DiaClas SDP 5

POVZETKI PRISPEVKOV
BOOK OF ABSTRACTS

1. DiaClas – Konferenca Narečna klasifikacija, Ljubljana, Atrij ZRC SAZU, 23.–25. april 2025
1st DiaClas - Dialect Classification Conference, Ljubljana, Atrij ZRC SAZU, 23th-25th April 2025

in / and

5. Slovenski dialektološki posvet, Ljubljana, Atrij ZRC SAZU, 22. in 25. april 2025
5th Slovenian Dialectological Symposium, Ljubljana, Atrij ZRC SAZU, 22th and 25th April 2025

Gotzon Aurrekoetxea
University of the Basque Country (UPV/EHU), Spain
gotzonaurre@gmail.com

Ariane Ensunza
University of the Basque Country (UPV/EHU), Spain
ariane.ensunza@ehu.eus

Jožica Škofic
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša (Fran Ramovš Institute of the Slovenian Language), Slovenia
jozica.skofic@zrc-sazu.si

Hans Van de Velde
Fryske Akademy, Utrecht University, the Netherlands
HVandeVelde@fryske-akademy.nl

DIACLEU - LOOKING BACK AND FORWARD

(Introduction paper)

This is a looking back paper of *DIACLEU*, Dialect classifications of languages in Europe, published in *Dialectología*. In this paper, we gather and describe which languages did we cover considering genetically (families), its status and some special cases, but we also explain which languages are in progress to be published in the forthcoming issues, and which ones are we lacking. Considering the published issues, we detail the type of classifications for each language and insight in motivation for this type of research (sometimes to include variation in standard, sometimes to get rid of it, showing it is a separate language), as well as shifts in data and varieties used. Finally, we also explain how we had to deal with conflicting principles and language conflict due to the influence that politics have in dialectology and dialect classifications.

Gotzon Aurrekoetxea
University of the Basque Country (UPV/EHU), Spain
gotzonaurre@gmail.com

Miguel Salicrú
University of Barcelona, Spain
msalicru@ub.edu

TECHNICAL ASPECTS FOR DIALECT CLASSIFICATION

The geographical variation of a language reflects the diversity of the groups that are in some way interrelated. To highlight the structure of groups, homogeneous within and heterogeneous between, dialect classification has used different theoretical-technical frameworks: ethnological or extralinguistic, isoglottic using a few linguistic features, perceptual and quantitative or dialectometric. Within these frameworks, the DIACLEU project (http://diacl.eu) has promoted the state of the art of the dialect classification of European languages. With the experience of the past, it is worth reflecting on the most appropriate techniques in dialect classification work in the future. Based on dialectometric techniques, both those developed so far and those under development, once the representativeness of the corpus is guaranteed and the robustness of the information is assured (which necessarily have to be dealt with from a linguistic point of view), the technical challenges focus on: establishing the number of groups, dialects, subdialects or varieties; obtaining homogeneous groups; identifying dialect groups and transition zones; extracting the most representative items of each group and distinctive items between them; quantifying the degree of discordance or separation between dialects, subdialects or varieties; and graphically representing the groups that make up the dialects, subdialects or varieties. In this contribution we propose a comprehensive way forward that encompasses these aspects, using data from various linguistic atlases. The aim is to achieve results that go beyond the current situation in the question of dialectal classifications and to increase the reliability of the results.

Stamatis Beis
Research Centre of Modern Greek Dialects, Greece
beis@academyofathens.gr

PROBLEMS OF CLASSIFICATION OF THE PERIPHERAL VARIETIES OF MODERN GREEK

There are very few overall classifications, or cartographical representations of Greek dialects available (cf. Trudgill 2003: 45; Tzitzilis 2016 and the new dialectological interactive map of the Research Centre for Modern Greek Dialects of the Academy of Athens). The most common way of classification distinguishes core / mainland Greek from the peripheral dialects, which are thus called 'proper dialects'. Mainland Greek is usually divided into Northern and Southern Modern Greek, based on one of the most accepted criterion: the raising of unstressed [e] and [o] to [i] and [u] respectively and the frequent deletion of unstressed [i] and [u], e.g. /peði/ > [piði] 'child', /kulúri/ > [klur] 'bagel' (Hatzidakis [1892] 1975: 342, Kontossopoulos [1981] 2008: 92). These two phenomena, called "northern vocalism'', characterise Northern Modern Greek varieties and distinguish them from Southern Modern Greek, in which they are absent. Apart from this isogloss, most Northern Modern Greek dialects share other features, such as the decline of use of the genitive plural as opposed to Southern Modern Greek (including Standard Modern Greek). This type of classification uses qualitative criteria (e.g. phonetic-phonological) and is typical of isoglossic dialectology.

Turning to peripheral dialects, their grouping in a single category is largely reminiscent of perceptual dialectology. It is a kind of ethnological classification, based on geographical and political criteria, since most of the peripheral varieties of Modern Greek are/were spoken in distant areas, outside the Greek state. Italiot Greek is spoken in Southern Italy (Calabria and Puglia), Pontic and Cappadocian used to be spoken in Turkey until 1924 and since then in Greece (apart from Muslim speakers in a few villages in Pontus) and the Mariupolitan varieties are still spoken in Ukraine despite being in decline. This presentation is an attempt to establish an isoglossic classification, to regroup the main features of peripheral dialects, and to locate the phenomena that make them unintelligible to a Standard Modern Greek or a core Greek dialect speaker.

Đorđe Božović
University of Belgrade, Faculty of Philology, Serbia
djordje.boz@gmail.com

ŠTOKAVIAN DIALECT CLASSIFICATION: A STRATIFICATIONAL-SCALING APPROACH

Being a part of the wider South Slavic dialect continuum and of the Central South Slavic geolinguistic area (see, e.g., Šekli 2013), Štokavian dialects are notoriously hard to define as a unit and to classify according to strictly genealogical criteria. This concerns both their external affiliation, on the one hand, especially the relationship of Štokavian with the Čakavian and Kajkavian diasystems or the balkanised Torlakian, and their internal subdivisions, on the other hand, both being highly obscured by various migratory processes and dialect shifts, but also by features spreading unevenly and at different historical periods from various innovation centres within the Štokavian dialect complex (cf. Ivić 1956; Brozović & Ivić 1988).

The classification devised by Belić (1909) and Ivić (1956; 1958) uses two perceptually salient phonological features as the classification criteria, namely the reflexes of the Proto-Slavic vowel yat (**ě* > *i*, *ije*/*je*, *e*, and various combinations thereof) and the accentual systems, in particular presence/lack of neo-Štokavian accent retraction, as a later, radially spreading innovation. However, while Ivić, pursuing a structural approach to dialectology, considers the two criteria simultaneously, Belić originally envisioned them in an explicitly stratificational manner, i.e. as two historically distinct dialect divisions, one of which (viz., the neo-Štokavian accentual and morphological innovations) has been subsequently overlaid on top of another, earlier one (viz., the original split based on yat-reflexes).

This paper proceeds to further develop a stratificational approach to Štokavian dialect classification, by utilising the concept of *geolinguistic scale*, and to discuss why this approach may be better suited for classifying Štokavian dialect formations, given its aforementioned elusiveness as a strictly genealogical unit. Instead of enumerating them in a taxonomic manner, various isoglosses that figure as possible classification criteria in the literature (i.e., the **dj* > *j* isogloss, šćakavism, yat-reflexes, the neo-Štokavian innovations, etc., as well as the so-called balkanisms in the morphosyntactic domain, as contact-induced innovations from a non-Slavic substratum/adstratum), are examined in the paper with an emphasis on their areality and historicity at different *scales*. In that way, a stratificational-scaling approach provides a more nuanced, multidimensional view of dialect classification, whereby, for historical as well as geolinguistic reasons, particular criteria may be more suitable for classifying certain dialect areas than others, and vice versa.

In addition, given its scalar character, the stratificational-scaling approach pursued here may be particularly suited for the present-day Štokavian dialect classification, as well, which is additionally complicated following the political dissolution of Yugoslavia and the sociolinguistic breakup of its Štokavian-based pluricentric standard language, as different isoglosses may not have the same value for dialect classification in different parts of the Štokavian area (e.g., the neo-Štokavian retraction being practically irrelevant for the classification of Bosnian dialects, as they are all heavily neo-Štokavianised; or the yat-reflexes being irrelevant for Montenegrin, as these are, with minor exceptions, in principle, all Ijekavian, cf. Čirgić 2011; or the balkanised : non-balkanised distinction as exclusively relevant for Serbian dialects, cf. Ivić 2009), while at the same time recognising the wider areal linguistic unity of Štokavian (Božović 2017).

Dunja Brozović Rončević
University of Zadar, Croatia
dbrozovic@unizd.hr

ADVANCES IN IN THE CLASSIFICATION AND CARTOGRAPHIC REPRESENTATION OF THE DISTRIBUTION OF CROATIAN DIALECTS BASED ON RECENT RESEARCH BY THE LINGEH PROJECT

This paper is based on systematic dialectological research conducted within the LinGeH project (*Linguistic Geography of Croatia in the European Context*), which was implemented at the University of Zadar from 2020 to June 2024. The network of points created for this project includes 173 Croatian local dialects of all three Croatian dialectal groups: 38 Kajkavian, 69 Čakavian and 66 Štokavian points. For 150 local dialects, brief linguistic descriptions which provide basic linguistic features at all levels were made. This was a necessary prerequisite for a possible revision of older dialect classifications (Brozović, Finka, Lisac, Lončarić, Lukežić, Moguš, Vranić et al.). The distribution and subsequent classification of Croatian local dialects is the result of dynamic historical events and centuries of constant migrations that have shaped a very indented linguistic mosaic that stands out from other Slavic languages in its complexity. The last war fought in Croatia in the 1990s once again reshaped this dialect patchwork.

The effort to create a new map of Croatian dialects based on the results of the LinGeH project has shown that the network of 173 points is not dense enough to accurately delineate the boundaries of the distribution of individual dialects. Therefore, for the purposes of this paper, data from other points for which we had available linguistic material, a total of 335 local dialects, are also analyzed. It is important to note that for the purposes of the LinGeH project, 40 local dialects were investigated for which we had no systematic linguistic data so far, and that some of this new data has actually contributed to the delineation of more precise boundaries (for example, the Štokavian dialects in the far east of Slavonia, the spread of the Kajkavian dialect islands in western Slavonia significantly further east than shown on older maps etc.). Furthermore, research has shown that there are no local dialects in Croatia today that belong to the East Bosnian dialect, although it has been registered by most Croatian language maps, most likely uncritically reproducing the older situation without conducting more recent research. Given that in the recent fieldwork research, attempts were made to conduct surveys with speakers of different generations as much as possible, changes that have occurred in local dialects over the past decades were also recorded. In order to gain insight into the stability and dynamics of Croatian dialectal patchwork, the material was also compared with older questionnaires for which research was conducted mostly in the 1960s, as well as questionnaires compiled for the Croatian

Linguistic Atlas. All the maps that will be presented are created by GIS related database of the LinGeH project.

Lars Bülow
ünch
lars.buelow@germanistik.uni-muenchen.de

Oliver Schallert
FAU Erlangen-Nürnberg, Germany
oliver.schallert@fau.de

N-GRAM-BASED CLASSIFICATION OF COHERENT DIALECT AREAS USING THE MAURER QUESTIONNAIRES

Aims: The key objective of this paper is to utilise the Maurer questionnaires, which were compiled in the 1930s, to examine the extent to which dialect areas in Bavaria can be determined through the application of quantitative and qualitative classification techniques. Furthermore, we investigate language variation and change in real-time by comparing the findings of the Maurer questionnaires with data from the Bavarian Language Atlas projects (e.g. the *Sprachatlas von Oberbayern* [= SOB]), which were conducted in the 1980s and 1990s.

Background: In the 1930s, Friedrich Maurer initiated an indirect dialectological survey in Southern Germany (Bavaria, Baden, and Alsace) with the objective of identifying lexical and phonological differences. The speakers were requested to provide translations in their local dialect of the written stimuli. The survey yielded a total of ca. 8,000 written questionnaires from the Alemannic, Bavarian, and East Franconian dialect regions, which have been preserved. Apart from the Alemannic materials, some of which have already been analysed (see, for example, Strobel 2021), the questionnaires from the Bavarian and East Franconian dialect areas have not been subjected to systematic examination to date, despite the fact that they have already been digitised.

Research questions and research design: A detailed analysis of the responses to two sentences from the Maurer questionnaire, comprising 2,000 locations, is presented in this paper. The sentences are FM 84 (*Der böse Bruder ist tot.* 'The mean brother is dead.') and FM85 (*Ich habe ein neues Kleid.* 'I have a new gown.'). The analysis addresses the following research questions:

RQ1: Can the data be used to determine dialect areas in Bavaria around 1934?

RQ2: How can these data be used to identify language change in real time?

Following Birkenes' (2019) work on Frisian, we use letter trigrams for investigating dialectal differences in the respective regions. The resulting frequency profiles are then compared using cosine distances and visualised using multidimensional scaling (Manning and Schütze 1999: 300; Heeringa, 2004). Furthermore, cluster analyses were conducted using the k-

means algorithm (2–4 solutions). One advantage of this approach is that it is exploratory in nature, meaning that it does not require annotation or alignment of the raw data. However, as an aggregative method, trigrams do not permit direct access to the different variables (see Vergeiner 2023). In order to cross-validate the results of the quantitative approach, a qualitative analysis was conducted of two phonological variables (MHG *ô* in *tot* 'dead', MHG *oe* in *böse* 'mean'). Furthermore, the results from the Maurer questionnaires were compared with data from the Bavarian Atlas projects from the 1980s and 1990s in order to analyse possible language change in real time.

Results: The results demonstrate that the dialect classification proposed by Wiesinger (1983) can be largely corroborated through the application of a trigram-based classification technique and qualitative analysis. A comparison with the data from the Bavarian Language Atlas projects on MHG *ô* and MHG *oe* reveals that the dialect regions associated with specific variants of these two variables have remained largely stable.

Stanisław Cygan
Instytut Literaturoznawstwa i Językoznawstwa Uniwersytetu Jana Kochanowskiego w Kielcach (Inštitut za literarne vede in jezikoslovje na Univerzi Jana Kochanowskega v Kielcah), Poljska
stanislaw.cygan@ujk.edu.pl

MIEJSCE GWAR KIELECKICH JAKO PODDIALEKTU MAŁOPOLSKI ŚRODKOWOPÓŁNOCNEJ W LITERATURZE DIALEKTOLOGICZNEJ

Autor referatu podejmuje próbę opisu niewielkiego obszaru (zespołu) gwarowego, charakteryzowanego w literaturze dialektologicznej w różny sposób i na podstawie różnych kryteriów (por. np. propozycję podziału dialektalnego polskiego obszaru gwarowego autorstwa Kazimierza Nitscha w *Dialektach języka polskiego* (1915), *Wybór polskich tekstów gwarowych* (wydanie I: 1929; wyd. II: 1960; wyd. III: 1968); Barbary Bartnickiej w publikacji: *Podstawowe wiadomości z dialektologii polskiej z ćwiczeniami*, Warszawa 1959 (dialekt kielecko-sandomierski); Stanisława Urbańczyka w *Zarysie dialektologii polskiej* 1981, wyd. szóste); Eugeniusza Pawłowskiego (por. artykuł *Podział gwar małopolskich na tle wzajemnych wpływów gwarowych oraz nowych tendencji językowych*, „Rozprawy Komisji Językowej WTN", t. VI, 1966, 191-202; Karola Dejny, najbardziej zasłużonego badacza gwar kieleckich (por. tegoż, *Atlas gwarowy województwa kieleckiego*, *Słownictwo ludowe z terenu byłych województw kieleckiego i łódzkiego* (1974-1981, t. XX-XXXI oraz dane zawarte w kilku Jego artykułach naukowych); Renaty Kucharczyk czy innych.

Z przedstawionych podziałów dialektalnych, opartych na różnych kryteriach, wynika, że Kielecczyzna to dialektalny obszar Małopolski (Małopolski środkowopółnocnej), charakteryzujący się trzema podstawowymi cechami językowymi: zanikiem rezonansu nosowego, ginącą metatezą grup *śr*, *źr*, częściowym lub całkowitym zanikiem *e* ścieśnionego, także szeregiem cech gwarowych wspólnych z innymi gwarami polskimi oraz kilkoma, częściowo odrębnymi od nich.

Gwary kieleckie można by określić jako poddialekt Małopolski środkowopółnocnej.

Agnė Čepaitienė
Institute of the Lithuanian Language, Lithuania
agne.cepaitiene@gmail.com

THE PRINCIPLES OF THE SPATIAL DATA INFRASTRUCTURE FOR LITHUANIAN LANGUAGE: THE CASE OF GEOSPATIAL ANALYSIS OF PERCEPTUAL DIALECTOLOGY DATA

The principles of multimodal dialectology of Lithuanian language are applied in the 21st century Lithuanian language variation studies (Mikulėnienė 2018; 2020: 11–33). This methodology covers the principles of processing and assessment not only linguistic data, but also the data of linguistic landscape, the socio-historical and socio-cultural situation of the location, the attitude of language users towards local language variants, and the language attitude of the local young people.

All of the above mentioned data sets are processed and analysed using the Lithuanian Language Variation data model (Lith. *Tarmynas*) developed on the basis of Geographic Information Systems (GIS) tools and methods (*Geolinguistics portal*, Mikulėnienė, Čepaitienė 2021). The geospatial data model is implemented as a geodatabase in the multi-user relational database management system based on ArcGIS technology (ArcGIS, ArcGIS geodatabase). The data management processing includes collection, digitalisation and storage of spatial data sets and layers, geospatial analysis of linguistic, socio-economic and socio-cultural phenomena, as well as publishing of the results through web based digital maps.

The functioning of the database is made possible by the Spatial Data Infrastructure for Lithuanian Language (SDI for LTL) based on Esri ArcGIS platform. The aim of the presentation is to demonstrate the potential of SDI for LTL and to present the analysis results of one type of linguistic entity stored in *Tarmynas* – the younger generation language attitude. The data was collected by applying the principles of perceptual dialectology (Handbook of Perceptual Dialectology vol. 1, 2, etc., Aliūkaitė et al. 2017, 2020).

The pilot study analyses the data from 100 responses to perceptual questionnaire collected from 16–20-year-old gymnasium students in Eastern Lithuania between 2023 and 2024. The material was collected using a web-based questionnaire developed in ArcGIS Online application Survey123 (Perceptual Questionnaire for Gymnasium Students). The questions cover the language use patterns of the subjects, their perception of the linguistic environment, (non-)dialectal language attitudes, associates of dialectal speakers, and a mapping task that determines the geolinguistic competence of the subjects, i.e., it estimates where young people think that standard and dialectal languages are spoken.

The data were processed and analysed in two stages: 1) the maps of the questionnaire were scanned and georeferenced, the polygons of dialectal and standard language distribution from each map were digitized, and the aggregated response frequency raster layers were generated; 2) the questionnaire responses were used to perform a geospatial analysis and to produce interpolated spatial distribution layers (cf. similar approaches used in the perceptive dialectology studies (Montgomery, Stoeckle 2013: 52-85; Calaza et al. 2015: 282-291, etc.).

