario – JACOB, Camille (2019): Revisiting borders: Named languages and de-colonization. In: *Language Sciences*, Vol. 76. Available at: https://doi.org/10.1016/j.langsci.2018.05.007

TILLY, Charles (1985): Retrieving European Lives. In: O. Zunz (ed.): *Reliving the Past: The Worlds of Social History*. Chapel Hill: University of North Carolina Press, pp. 11–52.

TRUMPER, John – MADDALON, Marta (1990): Il problema delle varietà: L'italiano parlato nel Veneto. In: M. A. Cortelazzo – A. M. Mioni (eds.): *L'italiano regionale. Atti del XVIII Congresso Internazionale di Studi. Padova – Vicenza, 14–15 settembre 1984*. Roma: Bulzoni, pp. 159–191.

VIETTI, Alessandro (2019): La varietà di lingua come insieme di tratti coerenti: Verso una caratterizzazione empirica. In: *Rivista Italiana di Dialettologia. Lingue dialetti società*, Vol. 43, pp. 11–32.

WANDRUSZKA, Mario (1975): Mehrsprachigkeit. In: H. Moser (ed.): *Sprachwissenschaft und Sprachdidaktik: Jahrbuch 1974 des Instituts für deutsche Sprache*. *Sprache der Gegenwart, 36*. Düsseldorf: Schwann, pp. 321–350.

WANDRUSZKA, Mario (1979): *Die Mehrsprachigkeit des Menschen*. München: Piper.

POKYNY PRE AUTOROV

Redakcia JAZYKOVEDNÉHO ČASOPISU uverejňuje príspevky **bez poplatku** za publikovanie.
Akceptované jazyky: všetky slovanské jazyky, angličtina, nemčina. Súčasťou vedeckej štúdie a odborného príspevku je abstrakt v angličtine (100 – 200 slov) a zoznam kľúčových slov v angličtine (3 – 8 slov).
Súčasťou vedeckej štúdie a odborného príspevku v inom ako slovenskom alebo českom jazyku je zhrnutie v slovenčine (400 – 600 slov) – preklad do slovenčiny zabezpečí redakcia.
Posudzovanie príspevkov: vedecké príspevky sú posudzované anonymne dvoma posudzovateľmi, ostatné príspevky jedným posudzovateľom. Autori dostávajú znenie posudkov bez mena posudzovateľa.
Technické a formálne zásady:

- Príspevky musia byť v elektronickej podobe (textový editor Microsoft Word, font Times New Roman, veľkosť písma 12 a riadkovanie 1,5). V prípade, že sa v texte vyskytujú zvláštne znaky, tabuľky, grafy a pod., je potrebné odovzdať príspevok aj vo verzii pdf alebo vytlačený.
- Pri mene a priezvisku autora je potrebné uviesť pracovisko.
- Text príspevku má byť zarovnaný len z ľavej strany, slová na konci riadku sa nerozdeľujú, tvrdý koniec riadku sa používa len na konci odseku.
- Odseky sa začínajú zarážkou.
- Kurzíva sa spravidla používa pri názvoch prác a pri uvádzaní príkladov.
- Polotučné písmo sa spravidla používa pri podnadpisoch a kľúčových pojmoch.
- Na literatúru sa v texte odkazuje priezviskom autora, rokom vydania a číslom strany (Horecký 1956, s. 95).
- Zoznam použitej literatúry sa uvádza na konci príspevku (nie v poznámkovom aparáte) v abecednom poradí. Ak obsahuje viac položiek jedného autora, tie sa radia chronologicky.

Bibliografické odkazy:

- knižná publikácia: ONDREJOVIČ, Slavomír (2008): *Jazyk, veda o jazyku, societa*. Bratislava: Veda, vydavateľstvo SAV. 204 s.
- slovník: JAROŠOVÁ, Alexandra – BUZÁSSYOVÁ, Klára (eds.) (2011): *Slovník súčasného slovenského jazyka. H – L.* [2. zv.]. Bratislava: Veda, vydavateľstvo SAV.
- štúdia v zborníku: ĎUROVIČ, Ľubomír (2000): Jazyk mesta a spisovné jazyky Slovákov. In: S. Ondrejovič (ed.): *Sociolinguistica Slovaca 5. Mesto a jeho jazyk*. Bratislava: Veda, vydavateľstvo SAV, s. 111 – 117.
- štúdia v časopise: DOLNÍK, Juraj (2009): Reálne vz. ideálne a spisovný jazyk. In: *Jazykovedný časopis*, roč. 60, č. 1, s. 3 – 12. DOI 10.2478/v10113-009-0001-3. [cit. DD-MM-RRRR].
- internetový zdroj: Slovenský národný korpus. Verzia prim-5.0-public.all. Bratislava: Jazykovedný ústav Ľudovíta Štúra SAV. Dostupné na: https://korpus.juls.savba.sk [cit. DD-MM-RRRR].

INSTRUCTION FOR AUTHORS

JOURNAL OF LINGUISTICS publishes articles **free of publication charges**.
Accepted languages: all Slavic languages, English, German. Scientific submissions should include a 100-200 word abstract in English and a list of key words in English (3-8 words).
Scientific articles in a language other than Slovak or Czech should contain a summary in Slovak (400-600 words) – translation into Slovak will be provided by the editor.
Reviewing process: scientific articles undergo a double-blind peer-review process and are reviewed by two reviewers, other articles by one reviewer. The authors are provided with the reviews without the name of the reviewer.
Technical and formal directions:

- Articles must be submitted in an electronic form (text editor Microsoft Word, 12-point Times New Roman font, and 1.5 line spacing). If the text contains special symbols, tables, diagrams, pictures etc. it is also necessary to submit a pdf or printed version.
- Contributions should contain the full name of the author(s), as well as his/her institutional affiliation(s).
- The text of the contribution should be flush left; words at the end of a line are not hyphenated; a hard return is used only at the end of a paragraph.
- Paragraphs should be indented.
- Italics is usually used for titles of works and for linguistic examples.
- Boldface is usually used for subtitles and key terms.
- References in the text (in parentheses) contain the surname of the author, the year of publication and the number(s) of the page(s): (Horecký 1956, p. 95).
- The list of references is placed at the end of the text (not in the notes) in alphabetical order. If there are several works by the same author, they are listed chronologically.

References:

- Monograph: ONDREJOVIČ, Slavomír (2008): *Jazyk, veda o jazyku, societa*. Bratislava: Veda, vydavateľstvo SAV. 204 p.
- Dictionary: JAROŠOVÁ, Alexandra – BUZÁSSYOVÁ, Klára (eds.) (2011): *Slovník súčasného slovenského jazyka. H – L.* [2. zv.]. Bratislava: Veda, vydavateľstvo SAV.
- Article in a collection: ĎUROVIČ, Ľubomír (2000): Jazyk mesta a spisovné jazyky Slovákov. In: S. Ondrejovič (ed.): *Sociolinguistica Slovaca 5. Mesto a jeho jazyk*. Bratislava: Veda, vydavateľstvo SAV, pp. 111–117.
- Article in a journal: DOLNÍK, Juraj (2009): Reálne vz. ideálne a spisovný jazyk. In: *Jazykovedný časopis*, Vol. 60, No. 1, pp. 3–12. DOI 10.2478/v10113-009-0001-3. [cit. DD-MM-RRRR].
- Internet source: Slovenský národný korpus. Verzia prim-5.0-public.all. Bratislava: Jazykovedný ústav Ľudovíta Štúra SAV. Available at: https://korpus.juls.savba.sk [cit. DD-MM-YEAR].

ISSN 0021-5597 (tlačená verzia/print) ISSN 1338-4287 (verzia online) MIČ 49263

JAZYKOVEDNÝ ČASOPIS

VEDECKÝ ČASOPIS PRE OTÁZKY TEÓRIE JAZYKA

JOURNAL OF LINGUISTICS

SCIENTIFIC JOURNAL FOR THE THEORY OF LANGUAGE

Objednávky a predplatné prijíma/Orders and subscriptions are processed by:
SAP – Slovak Academic Press, s. r. o., Bazová 2, 821 08 Bratislava
e-mail: sap@sappress.sk

Registračné číslo 7044

Evidenčné číslo 3697/09

IČO vydavateľa 00 167 088

Ročné predplatné pre Slovensko/Annual subscription for Slovakia: 12 €, jednotlivé číslo 4 €
Časopis je v predaji v kníhkupectve Veda, Štefánikova 3, 811 06 Bratislava 1