A fragmentary geo-spatial analysis of the material shows which areas are considered by the respondents to be the most dialectal and the least dialectal, the degree of prestige of the local language variant used by the younger generation, and the values of the associations of a dialect-speaking person. In addition, the results were compared with various socio-economic factors such as population, education, purchasing power, average age, ethnic composition.

Jaka Čibej
Faculty of Arts, University of Ljubljana, Slovenia
jaka.cibej@ff.uni-lj.si

CLUSTERING SLOVENE DIALECTS USING PHONEMIC PROFILES OF PHONETIC TRANSCRIPTIONS

Dialectology has been a very active branch of Slovene linguistics in the last century and the long tradition has resulted in many collections of dialectal recordings and transcriptions, which have been the basis for numerous dialectological studies (see e.g. Kenda-Jež 2002 for an overview). Despite this, the materials used have mostly been unavailable in a machine-readable format or even transcribed in full. However, a noteworthy effort in this regard has been made by the authors of the *Interactive Map of Slovene Dialectal Texts* (*Interaktivna karta slovenskih narečnih besedil*, IKNB), a website showcasing approximately 110 recordings of dialectal speech from different Slovene-speaking regions. The recordings are thematically similar (the informants were asked to describe old houses) and were collected and manually phonetically transcribed by students of Slovene linguistics at the Faculty of Arts (University of Ljubljana). The transcriptions have been assigned metadata on the basis of the *Map of Slovene Dialects* (Logar & Rigler 1983), which was based on the *Dialectological Map of Slovene* (Ramovš 1931).

Because of the lack of digital resources, Slovene dialectology has so far mostly relied on manual analyses, which are limited in scope. In computational and corpus linguistics as well as sociolinguistics, however, more advanced methods involving machine learning and statistical analysis have been used in recent years to infer insights from large quantities of data (Grieve 2012; an example for Slovene computational linguistics is described in e.g. Čibej 2021). Some of these methods have been implemented in dialectology and corpus dialectometry for other languages (e.g. Grieve 2016; Grieve et al. 2019; Donoso & Sánchez 2017; Doyle 2014).

In the paper, we present a computational analysis of Slovene phonetic transcriptions from IKNB. We first export relative frequencies of graphemic symbols of various phonemes and their combinations in order to compile vectorized representations of phonemic profiles of individual recordings. We perform k-means clustering to group similar transcriptions together, then compare the obtained clusters to the assigned metadata to identify potential discrepancies between automatic clusters and traditional dialectal categories. We also perform a statistical analysis of relative frequencies to identify the most salient phonemes (or combinations thereof) for each cluster.

While the method currently only takes phonemic profiles into account for a limited number of manually transcribed recordings, other features may be added in the future (on

more material) to provide a more accurate overview of Slovene dialects. We list several possibilities for future steps. In addition to recordings used for strictly dialectological purposes, an adapted version of the method could also be used on other transcriptions of speech recordings or possibly on quantified representations of sound recordings alone, bypassing transcriptions entirely and improving the scalability of the method. We expect the results to provide a new data-driven perspective on the current dialect classification system in use in Slovene dialectology.

Acknowledgments: The research described in this paper was conducted within the research project titled *Basic Research for the Development of Spoken Language Resources and Speech Technologies for the Slovenian Language* (J7-4642) and the research programme *Language Resources and Technologies for Slovene* (P6-0411), both funded by the Slovenian Research and Innovation Agency (ARIS).

Robert Engsterhold
Philipps-Universität Marburg, Research Center Deutscher Sprachatlas, Germany
engsterhold@uni-marburg.de

DIALECTOMETRY ON UNSUPERVISED DATA

This presentation addresses whether unsupervised, transcribed dialect data can be utilized effectively for dialectometric analyses. Traditionally, dialectometric studies rely on normalized datasets, enabling a focus on specific linguistic phenomena encoded in selected lemmas. However, datasets of written dialects exhibit significant heterogeneity (Lameli, Alfred, Elvira Glaser & Philipp Stöckle. 2020).

The “Wenkerbogen-App” project (Engsterhold, Robert & Elvira Glaser. 2023) provides a platform for transliterations of the so-called Wenker Sentences (WS). It also serves as a citizen science tool, allowing contributors to create their own transliterations using an online editor. To encourage participation, complex annotation rules were deliberately omitted, lowering the barrier for citizen scientists.

As of December 2024, 5,449 survey forms/locations comprising 170,495 individual sentences (approximately 10% of the WS corpus) have been transliterated. About half of this work was contributed by citizen scientists, with the remainder originating from scientific projects. The dataset is freely accessible via GitHub (Engsterhold, Robert. 2024).

Based on this material, the present study focuses on two primary research questions:

1. To what extent can known dialect structures be identified within the dataset?
2. Does the heterogeneity caused by diverse transliteration styles from multiple contributors significantly impact the results?

To this end, various quantitative analyses were conducted, with preprocessing minimized to ensure the algorithms ran stably: (1) Levenshtein Distance: Pairwise Levenshtein (Heeringa, Wilbert 2004) distances were calculated for individual sentences, then averaged across all sentences. The resulting distance matrices were visualized using clustering or ultidimensional scaling (MDS). Although computationally intensive, this approach is straightforward; (2) Bag-of-Words (BoW) Embeddings: Sentence embeddings were generated using {2,3}-grams (Birkenes, Magnus C Jürg Fleischer 2021), producing vectorized representations that are suited for analyses such as cosine similarity, dimensionality reduction, and clustering; (3) TF-IDF (Term Frequency–Inverse Document Frequency): A specialized variant of the BoW approach, TF-IDF embeddings were derived by weighting words inversely to their overall frequency in the dataset. These embeddings were subsequently analyzed using latent

semantic analysis (LSA) and clustering techniques. To identify optimal clustering configurations, the elbow method combined with silhouette scores was applied.

As will be shown, all methods produce regional clusters that broadly correspond to known dialectal regions. However, the influence of transliterator variability is evident, as higher numbers of clusters (larger k) were preferred by the elbow method. Notably, some clusters demonstrate high transliterator homogeneity, suggesting that individual transliteration styles significantly contribute to clustering outcomes. This effect can be mitigated by dimensionality reduction, which limits the amount of information encoded in the vector representations.

It is important to note that transliterator bias does not necessarily conflict with dialectal clustering. This is likely because citizen scientists tend to focus on regions, they are familiar with, which may align with dialectal boundaries.

Stefano Fiori
University of Turin, Italy
s.fiori@unito.it

Beatrice Colcuc
The Paris Lodron University of Salzburg, Ludwig Maximilian University of Munich, Germany
beatrice.colcuc@plus.ac.at

PERCEPTION AND SPATIAL CLASSIFICATION OF DIALECTS: A COMPARATIVE STUDY OF LADINIA AND THE QUATTRO PROVINCE THROUGH QUANTITATIVE ANALYSIS

This presentation seeks to provide a perceptual account of the spatial classification of dialect varieties in two distinct linguistic regions of the Italo-Romance domain. We compare data from two studies conducted with almost identical methodology and tools (Colcuc 2019 and Fiori 2023a), concerned, respectively, with the Ladin-speaking areas of Trentino Alto-Adige and Veneto and the Quattro Province, a region nestled in the Apennines north of Genova whose dialects are understudied and of difficult classification. Both studies employed an online questionnaire whereby respondents were asked to listen to short audio samples and place them on a map. The advantage of this free-placement procedure is that is allows the emergence of a "bottom-up", or "emic" geographical classification of linguistic varieties by laypeople (Krefeld & Pustka 2010), in contrast to forced-choice tasks, which rely on "etic" categories determined in advance by the researcher (Clopper & Pisoni 2006). However, the free classification task is not part of the usual five-point methodology devised by Preston (2002), and has been sparsely utilized in perceptual dialectology, most notably by Montgomery (2007) and, more recently, by Pinget & Voeten (2023).

In this contribution, we endeavor to quantify how respondents were able to differentiate the samples in space based on their perception, by calculating, for each pair of "point clouds" consisting of the placements of two dialect samples, the Bhattacharyya Coefficient (BC), a measure of overlap between two distributions, which is widely used in ecology to study land use by animals (Fieberg & Kochanny 2005), but has also been applied to linguistics, namely in acoustic phonetics, to measure vowel mergers (Stanley & Sneller 2023). Since the BC values range from 0 to 1, they can be interpreted as the (perceptual) distance between each pair of dialects represented by the samples; we arranged them in a symmetric matrix as a starting point for dialectometric analyses, such as probabilistic clustering and multidimensional scaling, which have been carried out with the Gabmap software (Nerbonne et al. 2011). Thus, this work falls into a line of research dedicated to apply dialectometry to perceptual data (see Goebl 2006 for Ladinia and also the recent

Calamai et al. 2022 on the perception of Tuscan varieties), and aims to provide a tool with which to compare “subjective” and “objective” linguistic classifications, both for areas where traditional geolinguistic and dialectometric literature are well established, such as Ladinia (Goebl 1981, Bauer 2009), as well as for less researched corners of the Italo-Romance domain, like the Quattro Province (Fiori 2023b).

Lea Fischbach
Research Center Deutscher Sprachatlas, Philipps-Universität Marburg, Germany
Lea.Fischbach@uni-marburg.de

Akbar Karimi
The Lamarr Institute for Machine Learning and Artificial Intelligence, University of Bonn, Germany
ak@bit.uni-bonn.de

Caroline Kleen
Research Center Deutscher Sprachatlas, Philipps-Universität Marburg, Germany
caroline.kleen@uni-marburg.de

Lucie Flek
Bonn-Aachen International Center for Information Technology (b-it), The Lamarr Institute for Machine Learning and Artificial Intelligence, University of Bonn, Germany
flek@bit.uni-bonn.de

Alfred Lameli
Research Center Deutscher Sprachatlas, Philipps-Universität Marburg, Germany
lameli@uni-marburg.de

DEEP LEARNING PIPELINE FOR GERMAN DIALECT RECOGNITION

The reliable classification of German dialects using Deep Learning (DL) methodologies, specifically Convolutional Neural Networks (CNN), without the need for manual transcription or other interventions in audio data, presents a significant challenge in the field of dialectology. Such a model not only offers efficiency in classifying German dialects but also holds the potential for seamless adaptation to dialects of other languages. Additionally, the long-term goal is to develop a classification system capable of analyzing speaker types within conversations, such as switchers, shifters and moveless. In dialectology, these terms generally describe different ways in which speakers vary or maintain their dialect usage depending on the context.

This study utilizes recordings from the Regionalsprache.de project (Schmidt & Herrgen & Kehrein & Lameli 2020ff), encompassing 20 dialects, classified according to Wiesinger (1983), from approximately 110 locations, 577 speakers across three generations (young, middle-aged, old), and five different recording scenarios. For this research, we

specifically utilized the recording scenario in which participants were required to translate 40 standardized sentences (“Wenkersätze”) from standard German into their respective dialects. The resulting dataset includes around 43 hours of usable audio.

Our pipeline comprises several key stages: preprocessing, automatic speaker diarization (ASD), data augmentation, feature extraction, and classification. Preprocessing ensures uniformity across the audio files. ASD separates and labels speakers, isolating the dialectal speaker from other speakers present in the audio (Fischbach 2024a). Data augmentation techniques such as pitch-shifting, segment removal, background noise addition, and segment swapping were employed based on established methodologies for text augmentation (Wei & Zou 2019). Additionally, voice conversion to a single target speaker was tested to enhance model robustness. Feature extraction was performed using Google’s TRILLsson model (Shor & Venugopalan 2022), leveraging transfer learning to obtain robust embeddings, which were then utilized to train a lightweight CNN. To account for speaker variability and prevent the model from learning speaker-specific features, we conducted multiple iterations where, in each run, different speakers were randomly assigned entirely to only one dataset partition (training, validation, or test). The complete pipeline is accessible on GitHub (Fischbach 2024b).

The model achieved a median weighted F1-score of 0.32, where the weighted F1-score balances precision and recall by considering the number of true instances for each class. Classification accuracy varied significantly across dialects, influenced by the number of available speakers and the similarity between dialects. Younger speakers were the least accurately classified (27.6%), likely due to their limited representation in the dataset and because they often do not master the dialect as proficiently. In contrast, middle-aged and older speakers achieved classification accuracies of 43.1% and 46.4%, respectively. A positive correlation was observed between speakers’ self-assessed dialect proficiency and classification performance.

Due to the variability in speaker and dialect selection, comparing these results with existing studies is challenging. Therefore, a future comparative study involving human dialect recognition rates is planned to establish a reference benchmark.

Januška Gostenčnik
Research Centre of the Slovene Academy of Sciences and Arts, Fran Ramovš Institute of the Slovene Language, Slovenia
januska.gostencnik@zrc-sazu.si

CLASSIFICATION OF THE SLOVENIAN LOCAL DIALECTS OUTSIDE NATIONAL BORDERS – A CASE STUDY

The contribution presents the methodology used in the research project Endangered Dialects in the Slovenian Linguistic Environment (the Bad Radkersburg Corner, Burgenland, Hum na Sutli and Surroundings, and Dubravica and Surroundings), which aimed to define in detail the borders of three Slovenian dialects outside Slovenia: the Prekmurje dialect in the Bad Radkersburg Corner and Burgenland in Austria, and the central Styrian and Kozjansko–Bizeljsko dialects in Croatia. In recent years, several clearly defined field studies using the same methodology had been carried out in the area of the Kostel and Čebranka dialects of the Lower Carniolan and Rovte dialect groups of the Slovenian language in the Gorski Kotar region of Croatia.

Dialectology as a part of genealogical linguistic deals with organic idioms, from their present (phonetical) state to their original (phonetical) state, i. e. their common (phonetical) denominator. Its main methods are the comparative method and the method of reconstruction. Its secondary task is to place a particular micro-organic idiom (i.e. the local dialect of a place) within a group of local dialects which may form a dialect at the next stage. In order to achieve this, it is necessary first to be familiar first with the synchronic state of the individual organic idiom, with the synchronic state of the neighbouring idiom, and so on, followed by the establishment of parallels, i.e. isoglosses. At this stage, these are isophones, i.e. isoglosses of phonetic phenomena, which are then grouped together into individual so-called isoglossic bundles.

So when determining the speaker's local dialect, it must be done independently of the speaker's place of residence in a particular country, the speaker's nationality, the literary language he or she (primarily) uses, and the speaker's self-perception of his or her local dialect.

Marija Gromova
Московский государственный университет М.В. Ломоносова, филологический факультет (Moskovska državna univerza M. V. Lomonosova, Filološka fakulteta), Rusija
marija.gromova@list.ru

STAVČNE NASLONKE NA ABSOLUTNEM ZAČETKU NEVPRAŠALNEGA STAVKA V SLOVENSKIH NAREČJIH

Prispevek analizira inovativni pojav slovenske narečne skladnje – umeščanje stavčnih (glagolskih in zaimenskih) naslonk na absolutni začetek nevprašalnih stavkov. Gradivo so sestavljali zapisi narečnega govora v obsegu 2505 strani iz 288 naselij. Analizirani so zapisi spontanega narečnega govora iz 107 znanstvenih del in skladenjsko zanesljivih poljudnoznanstvenih knjig iz zbirke *Glasovi*, zapisanih po pripovedi naravnih govorcev ustreznih narečij.

V prispevku so predstavljene značilnosti stavčnih naslonk in njihova sprejemljiva mesta v nevprašalnih stavkih v sodobnem slovenskem jeziku. Obravnavana je tudi zgodovina proučevanja problematike v slovenskem jezikoslovju.

S statistično in lingvogeografsko (arealno) metodo ugotavljam pogostost in največje število naslonk na absolutnem začetku nevprašalnih stavkov.

Na podlagi zbranih podatkov ugotavljam, da ostaja v vseh slovenskih narečjih, tako kot v sodobnem slovenskem knjižnem jeziku, glavno mesto stavčnih naslonk v nevprašalnem stavku položaj za prvo sestavino. V povprečju je pogostnost začetnega položaja naslonk 24 %, a se med narečji opazno razlikuje. Na absolutnem začetku nevprašalnih stavkov se naslonke nahajajo po celotnem slovenskem ozemlju in tvorijo neprekinjeno razpršeno območje, v katerem je mogoče ločiti cone razmeroma visoke frekvence. V začetnem položaju lahko najdemo tako glagolske kot zaimenske naslonke, pri čemer večina slovenskih narečij dovoljuje do štiri naslonke v nizu.

Aidas Gudaitis
Institute of the Lithuanian Language, Lithuania
aidas.gudaitis@lki.lt

Application of GIS and Geospatial Analysis Methods in Linguistic Landscape Studies for the Assessment of Ethnolinguistic Dynamics in Southeast Lithuania

When analysing the Lithuanian-Slavic ethnolinguistic dynamics in the multilingual region of Southeast Lithuania, it is necessary to acknowledge an obvious lack of quantitative studies. Understanding of ethnolinguistic processes is crucial for formulating appropriate education and regional development policies, as well as for assessing potential geopolitical threats. Most of ethnolinguistic research articles published on this region are fragmentary and primarily based on qualitative methods. A comprehensive quantitative study using modern information technologies would therefore contribute to a better understanding of language interactions and related processes in the region.

The aim of this study was to investigate the dynamics of ethnolinguistic boundaries over the past 100 years in the zone of intensive Lithuanian-Slavic ethnolinguistic interaction by applying GIS-based geospatial analysis methods.

A detailed study was conducted in the southern part of the Lithuanian-Slavic ethnolinguistic interaction zone, covering the territories of seven elderships across two municipalities. The selected study area spans 1,139 km² and has a current population of approximately 9,100 inhabitants.

Methodology: The study was based on the Linguistic Landscape approach as a method to explore how language use in a multilingual space represents ethnolinguistic groups. To gather more precise information about the linguistic attitudes of the population in the research area, a geospatial data model of tombstone inscriptions was developed. GIS-based information technology tools were employed for data collection, processing, and analysis. A mobile GIS application was created for the collection of gravestone inscriptions, enabling the rapid compilation of large amounts of geolinguistic data.

More than 6,500 samples of tombstone inscriptions were collected during two field surveys. The transcription of names, surnames, years of birth, and years of death written on the tombstones was carried out, and the language of each inscription was determined. The collected data were processed using statistical and geospatial analysis methods.

Results: The case study demonstrated that the collected data could be processed using the principles of quantitative analysis and is suitable for assessing ethnolinguistic dynamics from both temporal and spatial perspectives.

The analysis results showed that the language data from tombstone inscriptions correlated with statistical data on nationality and mother tongue from the population census. The percentage of tombstone inscriptions in Lithuanian in the study area was 62.4%, while, according to the 2021 census, 59.5% of the population in the study area identified themselves as Lithuanian.

The analysis revealed that the language of inscriptions in the study area has been shifting in a direction favourable to the Lithuanian language (Pearson's correlation coefficient $R=0.65$, $p<0.05$). It is statistically reasonable to conclude that the prestige of the Lithuanian language in the study area has been steadily increasing. The percentage of Lithuanian language inscriptions surpassed 65% in the last decade and is projected to reach 80% by 2040. Additionally, the Lithuanian-Slavic ethnolinguistic boundary has shifted 2–3 km to the southeast over the past 100 years.

The study demonstrated that the use of GIS-based quantitative methods enhanced the efficiency of data collection, processing, and evaluation. These methods allowed for a more objective assessment of trends in ethnolinguistic change.

Wilbert Heeringa
Fryske Akademy, the Netherlands
wheeringa@fryske-akademy.nl

Cesko Voeten
University of Amsterdam, the Netherlands
c.c.voeten@uva.nl

METHODS FOR PARTITIONING THE DIALECT CONTINUUM

Dialectology has long employed the practice of dividing dialect continua into distinct areas, often visualized through maps that differentiate these regions using lines or color coding. Notable examples include maps by Te Winkel (1901) and Daan & Blok, which highlight the various dialect areas within the Dutch continuum. These straightforward visual representations make the maps easily accessible and interpretable by a broad audience.
In this paper we consider several alternatives of creating similar maps dialectometrically. As a case study we use data from 50 locations and 166 items from the Series of Dutch Dialect Atlases. We calculated aggregated binary item distances. Then we applied five methods for partitioning the 50 variëties.

Method 1: Using this method, a dendrogram is created by hierarchical cluster analysis on the basis of the aggregated distances among the local dialects. Areas can be derived from the dendrogram by drawing a vertical line somewhere in the dendrogram and counting the horizontal lines matching it (see Figure 1). Local dialects connecting to the matching horizontal lines will then belong to the same cluster (or group or area). The vertical line should be drawn between the two successive nodes that are most distant to each other (see also https://statsandr.com/blog/files/Hierarchical-clustering-cheatsheet.pdf).

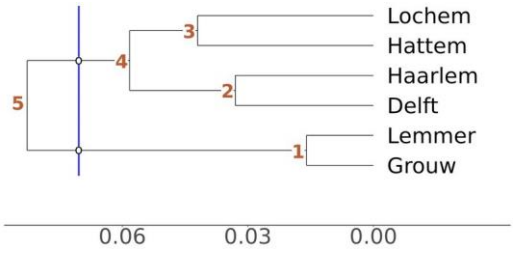


Figure 1: The largest distance between two successive nodes is between nodes 4 and 5. A vertical line is drawn in the middle of these two nodes and intersects with two horizontal lines that connect with two groups.

Method 2: This method uses bootstrap clustering to find dialect groups. It involves resampling data, performing hierarchical clustering to identify natural groups with the elbow method, and counting how often dialects co-occur in the same group. Local dialects are marked as connected if they appear in the same group in over 95% of iterations, resulting in networks that represent dialect groupings (Heeringa 2017).

Method 3: Similar as method 1, but instead of resampling data, noise is added to the distances.

Method 4: Affinity Propagation (AP) is a clustering algorithm that selects representative data points as exemplars and groups other data points into clusters based on their similarity to these exemplars.

Exemplars are actual points from the dataset. In contrast to k-means there is no need to specify the number of clusters in advance (Frey & Dueck 2007).

Method 5: DBSCAN is an algorithm that groups together points that are closely packed while marking points in low-density regions as noise (Ester 1996). HDBSCAN is generally considered more robust than DBSCAN, particularly in handling datasets with varying densities and noisy data (McInnes et al. 2017). The latter method only requires the user to specify the minimum number of points necessary to form a cluster.

We evaluate the results by comparing the partitions with the original distance measurements using the Silhouette score (Rousseeuw 1987). Figure 2 shows that partitions with a higher Silhouette score reflect the beam map more faithfully.

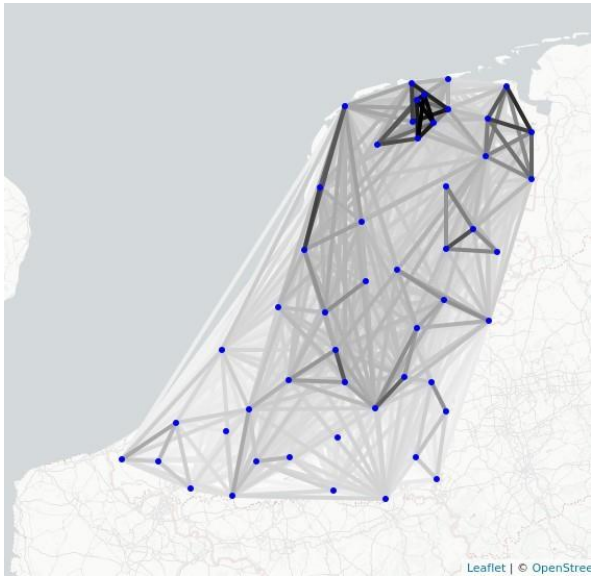


Figure 2a: Beam map: darker lines represent smaller distances, lighter lines represent larger distances.

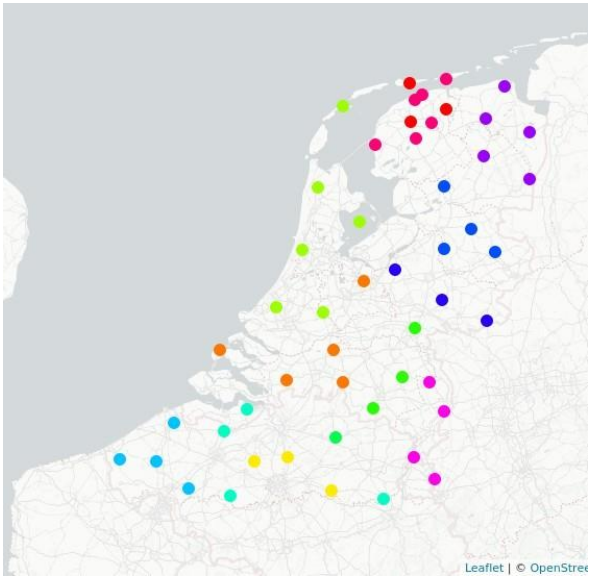


Figure 2b: Affinity Propagation. Silhouette score = 0.1282.

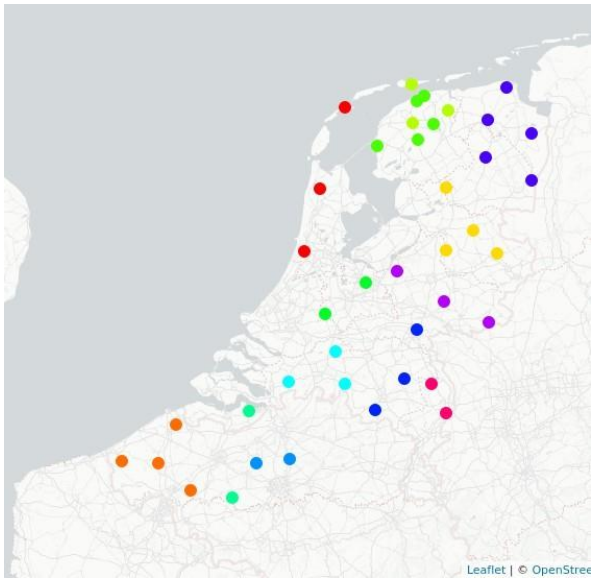


Figure 2c: HDBSCAN. Silhouette score = 0.1939.

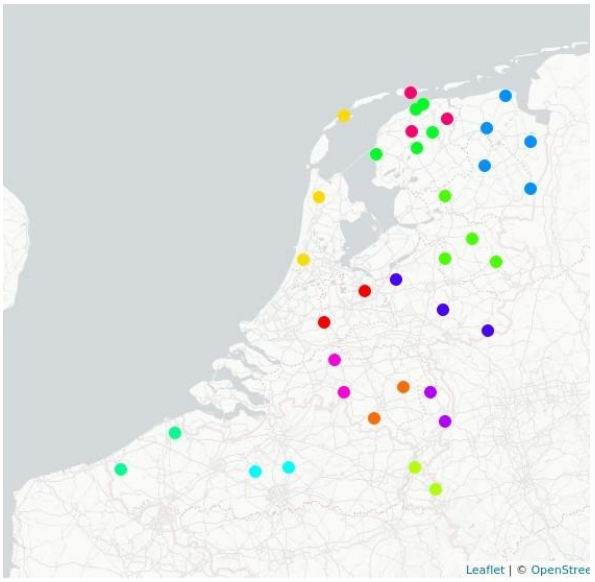


Figure 2d: Bootstrap Ward's clustering. Silhouette score = 0.2157.

Wilbert Heeringa
Fryske Akademy, the Netherlands
wheeringa@fryske-akademy.nl

Vincent van Heuven
Leiden University, the Netherlands
v.j.j.p.van.heuven@hum.leidenuniv.nl

Charlotte Gooskens
University of Groningen, the Netherlands
c.s.gooskens@rug.nl

FURTHER IMPROVEMENTS IN ACOUSTIC DIALECTOMETRY

The paper explores advancements in acoustic dialectometry, focusing on improving methods for measuring pronunciation distances among dialects. Traditional dialectometry relies on variables, phonetic transcriptions or text corpora, with Levenshtein distance as a popular algorithm for quantifying phonetic differences. However, the latter method may be influenced by inconsistencies in transcriptions. Therefore, Bartelds et al. (2020: 1-10) introduced the use of dynamic time warping (DTW) with Mel-frequency cepstral coefficients (MFCCs) derived from acoustic data to overcome this issue. Bartelds et al. (2022: 1-17) enhanced the approach using wav2vec representations, and found that the results more closely matched human similarity judgments.

Our study demonstrates that the original MFCC-based method can be refined to achieve results superior to those of wav2vec. We propose two main improvements. First, to mitigate the disruptive effects of gender differences, recordings by male speakers were transformed to resemble those by female speakers using Praat's gender changer, employing a formant shift ratio of 1.2 and doubling the median pitch. Second, the MFCC parameters were optimized to maximize correlation with perceptual distances derived from human judgments. We used recordings of the fable *The North Wind and the Sun* in 15 Norwegian dialects (https://www.hf.ntnu.no/nos/). Students in the corresponding regions rated the similarity of each recording to their native dialect on a 1–10 scale, generating a 15×15 perceptual distance matrix (Gooskens & Heeringa 2004: 189-207). Before measuring acoustic distances, the dialect recordings were split into a maximum of 58 word samples. By repeatedly applying the function melfcc from the R package tuneR to this data and correlating the results with the perceptual measurements, the parameters of the function were optimized. To reduce noise, we standardized the MFCC coefficients by removing their

mean and dividing by the standard deviation (i.e., z-transformation) across time and for each word sample (Ganapathy et al. 2011; Shafik et al. 2009: 52-58).

Unlike Bartelds et al. (2020: 1-10), who computed distances using Euclidean metrics, we used 1 minus Pearson's correlation, yielding interpretable distances between 0 and 1. We also used a more precise normalization method for DTW distances, dividing by the alignment length rather than by the sum of the representations' lengths.

The results (Table 1), show that the classical Levenshtein distance still outperforms DTW-based approaches in correlating with perceptual distances. However, our enhanced DTW method, particularly when incorporating gender normalization, are almost as effective and can be used in cases where phonetic transcriptions are unavailable or inconsistent. The DTW approach proposed in this paper is available in the web app LED-A (see https://www.led-a.org/docs/DTW.pdf and https://www.led-a.org/#examples under 'Text To Speech -- Germanic').

method	correlation r
Bartelds et al., 2020, DTW/MFCCs[1]	0.22
Bartelds et al., 2022, DTW/wav2vec[2]	0.49
DTW/MFCC, approach proposed in this paper	0.62
Plain Levenshtein distance[3]	0.68

Table 1: Pearson's correlation coefficients between four dialectometric distance measurements and perceptual distances that were obtained by using recordings of 15 Norwegian dialect translations of the fable *The North Wind and the Sun*.

[1] This correlation was reported by Bartelds et al. (2022).

[2] Actually two versions were used, namely DTW/w2v2-en, DTW/XLSR-sv; for both versions the same correlation was found.

[3] Substitutions had a weight of 1, insertions and deletions had a weight of 0.5.

Roeland van Hout
Center for Language Studies, Radboud University, Nijmegen, the Netherlands
roeland.vanhout@ru.nl

John Huisman,
Uppsala University, Sweden
john.huisman@lingfil.uu.se

Frens Bakker
Radboud University, Nijmegen, the Netherlands
frens.bakker@ru.nl

CLASSIFICATIONS AND CONTINUUMS: A UNIFIED ANALYSIS OF VARIATION ACROSS COUNTRIES AND DIALECT AREAS

We investigated the impact of geographic distance, a continuous variable, and discrete boundaries on linguistic differences across the North-South Lower Franconian borderland in the Netherlands and Germany. It is well known that linguistic differences gradually accumulate over geographic distance, but how does a linguistic continuum interact with discrete boundaries marking distinct dialect or language areas? In this contribution we will unify the analysis of the continuous character of linguistic distances and discrete language boundaries.

The area under investigation here has two such discrete boundaries, one linguistic (a dialect isogloss) and one non-linguistic. The linguistic boundary is the Uerdingen line, which is part of the Rhenish Fan and marks the northernmost extent of the High German consonant shift - with the shifted forms *ich* (vs. *ik*) '1.SG' and *auch* (vs. *ook*) 'also'. The Uerdingen line reflects the linguistic landscape before 1815. In that year, the national border between the Netherlands and then Prussia (Germany, from 1871 onwards) was established. This non-linguistic boundary put speakers in a new sociocultural situation, with different national identities, and education in different dominant standard languages: Dutch in the Netherlands, and German in the Germany. To what extent have these two boundaries influenced the linguistic differences we observe, and how do they interact with distance as a driver of linguistic differentiation?

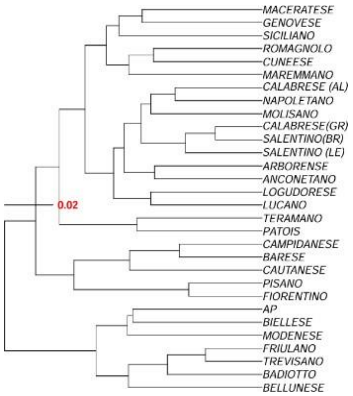
We analysed two datasets with dialectometric techniques containing extensive data for 44 linguistic varieties on both sides of the national border: one dataset comprising basic vocabulary, the other comprising the personal pronoun vocabulary. These data came from older dialect surveys in the Netherlands and Germany, including the Wenker questionnaire in

Germany. We computed pairwise linguistic distances between all varieties and conducted linear mixed-effects regression to assess the relative strengths of geographic distance, separation by linguistic boundary, and separation by national border. While geographic distance and separation by a linguistic boundary played a significant role in both datasets, the impact of the national border differed between the two. For the basic vocabulary, the national border significantly predicted linguistic distance, suggesting the effect of standard language contact on (part of) the lexicon. For the (generally more stable) pronoun vocabulary, the national border was not a significant predictor, indicating that new boundaries need time before their effects are found in all parts of the language system. Our results exemplify that a unified dialectometric analysis of continuity and discrete boundaries is possible and that such an analysis provides the best approach in analysing linguistic variation across countries and dialect areas.

Elena Isolani
University of Cambridge, United Kingdom
ei259@cam.ac.uk

WHEN THE PARAMETRIC COMPARISON MEETS THE CP: A PRELIMINARY TAXONOMY OF ITALO-ROMANCE VARIETIES

Background & Aims: The Parametric Comparison Method (PCM) is an innovative tool for language comparison aimed at reconstructing linguistic phylogeny and developing new linguistic taxonomies based on syntactic parameters (Longobardi 2005; Longobardi & Guardiano 2009, 2017; Guardiano et al. 2020). While previous studies involved the NP/DP and standard languages (Crisma et al. 2020), this study delves into an unexplored syntactic domain, the complementizer phrase (CP), selecting a language sample of Italo-Romance varieties. The goals are: (1) proving that the PCM can extend to other structural modules, given an effective parametric database, (2) categorizing the micro-variation in Italo-Romance varieties through a reliable parametric tool that captures minimal variational patterns and offering a revised taxonomy.
Methods & Results: The parametric database regularizing the CP was framed within the cartographic approach (Rizzi 1997). Each head of the split-CP was independently treated to identify salient variational patterns and the respective parameters regulating them. Core parameters were constructed according to a parameter schema (Longobardi 2005; Gianollo et al. 2008) asking whether a feature F is grammaticalized, checked, spread or strong. The valuation of parameters in the schema sheds light on other structural phenomena, displaying a significant cross-linguistic interplay among the languages tested.



Contrary to the previous PCM applications, this study prioritized non-standard varieties to determine the relatedness between languages that are unequivocally close given

their common ancestor. Thirty dialects from different areas of Italy were investigated, collecting data through a recorded oral translation task.

Each parameter received a value (+, -, 0) for each language tested and the syntactic distance for each language pair was computed. Results, analysed via the UPGMA method, align broadly with traditional classifications but highlight notable exceptions: (1) Genovese, Cuneese, and Romagnolo clustering within SIDs, (2) Sardinian varieties' distribution, (3) Tuscan varieties' distribution, (4) Valdôtain Patois idiosyncrasies.

Discussion: The phylogenetic tree allows some preliminary conclusions, alongside additional hypotheses for further research. First, the efficacy of the PCM in detecting cross-linguistic variation and in producing a reliable taxonomy is attested in relation to the CP as well. Second, by comparing Guardiano et al. (2020)'s analysis of the nominal domain on a set of Italo-Romance with the tree represented above, the DP/NP looks more compact, whereas the CP reveal several inconsistencies, predicting its higher sensitivity to cross-linguistic variation. Third, it is reasonable to advance a new classification of Italo-Romance based on the CP, in addition to the traditional taxonomies, mainly based on phonology or morphology. Lastly, this quantitative analysis seeks to uncover variational patterns that can serve as a foundation for typological analysis. A key, yet underdeveloped step, is to understand the source of this variation, explaining why the CP classification appears as it does.

Conclusions: Although the investigation is still in its early stages, the results are compelling and offer insights into the linguistic evolution of Italo-Romance. Notably, they reveal several unexpected patterns of variation, which have been effectively analysed using a detailed parametric approach. Further research could delve deeper into this classification to uncover the fundamental sources of variation.

Paweł Janczulewicz
Instytut Języka Polskiego Polskiej Akademii Nauk (Inštitut za poljski jezik Poljske akademije znanosti), Poljska
pawel.janczulewicz@ijp.pan.pl

Rafał Szeptyński
Instytut Języka Polskiego Polskiej Akademii Nauk (Inštitut za poljski jezik Poljske akademije znanosti), Poljska
rafal.szeptynski@ijp.pan.pl

ZACHODNIOSŁOWIAŃSKI MATERIAŁ DO REKONSTRUKCJI SUFIKSU WERBALNEGO *-*NY*-

Dla języka prasłowiańskiego rekonstruuje się sufiks werbalny *-*ny*- (obok lepiej poświadczonego wariantu *-*nǫ*- < *-*nu*-) w oparciu o domniemaną zgodność pewnych faktów zachodniosłowiańskich z danymi zachodniej podgrupy języków południowosłowiańskich (słoweński, dialekty bośniackie/chorwackie/czarnogórskie/serbskie). O ile południowosłowiańska postać *-ni-* pozwalałaby rekonstruować zarówno praformę *-*ni*-, jak i *-*ny*-, różnorodne postaci zachodniosłowiańskie wskazywałyby wyłącznie na tę drugą możliwość (w związku z zachowaniem opozycji między refleksami psł. **y* oraz **i*). Celem referatu jest weryfikacja twierdzeń o odzwierciedleniu w materiale zachodniosłowiańskim wariantu *-*ny*-.

W referacie poddamy analizie – częściowo w oparciu o starszą literaturę – wszystkie najważniejsze fakty zachodniosłowiańskie: górnołużyckie, polsko-pomorskie oraz połabskie. W pierwszych dwóch wypadkach analiza obejmie zarówno materiał dialektologiczny, jak i filologiczny; w trzecim z konieczności będzie się ona ograniczać do materiału filologicznego (pomimo obecnych w literaturze sugestii o odzwierciedleniu w interesujących nas zapisach zróżnicowania dialektalnego). Naszym zdaniem fakty żywych języków dają się lepiej objaśnić jako rezultaty niezależnych przekształceń innych postaci sufiksu: unosowionego *-*nǫ*- lub ewentualnie – w górnołużyckim – starszego *-*nu*-. Natomiast dla wymarłego języka połabskiego zachodzi uzasadnione podejrzenie o sztucznym charakterze zapisów interpretowanych dotąd jako refleksy wariantów innych niż pewnie poświadczony *-*nǫ*-.

Wątpliwości względem danych zachodniosłowiańskich otwierają drogę do dalszej rewizji rekonstrukcji psł. *-*ny*-. Skoro dane południowosłowiańskie nie wykluczają barwy *i* jako pierwotnej, należy się liczyć z możliwością lokalnego wpływu typu koniugacyjnego na *-i-* (jako podstawa analogii wystarczające mogły tu być wspólne formy rozkaźnika na *-i*).

Irena Jaros
Uniwersytet Łódzki, Wydział Filologiczny (Univerza v Lodžu, Filološka fakulteta), Poljska
irena.jaros@uni.lodz.pl

PODZIAŁ POLSKICH DIALEKTÓW WEDŁUG PROF. KAROLA DEJNY – HISTORIA I WSPÓŁCZESNOŚĆ

Autorka w pierwszej części artykułu przedstawia podział dialektów polskich autorstwa prof. Karola Dejny, pochodzący z lat siedemdziesiątych XX wieku, oparty na etniczno-językowych kryteriach. Omawia różnice w zakresie zasięgów terytorialnych poszczególnych zespołów dialektalnych w stosunku do wcześniejszych koncepcji podziału polskiego obszaru językowego, wskazuje tereny sporne. W drugiej części wystąpienia omawia zmiany, jakie dokonały się w zakresie omawianej klasyfikacji polskich dialektów (casus języka kaszubskiego) oraz przedstawia współczesne problemy z ugrupowaniem polskich gwar i dialektów.

5 Slovenski dialektološki posvet

Taja Jereb
Univerza v Ljubljani, Filozofska fakulteta, Slovenija
jerebtaja@gmail.com

HIŠNA IN LEDINSKA IMENA V VASI PLUŽNJE

V prispevku obravnavam hišna in ledinska imena v vasi Plužnje, ki leži na območju cerkljanskega narečja rovtarske narečne skupine. Glavni cilj raziskave je bil popis in zbiranje gradiva za nadaljnje jezikoslovne študije ter preučitev najpogostejših načinov tvorjenja tovrstnih imen s posebnim poudarkom na morebitnih posebnostih v njihovem nastanku in pomenu. Osrednji vprašanji sta bili: kakšna hišna in ledinska imena se pojavljajo v vasi Plužnje ter kakšni so glavni vplivi na njihovo tvorbo in uporabo?

Teoretični okvir raziskave temelji na monografiji *Zemljepisna in osebna lastna imena v kraju Livek in njegovi okolici* Mateja Šeklija, metodologijo pa sem povzela iz dela *Metode zbiranja hišnih in ledinskih imen* (Klinar idr. 2012). Terensko delo je vključevalo intervjuje z lokalnimi govorci, zbiranje podatkov o približno 75 imenih, ki sem jih nato fonetično transkribirala, poknjižila ter analizirala njihov izvor in pomen.

Rezultati kažejo, da so hišna imena, kot pričakovano, najpogosteje nastala iz osebnih imen, vzdevkov, priimkov ali poklicev gospodarjev in gospodinj, medtem ko so ledinska imena pogosto povezana z naravnimi značilnostmi, lego ali uporabo zemljišč. Posebnosti vključujejo vplive lokalnih narečnih značilnosti ter zgodovinskih in družbenih dejavnikov.

Prispevek poudarja pomen sistematičnega zbiranja in analize lokalnih poimenovanj kot pomembnega vira za dialektološke in onomastične študije.

Ana Marija Jevšenak
Univerza v Ljubljani, Filozofska fakulteta, Slovenija
anamarija.jevsenak@gmail.com

LEDINSKA IMENA V POTOPLJENI VASI DRUŽMIRJE

Družmirje je bila obcestna vas v zahodnem delu Šaleške doline, ki so jo zaradi ugrezanja območja kot posledice izkopavanja lignita porušili. Ugreznine je zalila voda, nastalo je Družmirsko jezero. V prispevku bodo predstavljena ledinska imena, ki so bila zbrana na tem, danes pretežno že potopljenem območju. Zbrana so bila s pomočjo informatorjev, ki so nekdaj živeli na tem območju, zemljevidov in zemljiških katastrov ter ankete o sedanjem poznavanju ledinskih imen na obravnavanem območju. Ledinska imena so zapisana in predstavljena glede na pomensko motivacijo, predstavljena v obliki slovarskih sestavkov ter prikazana na zemljevidu. Uporabljena je bila metodologija projekta FLU-LED.

Janoš Ježovnik
ZRC SAZU, Fran Ramovš Institute of the Slovenian Language, Slovenia
janos.jezovnik@zrc-sazu.si

THE MORPHOLOGICAL APPROACH TO THE CLASSIFICATION OF SLOVENIAN DIALECTS

The contribution describes an approach to the study of phonology and accentology in Slovenian dialects. Slovenian is a morphologically complex language characterised by a paradigmatic distribution (or, at the very least, remnants thereof) of accentual features – i.e., stress placement, vowel length and, where conserved, pitch, as well as vowel quality, as all of these are intrinsically intertwined – stemming from Common Slovenian (CSln.). Single-form lexical records, often featured in fieldwork questionnaires, are generally insufficient for a thorough understanding of synchronic characteristics and the genesis of Slovenian dialectal accentual features. Rather, a lexeme should always be regarded in the context of its corresponding inflectional pattern to account for potential intraparadigmatic differences, some of them inherited from CSln. and some results of later dialectal changes (e.g. the so-called *brata*-lengthening). On the other hand, comparison of accentual patterns (»accentual curves«, according to Dyboan terminology) of different paradigms is essential for differentiating between regular sound changes and possible analogies.

As Slovenian dialect classification is based primarily upon diachronic developments of certain cardinal vowel sounds (i.e. the diphthongisation of the CSln. *yat*, the merger (or not) of the CSln. nasal and mid-central vowels, the periodization of the *brata*-lengthening, the vocalisation of the long CSln. schwa and its merger either with **a* or **e*, to list only the most important factors), the correct identification of these sound changes is crucial for the correct classification of spoken local varieties, especially those situated close to or at dialect borders. The approach will be illustrated mostly using examples from the *Tersko* dialect of Slovenian.

Janoš Ježovnik
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
janos.jezovnik@zrc-sazu.si

Nace Pušnik
Institute of Creativity and Innovation, University for the Creative Arts, Xiamen University, Hongkong
nace.pusnik@uca.ac.uk

ZRCALO – ZRCOLA: NOVOSTI PRI VNAŠALNEM SISTEMU IN PRENOVA ČRKOVNE VRSTE

Vnašalni sistem ZRCola 2 uporabniku omogoča, da med pisanjem osnovne črkovne znake kombinira s poljubnimi diakritičnimi znaki in tako v besedilo na preprost način dodaja kompleksne znake. Sistem sestavljata vnašalni program in pisava ZRCola; kot sistem makrov v programu Microsoft Word ga je primarno za potrebe priprave *Slovanskega lingvističnega atlasa* (OLA) razvil in do leta 2015 vzdrževal Peter Weiss (prim. Weiss 2004). S temeljito prenovo v letu 2016 je vnašalni program postal neodvisen od programskega okolja, od tedaj pa je doživel nekaj posodobitev, ki še niso bile celovito predstavljene znanstveni javnosti. Prav tako se je leta 2023 začel projekt celostne oblikovne prenove črkovne vrste ZRCola. Črkovna vrsta *ZRCalo* je sodobno zasnovana neserifna pisava, ki bo vzporedno s pisavo *ZRCola* prirejena za uporabo v kombinaciji z vnašalnim programom.
V prispevku bodo predstavljene nekatere novosti v vnašalnem programu, npr. podpora za prečrkovanje, izpopolnjen iskalnik znakov, boljša podpora vnosu poljubnih sestavljivih znakov v skladu s standardom Unicode … Prav tako bodo predstavljena izhodišča za oblikovanje nove črkovne vrste, dosedanji koraki in trenutno stanje.

Darja Jonjic
Christian-Albrechts-Universität, Kiel, Germany
djonjic@roots.uni-kiel.de

Søren Wichmann
Christian-Albrechts-Universität, Kiel, Germany
wichmannsoeren@gmail.com

Matías Guzmán Naranjo
University of Freiburg, Germany
mguzmann89@gmail.com

QUANTITATIVE CLASSIFICATION OF RUSSIAN DIALECTS

This study presents a new quantitative classification of Russian dialects using *The Dialectological Atlas of the Russian Language* (DARJa). Published in Moscow from 1986 to 2005, DARJa represents over four decades of linguistic data collection, and includes 313 maps – each corresponding to a specific linguistic feature – covering 4196 locations. In 2015-16, researchers at Kazan Federal University extracted linguistic features and their values directly from the physical maps and created Excel files giving feature values across locations (Isaev et al. 2016).

We have processed these materials further, georeferencing the map of locations covered and manually extracting latitude and longitude coordinates of every location. Subsequently we have applied dialectometric methods based on interpolation of all feature values using the Hamming distance and hierarchical clustering, aiming to derive clusters that reflect dialectal divisions across the Russian linguistic landscape.

Central to this research is our new method for determining the optimal number of dialect clusters, addressing a gap in existing dialectometric studies. While approaches like Marchenko's MDS-based classification into six zones (2023) and Pšeničnova's structural-typological analysis with five zones (1996) provide valuable frameworks, they lack a systematic way to objectively identify the optimal number of clusters. To address this, we introduce a new method, applying a process in which we select a number of *k* clusters and evaluate the fit of the classification by calculating the Adjusted Rand Index (ARI) for each linguistic feature separately. By averaging the ARI across all features, we get a measure of how well each choice of *k* clusters fits the data. This iterative ARI-based method enables us to identify the optimal number of clusters, which in the case of DARJa is five.

These results advance the field of dialect classification by providing a replicable and objective clustering method, which can be applied to dialect continua in various linguistic contexts, combining both quantitative techniques and linguistic analysis for more precise classifications.

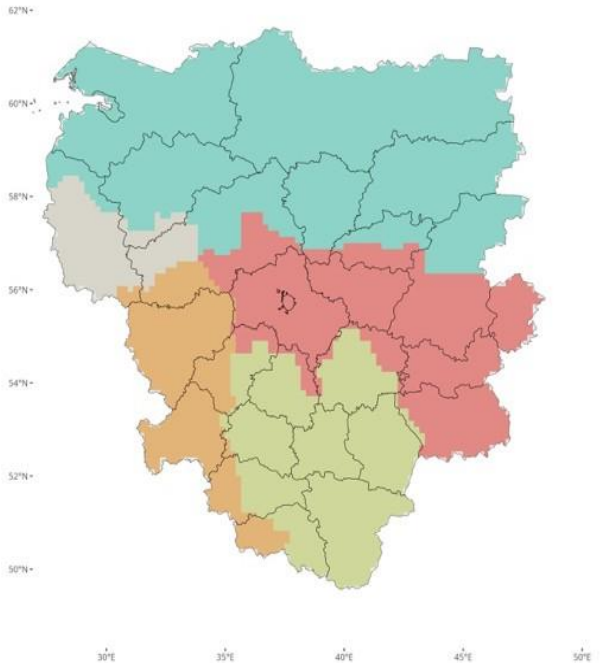


Figure 1: Map using interpolation and clustering on the DARJa data showing the optimal number of five clusters.

Karmen Kenda-Jež
ZRC SAZU, Fran Ramovš Institute of the Slovenian Language, Slovenia
carmen.kenda-jez@zrc-sazu.si

BACK TO THE FUTURE: THE ROLE OF PHONETIC VARIATION IN MODERN CLASSIFICATION OF SLOVENE DIALECTS

Despite the fact that the holistic prestructuralist approach of Fran Ramovš, the predecessor of today's dialect classification of Slovene and author of the Map of Slovene Dialects (1931), was criticised (Rigler 1975) and gradually abandoned by his successors – in the 1960s they abandoned various extralinguistic criteria, in the last decades the relevance of the so-called "(general) acoustic impression" (differences in "articulation, rhythm and tempo of speech") was reduced or rejected – the underlying linguistic criteria remain more or less the same. Therefore, the new boundaries on further versions of the Slovene dialect maps (Logar – Rigler 1983; Škofic et al. 2010, 2016, 2023) are predominantly the result of corrections to Ramovš's model based on new dialect data from underdescribed dialect areas, without systematic comparison of the classification principles of different authors over time.

The primary tool for a detailed classification analysis of Slovene dialects is phonetic tables of Common Slovene reflexes, from the early fieldwork expeditions of Ramovš and Logar (according to the collected manuscript material) to Šekli's (2009, 2018) isophonic genealogical classification, illustrated with the reflexes of the central Common Slovene vowels *$\breve{e}$, *$\grave{\breve{e}}$-, *$\bar{o}$, *$\grave{o}$, *$\bar{e}$, *$\grave{e}$-, *$\bar{\varrho}$, *ϱ-, *$\bar{o}$, *$\grave{o}$-, *$\bar{\partial}$, *$\grave{\partial}$-.

The first geolinguistic representation of the above reflexes used for the most recent dialect classification (Šekli 2018) – the set of individual and synthetic phonetic maps for the Slovene Linguistic Atlas (SLA) – aims to (1) show the current overlap of the phonetic data areas with the actual dialect boundaries (Škofic et al. 2023), (2) define the problematic areas (with the reconstruction of possible differences in the classification approaches during the process of map creation), (3) to create a plan for the proposal of the phonetic part of the project The Map of Slovene Dialects (3rd Generation).

The representation is a result of research, conducted within the research project titled *Basic research for the Development of Spoken Language Resources and Speech Technologies for the Slovenian Language* (J7-4642) funded by the *Slovenian Research and Innovation Agency* (ARIS).

Nataša Kolega
Rubin geoinformatika d. o. o., Slovenia
natasa@rubin.si

Jure Grahor
Rubin geoinformatika d. o. o., Slovenia
jure@rubin.si

DIATLAS – GEOSPATIAL DATABASE AND APPLICATION FOR DIALECTOLOGY

Primary objective for the development of DIAtlas was to cover the needs of Dialectology section of Fran Ramovš Institute of the Slovenian Language emerged from their continuous work on Slovenian Linguistic Atlas (SLA) but, as it was developed, it aims to cover also the needs of other dialectological work (different atlases, dictionaries and publications). SLA is composed of approximately 3000 questions and primary network of approximately 400 places in Slovenia, for which the data was collected through the years. Results are represented on dialectological maps with additional comments. DIAtlas is actually an answer for the need of developing a custom spatial geodatabase capable to manage huge amount of collected spatio-linguistic data and its various uses.

The development of DIAtlas started in 2022 and is still developing and upgrading. It is composed from database, web user interface and QGIS for map creation. It uses PostgreSQL/Postgis database in which we save dialectological data, spatial data, metadata, dialectology data sources (for example scans of notebooks with older dialect data, pictures, recordings etc.), created maps and additional geographic layers. For GUI we use stateless web user interface based on Django/Python code. Through the years the majority of data was collected in Excel tables, which are imported into DIAtlas with custom created procedures using Pandas.

DIAtlas also allows users to export their work for publishing by generating MS Word documents (for example map comments, indexes etc.) and QGIS projects (various types of dialectological maps) inside DIAtlas.

DIAtlas is intended for internal use of Dialectology section of Fran Ramovš Institute of the Slovenian Language; for public use we developed Interactive Slovenian Linguistic Atlas (iSLA) with integrated crowdsourcing possibility for public contribution of dialectology data.

DIAtlas is a powerful tool, highly adapted and customized for unique needs of dialectology research and can be used (as well as customized) for all other languages as well, not only Slovenian.

Katarzyna Konczewska
Instytut Języka Polskiego Polskiej Akademii Nauk (Inštitut za poljski jezik Poljske akademije znanosti), Poljska
katarzyna.konczewska@ijp.pan.pl

Диалекты польско-белорусско-литовского пограничья: Проблемы классификации и пути их решения

Современное польско-белорусско-литовское пограничье представляет собой зону не только латеральных (периферийных) балтийско-славянских контактов, но и активных языковых процессов как между различными группами (польский, белорусский, русский, литовский), так и внутри них. Гипотезы о процессе балто-славянских языковых контактов в латеральных зонах являются ключевым аргументом в изучении циркумбалтийского региона (Koptjevskaja-Tamm & Wälchli 2001: 615-750; Shmrecsanyi 2014: 81-92) и предметом активной научной дискуссии (Wiemer, Seržant & Erker 2014: 15-42). Процесс субстратного заимствования проходил здесь в условиях дву-, а иногда и трехъязычия, поэтому субстратно-адстратные отношения требуют чрезвычайно осторожного подхода, а определение и классификация диалектов значительно затруднены.

Установление политических границ после Второй мировой войны искусственно разделило единое некогда пространство и усложнило его изучение. Проблематика влияния границ на процессы языковой конвергенции и дивергенции является весьма актуальной (Auer, Hinskens & Kerswill 2005; Palander, Riionheimo & Koivisto 2018; Luděk 2023: 59-100). Изучение диалектов и контактов между ними в пограничных ареалах с неоднородной исторической и современной социолингвистической ситуацией является сложной задачей. По нашему мнению, наиболее продуктивным методом ее решения является микроареальный подход. Исследование микроареалов позволяет выделить локальные языковые особенности, проследить направления и глубину проникновения диалектов, определить процессы формирования языковых инноваций и установить ареалы моносемантов.

В выступлении мы обратимся к проблеме классификации диалектов в латеральных зонах на примере двух исследуемых нами микроареалов, на польско-белорусском и белорусско-литовском пограничье, и укажем на возможные пути решения задачи. Особое внимание уделим социолингвистическим переменным, поскольку результаты наших исследований отмечают выраженную корреляцию между современной социолингвистической ситуацией и историческими процессами освоения микроареалов, а также особенностями их сословной составляющей. Покажем, что при исследованиях пограничных ареалов наиболее продуктивным является сочетание

традиционной лингвистической методологии с наработками этнографов, антропологов и этноисториков (ср. Wiemer 2003: 212-229; Costley & Reilly 2021: 1035-1047) и с поиском новых методов полевых экспедиций, которые представим на примере наших наработок. Продемонстрируем, что анализ собранного нами во время полевых экспедиций лексического материала позволяет сделать выводы, что диалектный ландшафт пограничья косвенно отражает историческую изменчивость политических и культурных границ, а изоглоссы в пограничных ареалах следует рассматривать как границы интенсивности языковых контактов, связанные с фактическими географическими и политическими границами и границами культурных влияний. Представленные нами методы исследования диалектов пограничных микроареалов могут служить для изучения процессов инфильтрации в балто-славянском пограничном регионе (ср. Wiemer & Erker 2011: 184-216), сопоставительного описания диалектов в латеральных зонах (ср. Mikulėnienė 2015: 282-301) и определения влияния политических факторов на процессы языковой конвергенции и дивергенции, а также могут помочь получить новые данные о языковых контактах в малоизученных латеральных диалектных ареалах и описать регионолект пограничья.

Mojca Kumin Horvat
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
mojca.horvat@zrc-sazu.si

KLASIFIKACIJA SLOVENSKIH GOVOROV V RADGONSKEM KOTU V AVSTRIJI

Radgonski kot v Avstriji je območje, ki je na dveh straneh ločeno od Republike Slovenije, in sicer na eni strani z reko Muro, na drugi strani s potokom Kučnica. Z vidika narečjeslovne klasifikacije se govori tega območja uvrščajo v panonsko narečno skupino slovenskega jezika, kar potrjujejo tudi dosedanje, sicer maloštevilne obravnave nekaterih govorov Radgonskega kota (npr. Haberl Zemljič (2012), Križman (1989, 1997, 2006), Zorko (1988, 1989)). V referatu bodo predstavljeni rezultati narečjeslovne raziskave, ki je v letih 2021–2024 v sklopu projekta *Raziskave ogroženih narečij v slovenskem jezikovnem prostoru (Radgonski kot, Gradiščanska, Hum na Sutli z okolico, Dubravica z okolico)* potekala na terenu v petih vaseh, tj. v Žetincih – Sicheldorf, Dedoncih – Dedenitz, Gorici – Goritz bei Radkersburg, Zenkovcih – Zelting in Potrni – Laafeld, kjer smo na novo zbrali obsežen korpus gradiva in na njegovi podlagi dosedanje védenje nadgradili z novim.

Izhodiščno raziskovalno vprašanje je bilo klasifikacija slovenskih govorov Radgonskega kota znotraj panonske narečne skupine, in sicer, ali govori pretežno izkazujejo prekmurske ali slovenskogoriške definicijske lastnosti, saj je to območje tudi ozemeljsko stično z območji prekmurskega in slovenskogoriškega narečja, ali morda zaradi različnih zgodovinskih dejavnikov izkazujejo značilnosti še katerih drugih narečij panonske narečne skupine. Raziskava je potekala skladno z že ustaljeno uspešno metodologijo raziskav ogroženih narečij v slovenskem jezikovnem prostoru. Definicijske lastnosti obravnavanih govorov bodo kartirane na jezikovnih zemljevidih, izdelanih s pomočjo aplikacije DIAtlas.

Alfred Lameli
Research Center Deutscher Sprachatlas, University of Marburg, Germany
lameli@uni-marburg.de

DIALECT CLASSIFICATIONS IN TRANSITION – LESSONS FROM THE PAST, TOOLS FOR THE FUTURE

(Plenary paper)

Dialect classification has long been central to linguistic inquiry, evolving from early cartographic representations to sophisticated computational models. Over time, methodologies have shifted from manually constructed dialect maps to statistical and algorithmic approaches that leverage big data, GIS mapping, and machine learning. This paper explores the dynamic landscape of dialect classification, considering historical perspectives, methodological challenges, and the role of emerging tools in reshaping the field.

One of the central questions in dialectology is how dialects are best classified: does it makes a difference whether classifications prioritize phonological features, lexical differences or syntactic variation? Should we assume continuity across linguistic landscapes, or do political and geographical borders hold intrinsic significance in shaping dialectal boundaries? These questions are far from new. Yet, the advent of computational methods introduces new dimensions to this discussion, allowing for broader and more nuanced analyses.

A key methodological challenge remains the weighting of linguistic variants. Traditional dialectology often relies on predefined criteria to determine which linguistic features hold the most significance in classification. However, contemporary computational models, such as those using neural networks and clustering techniques, may assign different weights dynamically based on data-driven insights. This raises critical questions about whether classifications should adhere to long-established taxonomies or evolve based on emergent linguistic patterns. Furthermore, the question of data quality persists: can noisy, real-world linguistic data provide reliable classifications, or does dialectology require meticulously standardized data sets? This raises broader concerns: are all linguistic system levels equally important for classification? Should cognates be weighted differently from non-cognate variants? If dialectal classification is to reflect linguistic reality, it must account for both systematic differences and irregular linguistic phenomena. Moreover, the influence of perception and intersubjectivity in dialectology cannot be overlooked. As will be demonstrated on the example of German dialects, incorporating intersubjective perspectives can lead to new approaches to classification that challenge established typologies.

By examining both historical and contemporary methodologies, this talk will outline a forward-thinking approach to dialect classification. By leveraging the lessons of the past

while embracing the possibilities of the future, we can develop more flexible, data-driven, and theoretically sound classification models. This discussion will set the stage for a broader conversation on the future of dialectology, ensuring that new tools and frameworks align with both linguistic realities and scholarly rigor.

5 Slovenski dialektološki posvet

Karin Marc
Univerza v Ljubljani, Filozofska fakulteta, Slovenija
karin.marc@ff.uni-lj.si

EVFEMIZMI V SLIKOVITEM IZRAŽANJU ZAHODNIH SLOVENSKIH NAREČIJ

Namen prispevka o slikovitem izražanju v zahodnih slovenskih narečjih je prek frazeologije širše predstaviti značilnosti družbene kulture, kot jo razgrinjajo narečni evfemizmi in jezikovni tabuji. Doslej zbrano gradivo v narečjih primorske narečne skupine in skrajnega zahodnega slovenskega etničnega ozemlja (slovenskih narečij v Italiji) so pokazali, da se najbolj posredno govori o prepovedanih temah, ki zadevajo nosečnost, duševne bolezni oz. umsko omejenost, alkoholizem, revščino, pa tudi smrt.

Narečno frazeološko gradivo se zbira s pomočjo usmerjevalne vprašalnice, pripravljene na podlagi narečne in domoznanske literature, ter se z vsakokratnim terenskim obiskom še dopolnjuje, saj se vedno dobi kak nov frazem; narečni govorci se namreč po asociaciji spomnijo še na druge tovrstne jezikovne enote. Pri vsakem frazemu se zbirajo podatki, kot so poznavanje frazema in njegovega pomena, raba (tudi s ponazoritvijo v ponazarjalnem zgledu) ter konotativnost (negativna ali pozitivna) in morebitne pragmatične značilnosti.

V prispevku bo prikazan bogat besedni in miselni zaklad predvsem s pojmovnega področja smrti, in sicer tako na sestavinski ravni kot na ravni pomena. Narečni frazemi, ki se povezujejo s smrtjo, bodo obravnavani po kronološkem principu: 1. *Frazemi, ki se nanašajo na obdobje pred smrtjo* (obdobje med življenjem in smrtjo ali bolezensko stanje kot bližina smrti), 2. *Frazemi, ki se nanašajo na trenutek ob nastopu smrti* (konec življenja ali trenutek, ko napoči smrt), in 3. *Frazemi, ki so povezani z obdobjem po smrti* (stanje po smrti, ki pa se že povezuje tudi z drugim pojmovnim področjem, tj. verovanjem). Zbrano gradivo bo analizirano z vidika lingvokulturoloških pristopov, zlasti v primeru zahodnoslovenskih narečnih frazemov pa deloma tudi kontrastivno.

Herta Maurer-Lausegger
Univerza v Celovcu, Avstrija
herta.maurer-lausegger@aau.at

ZILJSKO NAREČJE IN FILMSKE DOKUMENTACIJE SLOVENSKEGA IN NEMŠKEGA GOVORA VASI BLAČE V ZILJSKI DOLINI

Spodnja Ziljska dolina se nahaja na zgodovinsko pomembnem vozlišču starih prometnih poti, ki od nekdaj omogočajo tesne kulturne in jezikovne stike. Govorno območje ziljskega narečja se nahaja na skrajnem severozahodu vseslovenskega jezikovnega prostora, kjer se stikajo slovenščina, nemščina, italijanščina in furlanščina. Razteza se od Potoč (Potschach) pri Šmohorju (Hermagor) prek Ziljske doline do Podkloštra (Arnoldstein), kjer se navezuje na prehodne govore med ziljščino in rožanščino, ter sega do loške fare (Latschach) in Marije na Zilji (Maria Gail) z okolico. Vključuje tudi govore slovenskih vasi v sosednji Kanalski dolini, ki je bila skupaj z Rabeljsko dolino do podpisa senžermenske mirovne pogodbe leta 1919 del habsburške kronske dežele Koroške. V srednjem veku je ziljsko narečno območje segalo v kraje od Rateč do potoka Belce v Zgornjesavski dolini. Rateče, ki so pripadale kronski deželi Kranjski, so bile do razpada Avstro-Ogrske samostojna občina v sodnem okrožju Radovljica, cerkvenopravno pa so bile do leta 1390 del prafare Marija na Zilji, ki ji je pripadalo kakih 25 podružnic iz okolice. Tamkajšnji govor je dialektološko opredeljen kot ziljsko-gorenjski. Rezijanščina, ki je bila v 6. in 7. stoletju poseljena s koroške strani, temelji na ziljski narečni osnovi, vendar je od 14. stoletja dalje postopoma razvila lastni narečni sistem.

V Spodnji Ziljski dolini so bili slovenski govori stoletja glavno sporazumevalno sredstvo domačega prebivalstva. Proti koncu 19. in na prehodu v 20. stoletje so družbenopolitične razmere močno povečale pritisk germanizacije, zlasti na severnih obrobjih slovensko-nemške jezikovne meje. Zavedni Slovenci so se germanizaciji upirali in skušali ohraniti lastno identiteto. V zadnjem desetletju 19. stoletja je na Zilji zaživelo slovensko društveno in zadružno življenje, ki pa je postalo žrtev nemškega nacionalizma. Jezikovna podoba Ziljske doline se je v 20. stoletju močno spremenila. Družbeno-strukturne spremembe, zgodovinski dogodki, prepoved slovenščine in prisilna izselitev slovenskih družin ter pregon slovenske inteligence v vojnem času, migracije, turizem, odseljevanje slovensko govorečih v večja središča, jezikovno mešane družine, nemška vzgoja otrok, upadanje jezikovnega znanja in drugi dejavniki so odgovorni, da je večina ziljskega prebivalstva v zadnjih desetletjih zamenjala jezik. V največ krajih razmere ne dopuščajo več sporazumevanja v slovenskem narečju. Nekdaj prevladujoči slovenski govori so postali hišni jezik najstarejših. Z njihovim odhodom so ziljski govori v mnogih krajih je že utihnili. Ziljščina sodi k najbolj ogroženim slovenskim narečjem.

Stoletja dolgo sobivanje slovenskega in nemškega jezika in stik z romanskim jezikovnim prostorom sta pustila sledove vzajemnosti na jezikovnem in kulturnem področju. Medsebojni vplivi se odražajo na vseh jezikovnih ravninah. Značilne so arhaične prvine. V referatu bodo predstavljeni kulturnozgodovinski razvoj Spodnje Ziljske doline, opredelitev ziljščine na podnarečja in filmske dokumentacije slovenskega in nemškega govora vasi Blače (Vorderberg) v občini Štefan na Zilji. Na primerih filmov *Štehvanje p nas v Bvačah* in *Kufenstechen bei uns in Vorderberg* (2023) ter *Kaleda p nas v Bvačah: krale pa base* in *Kaleda bei uns in Vorderberg: Dreikönigsbrauchtum Krale und Base* (2024) bodo prikazane nemške prvine v slovenskih filmih in slovenske prvine v nemških filmih (koroška nemščina).

Danguolė Mikulėnienė
Institute of the Lithuanian Language, Vilnius, Lithuania
danguole.mikuleniene@lki.lt

THE PAST IN THE FUTURE: LITHUANIA'S TRADITIONAL DIALECTS AND NEW DIALECTAL FORMATIONS

(Plenary paper)

Framework: *isoglottic dialectology*. With the rise of structuralism, a classification of Lithuanian dialects and subdialects based on the phonological features of vocalism was developed in the second half of the 20th century (Girdenis, Zinkevičius 1966). Known today as the traditional classification, it did not consider any other features (lexical, morphological or syntactical) but followed the 19th century Neogrammarian tradition, forming the basis of the most important works in Lithuanian dialectology (Zinkevičius 1966; Morkūnas 1977–1991) and the research of Girdenis' phonological school (Girdenis 1981, 2014). The school of Girdenis significantly expanded the boundaries of descriptivism, applying precise machine calculation methods to experimental dialectal phonetics. But traditional Neogrammarian principles (applying the non-mobile older rural male criteria (NORM)) to native dialect speakers) were used alongside.

Framework: *isoglottic*, *qualitative* and *quantitative* dialectology. In the early 21st century, Lithuanian dialectologists recorded a shift in traditional dialects towards larger formations: *geolect – regiolect – regional standard* (Auer, Hinskens 1996). These variants are defined mostly territorially as their users often lose the most typical dialectal features due to the impact of the standard language (Mikulėnienė, Meiliūnaitė 2014). Regional dialects (regiolects) are forming around the larger, more attractive centres based on traditional Lithuanian dialects and often include more than one subdialect area. The geolectal zones observed within these territories also exhibit some of the best-preserved features of traditional (sub)dialects. The new formations are characterised by varying degrees of sustainability, from strong to transitional.

Framework: *multimodal dialectology*, (*socio-*)*geolinguistics*. The latest development in Lithuanian dialectology is the application of the multimodal method in researching local linguistic variants. As well as focusing on specific dialectal characteristics, this approach is based on analysing the linguistic landscape and sociocultural networks, the residents' views on their own local dialect and its usage (Mikulėnienė 2020a, 2020b; Čepaitienė, Mikulėnienė 2021; Mikulėnienė, Čepaitienė 2023). The multimodal method can therefore help forecast the direction local variants are likely to change towards.

However, the research conducted in 2019–2023 within the standard language zone in the district of Kaunas (around Lithuania's former temporary capital) showed that the diverse nature of the Lithuanian language goes beyond the concepts of geolect, regiolect and regional standard (Mikulėnienė etc. 2023). Qualitative studies identified several variants of the Lithuanian language in the area: (1) *Standard language zone*; (2) *Dialectal language zone*;

(3) *Non-dialectal language zone*; (4) *Standard-dialectal language zone*; (5) *Standard and non-dialectal language zone*; (6) *Dialectal and non-dialectal language zone*. All of these variants will have to be included in any future classifications. In other words, it is the variant's locality rather than its dialecticism that will be the focus, opening opportunities for using modern socio- and geolinguistic methods.

Nonetheless, it must be noted that the old dialectological tradition is retained when defining these local variants as new dialectal formations. They are analysed primarily based on the qualitative and quantitative realization of the same phonetical (phonological) features typical of the traditional dialects, with other linguistic characteristics treated as supplementary markers only.

Gabriela Múcsková
Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences in Bratislava, Slovakia
gabriela.mucskova@juls.savba.sk

THE DIALECT CONTINUUM AND THE DIALECT CLASSIFICATION

The paper builds on the article on the classification of Slovak dialects prepared within the DIACLEU project (Múcsková 2024) and will focus on the problematic issues of that classification – i.e. on the identification of dialectal macro areas, basic areas, or transitional micro areas – in a situation where the dialectal diversity of the Slovak linguistic territory has the character of a dialectal continuum. Within the dialectal continuum, the boundaries between dialects are not clear-cut and divergent linguistic changes overlap. Nevertheless, all the classifications of Slovak dialects presented in this article (Šembera 1864; Vážný 1934; Štolc 1968; Krajčovič 1988) delimit and name individual dialectal areas and seemingly transform the continuum into a mosaic, thus creating a certain structure of linguistic diversity.

In the case of the naming and delimitation of Slovak dialect units, apart from linguistic phenomena, socio-political circumstances played a significant role, especially the historical division of the territory of today's Slovakia into counties in the Old Hungarian Kingdom. Is it therefore possible to speak of linguistic criteria and an isoglottic type of classification (i.e. a classification based on purely linguistic criteria and dialectal diversity)?

A similar question, that is, the proportion of linguistic differences and socio-political factors, arises in the case of the classification of languages and their boundaries. For example, within the group of Slavic languages (on their classification see Šekli 2023; Blažek 2020 and others), which are also perceived from a dialectal perspective as a Slavic dialectal continuum, extra-linguistic – political and cultural – factors played a significant role. Moreover, in the structure of the linguistic mosaic, languages and dialects come into relations within the hierarchy in which they are categorized.

The presentation will therefore further compare the common and different phenomena in modelling different types of classifications with respect to the contribution of extra-linguistic factors in defining linguistic and dialectal boundaries.

Floris Nijhuis
University of Groningen, the Netherlands
f.b.nijhuis@student.rug.nl

Entropy-Based Linguistic Distances Conditional Entropy and Rajski's Distance Applied to Pronunciation Data

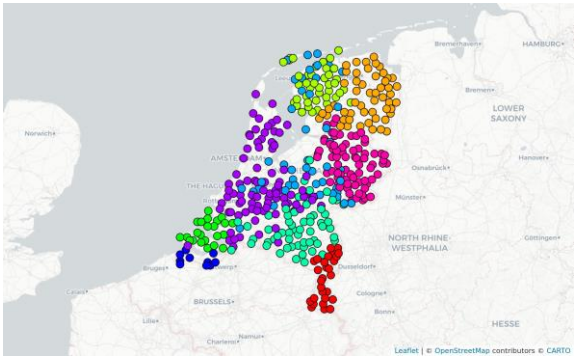
By design, common metrics of linguistic distances (e.g., the Levenshtein distance) are symmetric. However, this symmetry does not always reflect the linguistic reality. For example, Swedish and Danish show an asymmetrical intelligibility, with Danes understanding Swedish more easily than vice versa. Moberg et al. (2006) proposed a metric based on information theory to capture this asymmetry: conditional entropy (defined as $H(x|y)$, with x being the uncertainty given knowledge of y) and applied it successfully to demonstrate the Danish-Swedish intelligibility asymmetry using aligned phonetic word lists. The asymmetrical character of conditional entropy lies in the fact that it is based on a logarithmic transformation of a conditional probability, which is asymmetrical (i.e., $p(x|y) \neq p(y|x)$. For pronunciation data, this probability can, for example, refer to the chance of finding an /ɔː/ in language x given an /aː/ in language y (i.e., p(ɔː, aː)). Although not trivial, this concept can be extended to any aspect of linguistics, which can be defined in terms of a (discrete) probability distribution.

Furthermore, conditional entropy can be symmetrised straightforwardly using the following property: the sum of the entropy $H(x|y)$ and $H(y|x)$ constitutes the numerator of a common normalised distance measure in information theory. If divided by the joint entropy $H(x, y)$, this yields Rajski's distance (Rajski 1961), which has been applied to clustering (e.g, Kraskov et al. 2003). Although these metrics often have outcomes similar to other metrics, it should be noted that metrics derived from information theory work fundamentally differently as they measure the amount of uncertainty and not absolute distance.

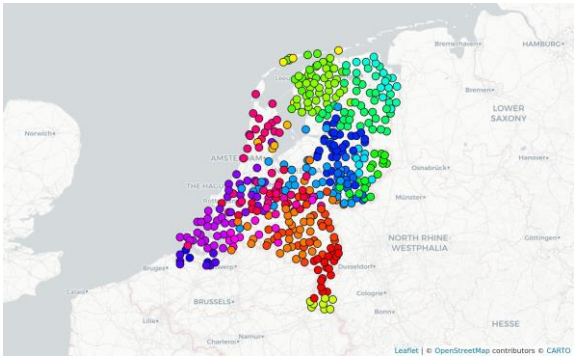
For this contribution, conditional entropy was extended to handle data with multiple realisations, and a normalised variant was introduced to compare unequal language (or, in this case, dialect) pairs. Conditional entropy and Rajski's distance were applied to the subset of the GTRP (Goeman-Taeldeman-Van Reenen-project) – a dataset of Dutch dialectal realisations containing 562 realisations for 422 places (Wieling et al., 2007). For every place pair, a joint probability distribution of all phonetic realisations was generated based on the alignments. The conditional entropy and Rajski's distance were calculated using these distributions for every place pair. Subsequently, the outcomes were visualised. For conditional entropy, the size of the asymmetry was visualised with a map. Given the

symmetrical nature of Rajski's distance, more common methods, such as cluster analysis and multidimensional scaling, were employed.

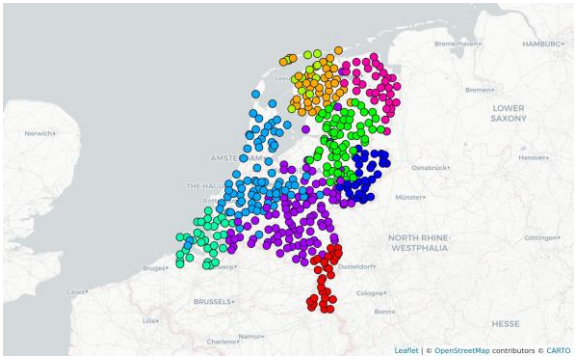
As visible in Figure 1, Rajski's distance yielded geographically more consistent clusters in the GTRP dataset compared to Levenshtein distance. To validate these results, two complementary approaches were employed: 1) validating clusters with Cdistance (Bartelds & Wieling 2022; Coen et al. 2010), and 2) using bootstrap methods to estimate the stability of the derived metrics. Furthermore, a broader comparison between these metrics and traditional linguistic distances was drawn.



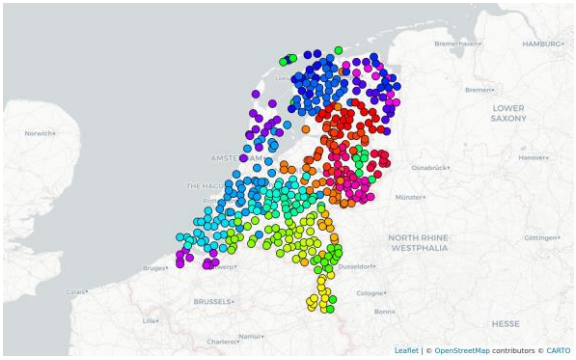
(a) Cluster map using Levenshtein distances (k = 9)



(c) Cluster map using Levenshtein distances (k = 5)



(b) Cluster map using Rajski's Distance (k = 9)



(d) Cluster map using Rajski's Distance (k = 25)

Figure 1: Comparison of cluster maps generated with Ward's method of Levenshtein (variance explained: 49.5%) and Rajski's Distance (variance explained: 55.8%)

Floris Nijhuis
University of Groningen, the Netherlands
f.b.nijhuis@student.rug.nl

John Huisman
Uppsala University, Sweden
john.huisman@lingfil.uu.se

Roeland van Hout
Center for Language Studies, Radboud University, Nijmegen, the Netherlands
roeland.vanhout@ru.nl

NEW WAYS TO ANALYSE THE RELATIONSHIPS OF DIALECT CLASSIFICATIONS IN A LARGER LANGUAGE AREA: THE CLASSICAL CASE OF THE DUTCH ZUIDERZEE REVISITED

Current dialectometric methods (cluster analysis, multidimensional, mixed-effects models) have restrictions on unravelling contact relationships between dialects and the overall role of geographic distance. In this presentation, we want to deal with a larger area of mutual relationships between dialects and dialect groups and the role of geographic distance between dialects. We use Admixture, a Bayesian population genetic approach, which assigns to each 'individual' (in our case, a dialect) proportions for a fixed number of ancestral populations (in our case, dialect clusters or groups). Admixture analysis provides a suitable method for modelling the horizontal transfer of genetic (or, here, linguistic) material. Crucially, it allows dialects to have multiple dialect ancestors or origins, i.e. they can be admixtures of multiple populations. We used data from the Zuiderzee (Southern Sea) area, a former inlet sea in the heart of the Netherlands. The Zuiderzee area is an intersection of three major regional languages: Hollandish-Brabantish, Frisian, and Low Saxon. A sea may act as a barrier to linguistic contact, but earlier work indicated instances of contact over water in the area under investigation. Classical dialectologists like Kloeke and Van Ginneken already emphasised the unifying forces in this area. Specifically, we looked at the realisations of the Proto-West Germanic vowels *$\bar{a}$ and *a for 40 words and 121 localities. The developments in these two vowels in this region were investigated in two words by Kloeke (1934): water and sheep. The Admixture analysis revealed nine linguistically ancestral populations, showing linguistic contact and migration over water and substrate and superstrate effects. We used these results to investigate the relationship between geographic and linguistic distance with a mixed-effects model. We used a matrix approach to define the geographic distance between each pair of localities/dialects. This matrix approach is new (see Huisman and van

Hout (2023)). Our findings show a typical relationship: where geographic distance increases, linguistic distance also increases, indicating a relation between physical separation and linguistic divergence. However, the strength of this correlation varies depending on the ancestral population and its scattering and contiguity. We conclude that the distances over water in the Zuiderzee were large enough to maintain the existence of three regional languages despite the combined influx of the relatively prominent Hollandish and the Dutch standard language. We expect this to be common in areas with a history of intensive and sustained shipping traffic.

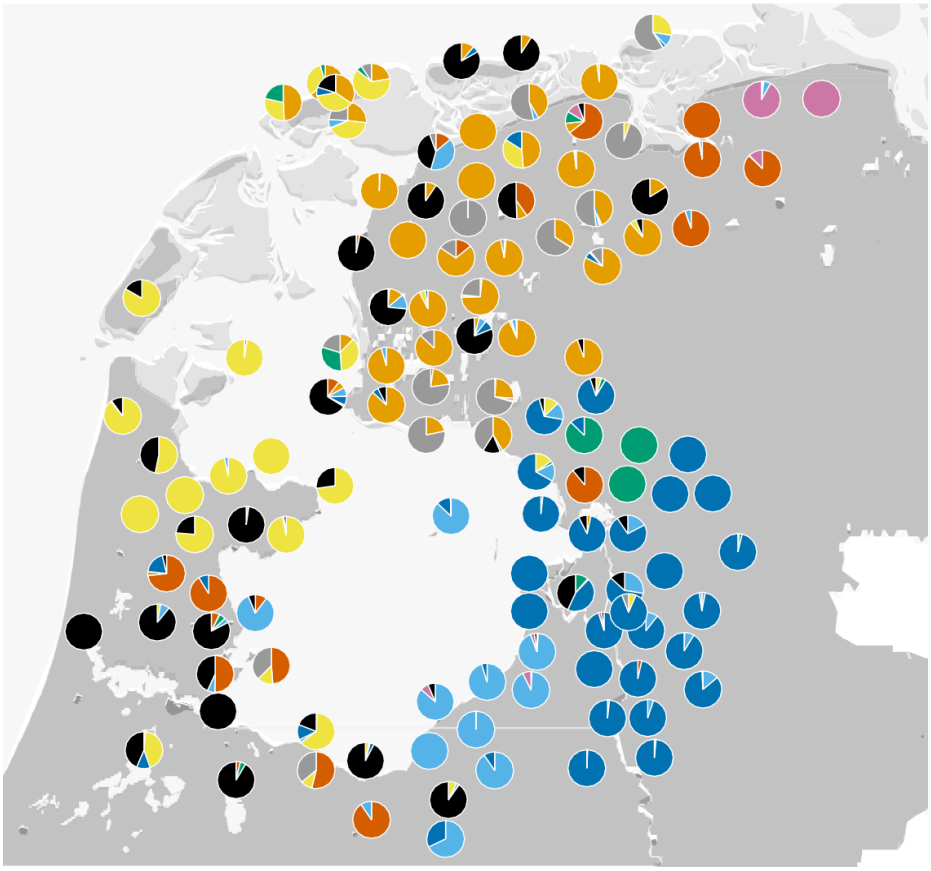


Figure 1: Results of Admixture Analysis with nine clusters (k = 9)

Stella Ondrejčiková
Jazykovedný ústav Ľudovíta Štúra SAV (Inštitut za jezikoslovje Ľudovíta Štúra, Slovaška akademija znanosti), Slovaška
stella.ondrejcikova@juls.savba.sk

KLASIFIKÁCIA A MAPOVANIE SYNTAKTICKÝCH JAVOV V SLOVENSKÝCH NÁREČIACH SO ZAMERANÍM NA PREDLOŽKOVÉ KONŠTRUKCIE

V slovenskej dialektológii sú zmapované hláskoslovné, tvaroslovné, slovotvorné a lexikálne javy v štvorzväzkovom *Atlase slovenského jazyka* (1968 – 1984). V minulosti bol pripravený dotazník pre spracovanie atlasu syntaktických javov slovenských nárečí, k zostaveniu tohto zväzku však nedošlo. Dotazník upozorňuje na špecifiká nárečových syntaktických javov a v Jazykovednom ústave Ľ. Štúra SAV v Bratislave existuje tiež samostatná kartotéka syntaktických javov. Syntaktickej rovine sa v kontexte nárečových výskumov venuje okrajová pozornosť, početné zastúpenie gramatických a syntaktických prác však nachádzame v tvorbe A. Ferenčíkovej (súhrnne 2019) a v menšom rozsahu sa vybraným javom venovali aj iní dialektológovia (napr. A. Habovštiak, E. Jóna, J. Oravec, J. Štolc).

Predmetom nášho výskumu sú primárne predložky, pričom sa zameriavame predovšetkým na ich základný priestorový význam a od neho odvodený časový význam, ako aj ich čiastkové významy. Doklady syntaktických konštrukcií s primárnymi predložkami excerpujeme z historických a nárečových textov, archívov a dostupnej literatúry. Analýzou nášho materiálu sa nám podarilo identifikovať osobitosti v syntaktických konštrukciách s primárnymi predložkami, ako aj ich osobité významy, ktoré sa často viažu len na konkrétnu nárečovú oblasť a vytvárajú tak areály. V príspevku sa budeme zaoberať možnosťami mapovania areálového rozšírenia niektorých predložkových konštrukcií z hľadiska ich formy a významu.

Tijmen Pronk
Leiden University, LUCL, the Netherlands
t.c.pronk@hum.leidenuniv.nl

DETECTING DIALECT CONTINUA IN THE PAST

One of the arguments used in the discussions of synchronic and diachronic language and dialect classification is the genealogical argument. Languages or dialects that have carried out shared innovations should be classified as belonging more closely together than those that have not taken part in the same innovations. That this criterium can be difficult to use in a dialect continuum has been observed over a hundred years ago. Johannes Schmidt (1872) famously argued that the way innovations spread across dialect continua does not necessarily lead to clear linguistic boundaries. In such cases, it is more adequate to describe the linguistic landscape in terms of centres from which innovations are able to spread, like waves formed by stones thrown into a pond. This is Schmidt's Wellentheorie or wave model.

The wave model is an adequate model for describing situations in which we can observe the existence of a dialect continuum in the present or in the recorded past. In assessing the relationship between related languages in prehistory, however, there is no clear methodology for establishing whether they formed part of a dialect continuum or not. Within the Indo-European language family, e.g., Proto-Indo-European as well as various proto-languages of the individual branches of the family have often been hypothesized to have formed dialect continua, but there is no consensus on the type of arguments that can be used to support or refute such hypotheses.

In this paper, we will try to answer the question whether there are any linguistic characteristics of known dialect continua that can be used to identify dialect continua in the past. It will be argued that one of the hallmarks of a dialect continuum is variation that has arisen as a result of sequences of innovations that affect different parts of the continuum in a different chronological order. This appears to be especially characteristic for the early phase of a dialect continuum, when a high degree of mutual intelligibility makes the spread of innovations across larger areas relatively easy. When identical linguistic innovations affect two different areas in a different order, this leads to different outcomes of what are essentially the same innovation. This will be illustrated by examples from Slavic and Baltic languages, specifically Slovenian, Croatian, Russian, Latvian and Lithuanian based on previous research into the history of these languages (Vermeer 1983, 1986; Kalniņš 2020; Pronk 2021).

Ana Savić Grujić
Institut za srpski jezik SANU (Inštitut za srbski jezik SANU), Srbija
sandrasavic0302@gmail.com)

Sandra Savić
Institut za srpski jezik SANU (Inštitut za srbski jezik SANU), Srbija
sandrasavic0302@gmail.com

LINGVOGEOGRAFSKI POTENCIJAL LEKSEME *MRTVAC* U GOVORIMA PRIZRENSKO-TIMOČKE DIJALEKATSKE OBLASTI

Govori prizrensko-timočke dijalekatske oblasti zauzimaju periferno područje srpskog jezičkog prostora, protežući se na teritoriji istočno od linije Prizren – Zaječar. Na jugu su omeđeni makedonskim govorima, sa istoka bugarskim, dok su na jugozapadnoj strani isprekidani albanskim jezikom. Kao što je u nauci poznato, turska okupacija u XIV veku uslovila je njihovo izdvajanje iz štokavske mase, zbog čega se govori prizrensko-timočke dijalekatske oblasti odlikuju arhaizmima i brojnim inovacijama balkanističke prirode, koje ih u pojedinim segmentima potpuno ili delimično diferenciraju u odnosu na celinu srpskog jezika. Pored toga, poznate su i osobine koje pokazuju unutrašnju diferencijaciju na osnovu kojih se izdvajaju tri dijalekta u okviru prizrensko-timočke dijalekatske oblasti: timočko-lužnički, svrljiško-zaplanjski i prizrensko-južnomoravski dijalekat. Specifičan razvoj uslovio je da se u likovima pojedinih leksema često očitava i više od jedne osobenosti, nekada svojstvene celom prizrensko-timočkom području, a nekada razlikovne unutar samog područja, što čini verodostojan ilustrativan jezički materijal za lingvogeografsko predstavljanje.

Dijalekatski materijal za leksemu *mrtvac* (*mrtvьc*, *mrtvac*, *mrtvьk*, *mrtvak*, *mrtьvьc*, *mrtavьc*, *mrtavac*, *mrtevac*, *mrtovac*, *mrtoac*, *mrtovec*, *mьrtvьk*, *mьrtvak*, *mьrtьvьc*, *mьrtavьc*) prikupljen za potrebe *Srpskog dijalektološkog atlasa* u 137 punktova širom prizrensko-timočke dijalekatske oblasti, te upoređen s leksičkom građom u mnogobrojnim rečnicima s istog govornog područja, primarno pruža uvid u razvoj praslovenske grupe **ьr* na prizrensko-timočkom arealu: *mrt-* : *mьrt-*, paralelno svedočeći i o egzistenciji različitih sufiksa: *-ьk/-ak* : *-ьc/-ac* i infiksa: *-v-/-av-/-ev-/-ov-*. Iako je akcenat na celoj teritoriji ekpiratoran, bez opozicije po kvalitetu i kvantitetu, zabeležene akcenatske varijante (s akcentom u korenu reči, na unutrašnjem slogu ili na sufiksu) govore o njegovom čuvanju na starom mestu ili o povlačenju u nekim zonama). Dakle, predočene jezičke datosti predstavljaju pouzdan materijal za izradu čak tri lingvističke karte.

U materijalu za *Opšteslovenski lingvistički atlas*, Tom 3. dati su kontinuanti praslovenskih grupa **ьr*, **ъr*, **ьl*, **ъl* na slovenskoj jezičkoj teritoriji. Za naš aspekt istraživanja posebnu važnost ima terenska građa iz prizrensko-timočkih punktova Orahovac, Pljačkovica, Donji Dušnik, Kalna, čija je interpretacija predstavljena na karti 14 F 2331 **mьrtvъ(-jь)*.

Mojca Schlamberger Brezar
Univerza v Ljubljani, Filozofska fakulteta, Slovenija
mojca.schlambergerbrezar@ff.uni-lj.si

Urška Stanković Elesini
Univerza v Ljubljani, Naravoslovnotehniška fakulteta, Slovenija
urska.stankovic@ntf.uni-lj.si

Primož Weingerl
Univerza v Ljubljani, Naravoslovnotehniška fakulteta, Slovenija
primoz.weingerl@ntf.uni-lj.si

Valentina Teršek
Univerza v Ljubljani, Naravoslovnotehniška fakulteta, Slovenija
vt33913@student.uni-lj.si

LUCIEN TESNIÈRE IN NJEGOV ATLAS ZA ŠTUDIJ DVOJINE: AKTUALIZACIJA PO 100 LETIH IZIDA

Lucien Tesnière, prvi lektor francoskega jezika na novoustanovljeni Univerzi v Ljubljani med letoma 1920 in 1924, hkrati pa Meilletov doktorski študent, je leta 1925 na Univerzi v Strasbourgu doktoriral s tezo *Les formes du duel en slovène* in sekundarno tezo *L'Atlas pour servir à l'étude du duel en slovène*. Prva od tez do sedaj ni bila prevedena, Atlas pa je leta 2022 v prevodu izšel pri Znanstveni založbi Filozofske fakultete Univerze v Ljubljani. Tesnière, za časa svojega življenja premalo cenjen in znan v slovenskem prostoru, katerega jeziku je posvetil temeljno delo, je kot pionir geolingvistike, saj gre za prvi opis kakega slovanskega jezika v tem okviru, danes pri nas praktično pozabljen. Prevod atlasa, ki ga v originalu hranijo v OHK FF UL v le še dveh izvodih, je povečal dostopnost Tesnièrjevega dela tako med strokovno kot širšo javnostjo.

V povzetku predstavljamo interdisciplinarno sodelovanje med NTF in FF, da bi atlas na eni strani predstavili širši javnosti, na drugi pa poenostavili njegovo rabo. V ta namen je bil izdelan prototip omenjenega atlasa v digitalni, interaktivni obliki. Z digitalizacijo atlasa smo dodatno prispevali k ohranjanju dragocene jezikovne dediščine francoskega lingvista, hkrati pa omogočili njegovo širšo rabo.

Izdelava prototipa je zajemala naslednje korake:

1. zasnova: zbiranje podatkov, določitev obsega in načina digitalizacije, analiza tiskane različice, pregled obstoječih rešitev, vektorizacija slikovnih predlog;

2. analiza uporabnikov: izvedba intervjujev, anketiranje uporabnikov, definiranje person;
3. načrtovanje uporabniške izkušnje: informacijska arhitektura, kreiranje različnih konceptualnih rešitev, izdelava mrežnih diagramov oz. skeletov, načrtovanje interaktivnosti in poteka uporabnika;
4. oblikovanje uporabniškega vmesnika: oblikovanje grafičnih elementov, izbira tipografije, oblikovanje ikonografije, izdelava in obdelava slikovnih predlog, oblikovanje posameznih komponent uporabniškega vmesnika, oblikovanje posameznih strani za različne velikosti predstavitvenih polj (zaslonov);
5. prototipiranje/izdelava interaktivnega prototipa v programu Figma.

V prihodnosti načrtujemo izdelavo končnega izdelka – atlasa v digitalni, interaktivni obliki, ki bo na spletu dostopen ciljni skupini in širše. Taka dostopnost bo omogočala lažje iskanje vzorcev v atlasu v povezavi z izreko, kar bo omogočalo širšo in učinkovitejšo rabo med zainteresirano publiko ter lažji dostop do Tesnièrjeve dediščine.

Robert Marcello Sevilla
The University of Hong Kong, China
rsevill2@hku.hk

ON THE (NON-)EXISTENCE OF XIANG: SUBGROUPING AND LEXICAL-MORPHOLOGICAL FEATURES

The Sinitic languages ('Chinese') represent around the order of 100 varieties with limited mutual intelligibility (Norman 2003), traditionally grouped into 7 to 10 subfamilies (Handel 2017), many of which have resisted attempts at subgrouping using the traditional criterion of distinctive shared innovations. Xiang (a traditional subgroup located in Central Hunan and Northern Guangxi spoken by around 30 million people) presents a particular conundrum: despite the received opinion that it represents a valid taxonomic unit (Yuan 1960, 1983, 2003; Wu 2005; Chen & Bao 2012; Bao 2017), the search for unique innovations defining the grouping (and its subgroups) has not had positive results. This has led some (Zhou & You 1988; Coblin 2011; Sevilla 2023) to claim that the grouping is in fact merely a taxonomic convenience and, for the purposes of phylogeny, should be divided into smaller subgroups which do present unique innovations.

The present study attempts to contribute to the search for distinctive shared features for Xiang by supplementing traditional phonological features with lexico-morphological properties. Of the 510 features in *The Atlas of Chinese Dialects* (Cao 2008), 428 (150 phonological, 195 lexical, 83 grammatical) were directly evaluated for six locations, four of them representing core varieties of Xiang, and then cross-checked with published materials, resulting in 1712 data points for analysis. Of the original 428 features, only 10 were both unique and general to Xiang, but upon further investigation it was found that all but one had wider distributions outside of Hunan. Xiang cannot therefore be shown to exist as a valid subgroup with the available evidence; while isolated instances of innovations can be found (老虫 for 'tiger', aspirated initial in 'kneel', *man-* kinship prefix, etc.), there is no core vocabulary that can identify the subgroup, nor are there any exclusive phonological innovations, supporting the claim that the grouping is geographical-cultural rather than linguistic.

However, it was found that clusterings of lexico-morphological isoglosses do allow us to identify several smaller groups – a northern group (including the dialects in and around Changsha), roughly corresponding to the traditional 'New Xiang' dialects, a group approximately corresponding to Coblin's (2011) Common Central Xiang (CCX), and a more southerly subgroup represented by the dialect of Shaoyang. Further investigation using Kalyan & François (2018)'s 'historical glottometry' model for measuring subgroup tenability turned up weak support for only one of the subgroups (Central Xiang), but reliable

descriptive materials which would further bolster this approach are lacking. The results of this study support three major conclusions: (1) Xiang cannot be recovered as a subgroup of Sinitic and is most likely paraphyletic; (2) there is some evidence that smaller groups do represent valid subgroups, although evidence is weak at best and more detailed linguistic descriptions of Hunanese Sinitic varieties are necessary; and (3) the tree model and the glottometric approach to subgrouping could be enriched by a looser, nucleus-periphery approach to lexico-morphological innovations and accounting for the rarity of particular innovations, respectively.

5 Slovenski dialektološki posvet

Vera Smole
Univerza v Ljubljani Filozofska fakulteta, Slovenija
vera.smole@guest.arnes.si

IZ DOLENJSKE DOBREPOLJSKE FRAZEOLOGIJE IN PAREMIOLOGIJE

V prispevku bodo obravnavani neznani ali manj znani frazemi in pregovori v govoru Vidma v Dobrepoljski dolini. Že več kot deset let jih zbira domačinka Anica Štrubelj, ki želi izdati tudi besedni narečni slovar diferencialnega tipa, poleg tega pa še vrsto spominskih zgodb, starih molitvic in drugega gradiva. Da to ne bo le eden od ljubiteljskih slovarjev, ji pri delu že drugo leto pomagamo s študenti izbirnih predmetov na ljubljanski slovenistiki. Iz nabora okrog 450 frazemov bodo predstavljeni predvsem živalski, od okrog 180 pregovorov pa aktualni tudi za današnjo rabo. Opozorjeno bo na določene glasovne pojave, ki jih v drugih bližnjih govorih doslej še nismo zasledili rabljene tako sistemsko.

Namen prispevka je prikazati 1) postopke vključevanja in povezovanja domačinskih raziskav v študijski proces; 2) pridobljene izkušnje ob tem; 3) nekaj rezultatov skupnega dela in s tem novega gradiva.

Xulio Sousa
University of Santiago de Compostela, Spain
xulio.sousa@usc.gal

Francisco Dubert Garcia
University of Santiago de Compostela, Spain
francisco.dubert@usc.es

OLD DATA AND NEW CLASSIFICATIONS: GALICIAN VARIETIES FROM THE ALPI

Until the publication of Fernández Rei's work in 1990, dialectal variation within the Galician linguistic domain remained a largely unexplored field. This foundational manual greatly benefited from the data collected for the creation of the *Atlas Lingüístico Galego* (ALGa). The ALGa surveys were conducted between 1974 and 1977 and are still being published. The ALGa provides a comprehensive depiction of Galician dialectal variation in the regions where the language is spoken (Galicia and the westernmost fringes of the provinces of Oviedo, León, and Zamora). This data enabled Fernández Rei to develop what is now considered the canonical classification of Galician dialects into three major dialectal blocks.

Subsequently, Hans Goebl's dialectometric method was applied to the ALGa data in various studies (Álvarez Blanco, Rosario, Francisco Dubert García, and Xulio Sousa 2006; Dubert García 2011, 2012; Sousa 2006, 2017, 2020) to generate quantitative analyses of Galician dialects. These qualitative and quantitative studies collectively allowed researchers to present Galician dialects to the linguistic community in a systematic manner, classify them, and identify dialectal patterns embedded in the data.

Nevertheless, both the qualitative and quantitative approaches have treated Galician as an isolated, self-contained language. In reality, Galician is part of the Romance linguistic continuum, geographically connected to Portuguese in the south and to Asturian-Leonese in the east. Galician possesses its own distinct linguistic features, as well as many characteristics shared with neighbouring varieties. The ETLEN project (Andrés et al. 2017), which surveyed varieties spoken along the Galician-Asturian linguistic border, sought to quantitatively establish the boundaries between these two languages.

Additionally, data from the *Atlas Lingüístico de la Península Ibérica* (ALPI; García Mouton et al., 2016) allows us to situate Galician varieties within the broader context of other Romance varieties with which Galician is in contact. We will apply the dialectometric method to the ALPI data from northwestern varieties to develop a classification of Galician dialects in their broader context: the northwestern constitutive dialectal varieties. The Romance varieties spoken in the northern Iberian Peninsula are referred to as *constitutive*

because it is in this region that the Ibero-Romance languages originated. Thus, northwestern Romance varieties can be quantitatively compared, enabling the analysis and classification of Galician, Asturian-Leonese, Castilian, and Portuguese within their shared geolectal context.

Irena Stramljič Breznik
Univerza v Mariboru, Filozofska fakulteta, Slovenija
irena.stramljic@um.si

NAREČNA RASTLINSKA IMENA MED ETIMOLOGIJO, BESEDOTVORJEM IN LJUDSKO ETIMOLOGIJO

V sodobnih splošnih slovarjih slovenskega jezika (SSKJ2, eSSKJ, Sinonimni slovar, Pravopis in Sprotni slovar) portala Fran je trenutno zajetih 195 rastlinskih imen z označevalnikom nar(ečno).

Za natančnejšo analizo smo izbrali šmarnico (*Convallaria majalis*) in marjetico (*Bellis perennis*), ki ob upoštevanju starejših virov (Bevk 1929; Barle 1937; Cilenšek 1892; Šulek 1879) kažeta zelo bogata variantna ali sinonimna ljudska poimenovanja.

Analize njunih različnih poimenovanj se bomo lotili tako, da bomo sledili razpoložljivim diahronim etimološkim razlagam in hkrati prepoznavali sinhrono motivacijo z besedotvornim pristopom, podprtim s širšim kulturološkim vidikom. Pri tem se zavedamo, da pri ljudskih rastlinskih imenih prihaja do različnih etimoloških variacij in ljudskoetimoloških križanj, kar zelo zamegljuje njihov dejanski pomen.

Namen prispevka je po eni strani izpostaviti bogastvo ljudskih rastlinskih imen, v kateri se skrivata tako jezikovna kot kulturna dediščina, po drugi strani pa kočljivost takih raziskav, ki se jih z Bezlajevimi besedami »bolj ali manj vsi bojimo ter se jih izogibamo kjer in dokler moremo.«

Omejitev dialektološke vrednosti raziskave predstavlja dejstvo, da večina poimenovanj, z izjemo nekaj primerov iz starejših virov (npr. Pleteršnik 1894–1895), nima zabeleženih podatkov o današnji krajevni rabi, kar je zelo bistven podatek za narečno raziskovanje.

Ho Wang Matthew Sung
Leiden University, the Netherlands
h.w.m.sung@hum.leidenuniv.nl

Jelena Prokić
Leiden University, the Netherlands
j.prokic@hum.leidenuniv.nl

ADVANCES IN USING MULTIPLE SEQUENCE ALIGNMENT IN DIALECT CLASSIFICATION

Quantitative methods in dialect classification have been in use since Séguy's pioneering work on French dialects in the 1970s (Séguy 1971; Séguy 1973). Over time, this area of research, known as *dialectometry*, has evolved and been further developed with a variety of statistical and computational techniques. Quantitative approach to dialect classification enables researchers to utilize all available dialect data without the need to manually select specific features, which can often be subject to selection biases. As a result, it is a much faster method that yields classifications comparable to those obtained through traditional techniques.

In this talk, we present the benefits of using automatic multiple sequence alignment in quantitative dialect classification. Multiple Sequence Alignment (MSA) is a technique used to align multiple strings simultaneously (e.g., IPA transcriptions). Commonly employed in bioinformatics for tasks like aligning DNA sequences to assess evolutionary distances between species, MSA has also demonstrated its effectiveness in dialectometry by providing more accurate alignments than pairwise alignments (Prokić at al. 2009). It is particularly useful for identifying and processing sound correspondences, as well as uncovering diachronic patterns that are often difficult to detect in pairwise alignments (Nerbonne et al. 2010).

In addition to offering a more transparent alignment for distance calculation compared to pairwise alignment (Levenshtein 1966; Heeringa 2004), multi-aligned transcriptions can be highly efficient for automatic feature extraction. Sung and Prokić (2024a) applied normalized Pointwise Mutual Information (nPMI) to multiple aligned sequences to identify distinctive features specific to individual dialect areas. They also showed that MSA can be combined with Factor Analysis in a bottom-up fashion to automatically extract dialect features.

Another benefit of the MSA approach is its ability to effectively integrate linguistic and spatial data, which is valuable for identifying focal and transitional dialect areas. Inspired by Séguy (1973), Sung & Prokić (2024b) rely on multiple aligned sequences to extract the

most distinctive dialect features, and, with the aid of local spatial autocorrelation, they identify both core and transitional dialect areas. Sung (2024, in progress) on the other hand used the so-called *dialect typicality analysis* to identify the presence and absence of exclusive features in transitional areas between two dialect groups, in a totally automatic fashion.

There are several other uses of MSA in dialectometry. For example, multiple aligned sequences form the foundation of the method presented in Prokić & Cysouw (2013), which can be used to trace the spread of dialect innovations and investigate the dynamics of dialect change.

Overall, MSA converts the original transcription into a clearer format, providing a linguistically transparent foundation for dialect classification and various types of research, such as detecting dialect features and analyzing the dynamics of dialect variation. When applied to existing dialect data, MSA not only enhances traditional methods of dialect classification, but also opens up new opportunities for quantitative research on dialects in the future.

Miha Sušnik
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
miha.susnik@zrc-sazu.si

SLOVENSKO NAREČNO GRADIVO V ZAPUŠČINI LUCIENA TESNIÈRA

Lucien Tesnière je za dialektološki del svoje disertacije pripravil mrežo 284 točk, v katero je vključil vse že objavljeno slovensko narečno gradivo, za 87 točk pa ga je v letih 1921–1922 na terenu zapisal sam po lastni vprašalnici s 424 vprašanji – te točke torej tvorijo zaključen in enoten podsistem mreže s precejšnjo količino zbranega gradiva, ki razen posameznih primerov v Tesnièrovi disertaciji in znanstvenih prispevkih ni bilo nikoli v celoti objavljeno.

V Atlasu je predstavil le 84 od vseh 424 vprašanj, zato doslej ni bilo mogoče niti predvidevati, kaj bi to gradivo lahko zaobjemalo. Ker se je vse jezikovno gradivo ohranilo v njegovi zapuščini (*Bibliothèque nationale de France*), se zastavlja vprašanje, kaj to gradivo vsebuje, kakšna je njegova vrednost za slovensko dialektologijo in za kakšne raziskave ga bo mogoče uporabiti v prihodnosti kljub dejstvu, da je bila vprašalnica zasnovana za študij dvojine.

V prispevku najprej predstavljam Tesnièrovo zapuščino s poudarkom na gradivu v zvezi z njegovo disertacijo, nato pa se osredotočam na podmrežo govorov, zapisanih po vprašalnici, pri čemer primerjam koncept in metode projekta SLA. Točke podmreže so sorazmeroma enakomerno geografsko razporejene, predvsem na vzhodu je večkrat zapisal govore bližnjih krajev, kjer je že mogoče opazovati razlike med govori, npr. Podgorje in Šentilj pod Turjakom, Družmirje in Velenje, Podčetrtek in Poljana Sutlanska, Brestanica in Krško ter prleške točke Vareja, Markovci in Dornava. Bistven je tudi čas zapisa: če je Tine Logar marca 1954 v Podčetrtku zapisal samo odraz *ȇ* za dolgi jat, kažejo Tesnièrovi primeri *svȇjča*, *dȇjla* (3. ed.), *lȇjp*, *dvȇ*/*dvȇj*, *ọbȇj*, *brẹ́jza*, *smrẹ́jka*, *mẹ́jsto*, *cẹ́jsta* še prevladovanje dvoglasnika *ȇi̯*. Tesnièrovo gradivo iz Dobove ima kratke odraze cirkumflektiranih dolžin, prim. *lẹ̀p*, *snjẹ̀k*, *blẹ̀t* ‘bled’, *brẹ̀k*, *mjẹ̀t* ‘med’, *vrȁt* (im. ed.), *mù̯st*, *bùk* ‘bog’, *vùk* ‘volk’ proti *xíšọ* (tož. ed.), *brẹ́za*, *smrẹ́ka*, *mjẹ́stu*, *cẹ́sta*, *mlẹ́ku*, *svẹ́ća*, *bẹ́w*, *grjẹ́x*, *ljẹ́po dijẹ́te*, *gnjẹ́zdọ*, *slẹ́pa*, *víẹm*, *sjẹ́dę̣*, *kráva*, *trávu*, *xrȃst*, *vrȃt* (rod. mn.), *túčẹ* (rod. ed.), *gústọ*, *súsẹt*, *vọ́zi* (3. ed.), *úsẹm*, *prúsn̨* (1. ed.), *gdû*, *ọ́trúk* (rod. mn.), kar je bilo doslej znano le iz Toporišičevega opisa govora Mosteca.

Gradivo je dragoceno tudi zaradi narave vprašalnice, saj se vprašanja večinoma pojavljajo v večbesednih zvezah ali stavkih (skupno število besed v knjižni različici je nad 1700), kar omogoča tudi opazovanje skladenjskih značilnosti krajevnih govorov, npr. obliko pridevnika v prilastkovnem in povedkovnem položaju, ujemanje ali npr. izražanje dopustnosti, prim. *Četudi je oreh zelen, je vendar že zreł*. Zaradi zasnove vprašalnice v stavkih se pojavlja

več pridevnikov, prislovov in členkov, ki navadno niso vključeni v dialektološke vprašalnice, prim. *kurji*, *pasji*, *mačji*, *kravji*, *božji*, *puškin*, *višnjev*, *kraljev*, *hudičev*, *krvav*, *divji*, *večeren*, *orožen*, *nadležen*, *ubran*, *teman*, *droben*, *zrel*, *polagoma*, *nenadoma*, *popolnoma*, *kmalu*, *doslej*, *torej*, *zelo*, *vsaj*, *morda*. Vprašalnica zajema več samostalnikov, ki jih v vprašalnici za SLA ni, prim. *čeljust*, *obraz*, *zrno*, *krma*, *milo*, *sukno*, *zapah*, *mezda*, *krčma*, *vrsta*, *sluga*, *junak*, *kletev*, *zarja*, *odmev*, *holm*, *reka*, *mavrica*, *smokva*, *brlog*, *vrana*, *sraka*, *sova*, *sokol*, *vrabec*, *ščinkavec*, izpeljanke *trgovec*, *kupec*, *delavec*, *sejalec*, *plesalec*, *starka*, *gobar*, *bolnik*, *bogatin*, *lenuh*, *oderuh*; nekateri od teh so prisotni v vprašalnici za OLA.

Татьяна Шалаева
Институт славяноведения РАН (Inštitut za slavistiko Ruske akademije znanosti), Rusija
koulkuk@gmail.com

БЕЛОРУССКО-РУССКИЕ ИЗОЛЕКСЫ *ШО́ПА*, *ШО́ФА* И *ШО́ХА*, *ШОХА́* ‘НАВЕС’

В докладе анализируется случай пересечения изоглосс двух близких фонетически и идентичных семантически, но неродственных лексем *шо́па*, *шо́фа* и *шо́ха*, *шохá* ‘навес’, зафиксированных в русских и белорусских диалектах. Наложение их ареалов будет наглядно продемонстрировано на специально составленных картах-схемах.

В русских говорах регулярно фиксируются лексемы *шох*, *шо́ха*, *шохá*, *ша́ха*, *шахá*, называющие навес на четырех столбах или сарай, куда складывают сено и снопы. Они концентрируются в южной части севернорусского наречия и в западной части средне- и южнорусских говоров.

Об их происхождении высказывались разные версии. Во-первых, они считаются заимствованиями из финно-угорских языков, в частности из карельского, с которым русские говоры находятся в контакте (ср. карел. *šaiho* ‘оградка или клеть для хранения мякины, соломы, листьев’). Во-вторых, высказывалась гипотеза об их исконно славянских истоках и производности от лексемы *сохá*, обозначающей различные жерди, колья и подпорки. Мена согласных *с*/*ш* при этом определяется как результат влияния финно-угорских языков (Мызников 2019: 914–915).

Однако *шо́ха*, *шохá* фиксируются также в южнорусском наречии, где связь с финно-угорскими языками отсутствует. Как представляется, этот факт полностью опровергает версию об их заимствованном происхождении и указывает на необходимость трактовать данные формы как исконные образования, а именно возводить к корню *сох*-/*сош*-.

Помимо ареалогических доказательств, эта этимология подтверждается и фонетически. Так, отмечены бесспорные дериваты корня *сох*-/*сош*- с начальным *ш*, зафиксированные за пределами севернорусского наречия: ср. *шошни́к* ‘резец, лемех сохи; сошник’, *шо́шка* ‘жердь с развилкой на конце, служащая подпорой к чему-либо’, ‘сооружение из жердей для просушки сена’ при *со́шка* ‘сооружение из кольев, палок и т.п. с раздвоенным концом и положенной на них горизонтальной жердью’, ‘сооружение из шестов, кольев, служащее основой для стога, шалаша и т.п.’. Вероятно, в указанных производных корня *сош*- произошла ассимиляция, приведшая к появлению начального звука *ш*, а затем у форм типа *шошни́к*, *шо́шка* образовалась ложная производящая основа *шох*- по аналогии с парой *сошка* – *соха*.

Определять исконную основу *сох*- в *шо́ха*, *шохá* заставляют не только указанные структурные явления, но и семантика, поскольку *сохá*, помимо палок и жердей,

называет еще и постройки и их части – такие же, какие обозначают *шóха*, *шохá*: ср. *сохá* ‘навес на столбах для хранения снопов соломы, сена и т.п.’. В свою очередь, *шóха*, *шохá* служат наименованиями не только навесов и сараев, но и палок и жердей: ср. *шóха* ‘жердь с сучками, вокруг которой мечут стог, если сено не просохло’, *шохá* ‘сухие деревья’. Таким образом, для возведения *шóха*, *шохá* к *сохá* имеются и структурные, и семантические, и лингвогеографические основания.

Кроме западных русских говоров, *шахá*, *шáха* отмечаются в восточных белорусских диалектах: ср. *шахá* ‘здание с плетеными стенами для сельскохозяйственного инвентаря’, ‘навес’, *шáха* ‘постройка на четырех столбах’. Однако повсеместно в белорусском языке и изредка в западных средне- и южнорусских говорах встречается фонетически и семантически близкая к *шóха* форма *шóфа*, *шóпа*: ср. белорус. диал. *шóпа* ‘навес’, *шóфа*, *шоф* ‘навес’, русск. диал. *шóпа* ‘сарай или большой навес, особенно на базаре’, ‘помещение для хранения льна’, *шóфа* ‘сарай для хранения льна’, *шафá* ‘сарай для сушки сена’. Западнорусские и белорусские формы *шóпа*, *шóфа* вместе с укр. *шóпа* ‘навес; сарай; хлев’ являются заимствованиями из польского, где *szopa* ‘временная хозяйственная постройка’ восходит к ср.-в.-нем. *schopf(e)* ‘строение без стен, крыша на столбах’ (Мызников 2019: 915; Фасмер 4: 466; ЕСУМ 6: 453; Brückner: 551; Boryś: 605).

Таким образом, восток Белоруссии и запад России оказываются территорией пересечения ареалов не связанных генетически, но близких по структуре и значению лексем *шóха*, *шохá*, *шахá*, *шáха*, с одной стороны, и *шóпа*, *шóфа* – с другой.

5 Slovenski dialektološki posvet

Matej Šekli
Univerza v Ljubljani, Filozofska fakulteta, Slovenija
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
matej.sekli@guest.arnes.si

O NASLONSKIH NIZIH PREGIBNIH NASLONK V REZIJANSKEM NAREČJU SLOVENŠČINE

V prispevku so obravnavani naslonski nizi pregibnih naslonk v rezijanskem narečju primorske narečne skupine slovenščine, in sicer z zgodovinskojezikoslovnega vidika, saj je zgodnji popraslovanski sistem pregibnih naslonk v tem slovenskem narečju doživel korenite spremembe. V zgodovini rezijanščine namreč ni prišlo samo do povečanja števila pregibnih naslonk, temveč sta se spremenila tudi njihov položaj v stavku in zaporedje v naslonskem nizu, slednji pa se je poleg tega razcepil v dva naslonska niza. Tovrstne spremembe so posledica jezikovnega vpliva sosednjih romanskih jezikov in njihovih narečij, v prvi vrsti furlanščine, kot tudi nekaterih razvojnih teženj v širšem slovenskem jezikovnem prostoru.

Število pregibnih naslonk v rezijanščini se je povečalo zaradi nastanka novih naslonskih oblik imenovalnika osebnih zaimkov (npr. *ún* vs. *an* ‘on’, *oná* vs. *na* ‘ona’, *uní* vs. *ni* ‘oni’, *onè* vs. *ni* ‘one’) ter novih nenaglašenih glagolskih oblik, tj. novih pomožnih glagolov (npr. *(an) ćé (iso rič)* ‘(on) hoče (to reč)’ vs. *(an) ćë dëlat* ‘(on) bo delal’), in glagolskih oblik, ki težijo k nenaglašenosti, tj. glagolov z oslabelim pomenom (npr. *(an) má (no lipo hïšo)* ‘(on) ima (eno lepo hišo)’ vs. *(an) ma dëlat* ‘(on) mora delati’). Poleg tega Wackernaglov zakon o stavi zaslonk na drugem mestu v stavku za rezijanščino ne velja več. Pregibne naslonke se posledično ne pojavljajo samo na drugem mestu v stavku, temveč težijo k pojavljanju v t. i. *obglagolskem položaju*, tj. pred ali za naglašeno glagolsko obliko, tj. pred polnopomenskim glagolom, ali pred nenaglašeno glagolsko obliko (če se ta seveda pojavlja), torej tudi na prvem mestu. Nadalje se pregibne naslonke lahko nahajajo v dveh različnih naslonskih nizih, in sicer sta to t. i. »glagolski« naslonski niz, ki lahko vsebuje naslonsko obliko imenovalnika osebnih zaimkov in nenaglašeno glagolsko obliko ali glagolsko obliko, ki teži k nenaglašenosti, ter zaimenski naslonski, ki je lahko sestavljen iz več kot ene oblike naslonskih oblik odvisnih sklonov osebnih zaimkov in povratnega osebnega zaimka. Naslonska niza sta lahko ločena z eno ali več naglašenih prvin. Ob odsotnosti slednje pa navidezno oblikujeta en sam naslonski niz zaimenskih in glagolskih naslonk.

Medtem ko je nastanek naslonskih oblik imenovalnika osebnih zaimkov, odprava Wackernaglove stave zaslonk na drugem mestu in težnja k obglagolskemu položaju pregibnih naslonk v stavku ter posledičen razcep prvotnega naslonskega niza v »glagolski« in zaimenski naslonski niz v rezijanščini najverjetneje posledica bližnjega romanskega jezikovnega vpliva, pa preoblikovanje pregibnih naslonk v samem naslonskem nizu zelo verjetno odraža težnjo

po premikanju oz. »dviganju« glagolskih naslonk (angl. *clitic climbing*) v položaj pred zaimenske naslonke, ki je značilen za širši slovenski jezikovni prostor.

Kot je znano, glagolske naslonke, ki so se v zgodnjem popraslovanskem obdobju oblikovanja posameznih slovanskih jezikov v naslonskem nizu pojavljale za zaimenskimi naslonkami (npr. **mu ga sem*/*si*/*je*/.../*so*), v slovenskih narečjih izkazujejo težnjo premikanja iz prvotnega položaja na koncu naslonskega niza proti začetku slednjega, torej v položaj pred zaimenske naslonke (npr. **sem*/*si*/*je*/.../*so mu ga*). Medtem ko osrednja slovenska narečja in s tem posledično tudi slovenski knjižni jezik v trdilnih stavkih odražajo vmesno stopnjo premikanja glagolskih naslonk v položaj pred zaimenske naslonke – v nikalnih stavkih je ohranjen starejši besedni red, pri čemer pa je zaradi skrčenja nikalnega členka **ne* in pomožnika prišlo do nastanka novih naglašenih skrčenih oblik –, so se glagolske naslonke v rezijanskem narečju že v celoti premaknile v položaj pred neimenovalniške naslonske oblike osebnih zaimkov in povratnega osebnega zaimka. Tako v knjižni slovenščini pomožnik *bíti* ohranja prvotno mesto v tretji osebi ednine in v vseh oblikah prihodnjika v trdilnih stavkih ter v vseh oblikah sedanjika in prihodnjika v nikalnih stavkih (npr. knj. sln. *brȁt mu ga* ***je*** *dâl*, *brȁt mu ga* ***nī*** *dâl*, *bratje mu ga* ***nīso*** *dâli*; *brȁt mu ga* ***(ne) bo*** *dâl*, *brátje mu ga* ***(ne) bodo*** *dâli*), a se je premaknil pred zaimenske naslonke v vseh ostalih oblikah sedanjika v trdilnih stavkih (npr. knj. sln. *brátje* ***so*** *mu ga dâli*). Nasprotno pa se v rezijanščini glagolske naslonke nahajajo v položaju pred neimenovalniškimi naslonskimi oblikami zaimkov (npr. rez. *bratar an* ***jë*** *mu ga dal*, *bratar an* ***nǐ́*** *mu ga dal*, *bratri ni* ***so*** *mu ga dali*, *bratri ni* ***nǐ́so*** *mu ga dali*; *bratar an* ***(ni) bo*** *mu ga dal*, *bratri ni* ***(ni) bodo*** *mu ga dali*).

Matej Šekli
University of Ljubljana, Faculty of Arts
ZRC SAZU, Fran Ramovš Institute of the Slovene Language, Slovenia
matej.sekli@guest.arnes.si

DEFINING GEOGRAPHICAL LINGUISTIC BORDERS WITHIN THE FRAMEWORK OF GENEALOGICAL LINGUISTICS: THE CASE OF SLOVENE AND CENTRAL SOUTH SLAVIC

The contribution aims to illustrate the theory and methodology of defining geographical linguistic borders within the framework of genealogical linguistics, using the example of the geographical linguistic border between Slovene (i.e. North-Western South Slavic), and Central South Slavic, which includes Kajkavian, Čakavian, Western and Eastern Štokavian.

In Slavic linguistics, the genealogical linguistic classification of geolects (i.e. geographical or spatial linguistic varieties), more specifically languages as well as their dialect macro-areas and dialects, has long been established as the standard approach in the classification of lects. Such dialect classification is based on linguistic criteria, taking into consideration only the linguistic features of the individual geolects. In this respect it is not solely the geographical distribution of the various linguistic features (such as mirrored by the isoglosses) that is considered but also the relative chronology of their nascence. Historical phonetics/phonology is undeniably the most important criterion for accurate genealogical linguistic classification as far as different levels of the linguistic system are concerned, being also the most systematic among the processes of linguistic change.

The national border between Slovenia and Croatia does not coincide with the linguistic boundary between Slovene and Central South Slavic, specifically Kajkavian and Čakavian. The dialectal delimitation between Slovene on the one hand and Kajkavian and Čakavian dialects on the other has been discussed in the last decades (cf. Vranić 2006, 2009; Šekli 2013; Gostenčnik 2018, 2023, 2024; Eterović 2019 etc.). However, many question concerning the mentioned linguistic border remain open. The contribution attempts to delimit Slovene from Central South Slavic, basing on the linguistic material of the Slovenski linguistični atlas 'Slovene Linguistic Atlas' (2011–) and using the theoretical-methodological framework of genealogical linguistics.

Valentina Teršek
Univerza v Ljubljani, Naravoslovnotehniška fakulteta, Slovenija
vt33913@student.uni-lj.si

Primož Weingerl
Univerza v Ljubljani, Naravoslovnotehniška fakulteta, Slovenija
primoz.weingerl@ntf.uni-lj.si

Mojca Schlamberger Brezar
Univerza v Ljubljani, Filozofska fakulteta, Slovenija
mojca.schlambergerbrezar@ff.uni-lj.si

Urška Stanković Elesini
Univerza v Ljubljani, Naravoslovnotehniška fakulteta, Slovenija
urska.stankovic@ntf.uni-lj.si

IZDELAVA INTERAKTIVNEGA TESNIÈRJEVEGA LINGVISTIČNEGA ATLASA ZA ŠTUDIJ DVOJINE V SLOVENŠČINI

V sodelovanju s Filozofsko fakulteto nastaja prototip interaktivnega lingvističnega atlasa, ki vključuje digitalizacijo lingvističnih kart. Projekt obsega analizo podobnih rešitev, raziskavo uporabnikov, načrtovanje uporabniške izkušnje, oblikovanje vmesnika in izdelavo prototipa. Namen je izdelati interaktivni atlas dvojine v slovenščini, ki bo uporaben v šolah, na fakultetah in širše, ter prispeval k ohranjanju kulturno-jezikovne dediščine.

Teoretični okvir in cilj: Lucien Tesnière, ki je bil prvi lektor za francoski jezik na pravkar ustanovljeni Univerzi v Ljubljani med letoma 1920 in 1924, je v tem obdobju zbral gradivo za doktorsko disertacijo o dvojini v slovenskem jeziku, pri čemer je informacije o jezikovnih oblikah vnašal na geografske karte. Tesnièrjeva doktorska disertacija, izdana leta 1925 v dveh knjigah, *Oblike dvojine v slovenščini* in *Lingvistični atlas za študij dvojine v slovenščini*, predstavlja dragoceno dediščino. Leta 2022 je bil Lingvistični atlas preveden v slovenščino (Tesnière 2022: 7). Namen raziskave je digitalizirati Tesnièrjev Lingvistični atlas za študij dvojine v slovenščini in izdelati njegovo interaktivno različico, ki bo v nasprotju s tiskano verzijo dostopna na svetovnem spletu ter bo tako omogočila enostavnejšo in uporabniku prijaznejšo dostopnost širšemu krogu uporabnikov.

V Sloveniji je na primer na voljo interaktivni atlas slovenskih narečij (iSLA), izdelan med letoma 2020 in 2024, ki je povezan z dialektološko podatkovno zbirko DIAtlas in je nadgradnja tiskanih publikacij SLA 1, SLA 2 in SLA 3, omogoča pa interaktivni prikaz

jezikovnega gradiva (Škofic 2013). Podoben projekt so leta 2022 realizirali v Franciji z zvočnim atlasom regionalnih jezikov Francije (Mareüil 2018).

Metodologija: Izdelavo delujočega interaktivnega prototipa smo izvedli po naslednjih korakih z navedeno opremo:

1) Zasnova in analiza je vključevala zbiranje podatkov ter določitve obsega in načina digitalizacije. Pregledan je bil lingvistični atlas in izbranih 65 zemljevidov za vektorizacijo. Pregledane so bile obstoječe rešitve vizualizacij informacij, npr. SLA ter zvočni atlas regionalnih jezikov Francije. Sledila je analiza mnenj uporabnikov (študenti na NTF in FF ter učitelji), ki smo jih kontaktirali z izdelanimi vprašalniki in intervjuji. Na podlagi rezultatov so bile kreirane različne konceptualne rešitve, nato pa je bila izdelana informacijska arhitektura.

2) Načrtovanje uporabniške izkušnje je zajemalo izdelavo mrežnih diagramov oz. skeletov v programu Figma. Najprej smo naredili manj podrobne (ang. *Low Fidelity*) in nato še bolj podrobne (ang. *High Fidelity*) skelete. Nadalje smo načrtovali interaktivnosti med atlasom in uporabnikom.

3) Oblikovanje uporabniškega vmesnika je potekalo skozi oblikovanje grafičnih elementov v programu Adobe Illustrator, izbiro tipografije (posebnosti, ki jih ima slovenščina v zapisu besed, so privedle do pisave iz istoimenskega vnašalnega sistema, ki ga je na SAZU razvil Peter Weiss – ZRCola), oblikovanjem ikonografije ter izdelavo in obdelavo slikovnih predlog v programu Adobe Illustrator. V programu Figma so se nato oblikovale posamezne komponente in posamezne strani za velikost zaslona 1728x1117px ter naredile povezave med njimi.

(Predvideni) rezultati: Projekt napreduje skladno z načrtom, čeprav se srečujemo z izzivi, kot so poravnava skeniranih kart, risanje vektorjev in vzpostavljanje interaktivnosti. Z dobro organizacijo in tehnično podporo razvijamo funkcionalen prototip, ki bo temelj za interaktivni Tesnièrjev atlas. Takšna vizualna komunikacija bo pripomogla k boljšemu razumevanju jezikovnih značilnosti slovenščine in olajšala poučevanje ter učenje jezikovnih pravil.

Suzana Todorović
Univerza na Primorskem, Pedagoška fakulteta, Slovenija
suzana.todorovic@upr.si

SLA IN SiLA KOT VIRA ZA OPAZOVANJE SPREMINJANJA POSEBNOSTI SLOVENSKIH ISTRSKIH GOVOROV

V prispevku se naslanjamo na dve geolingvistični študiji oz. raziskavi – na nacionalni lingvistični atlas SLA in subregionalni lingvistični atlas SiLA, osrednji del raziskave pa se umešča v področje spreminjanja narečij, ki v slovenski dialektologiji še ni sistematično raziskano.

Slovensko istrsko narečje, ki ga je v času fašizma želela zadušiti italijanščina, po vojni – v času tako imenovane (jugo)slovenizacije – pa izpodriniti ideja o nujnosti vzpostavitve čim bolj enotnega (slovenskega) prostora in jezika z brisanjem regionalnih posebnosti, se sicer od devetdesetih let 20. stoletja spet sliši tudi v javnem prostoru, a se neustavljivo spreminja. V preteklosti zapisano besedje priča o nekdanjem življenju njegovih govorcev, zato je nedvomno pomembno tudi za zgodovino, etnologijo, antropologijo in sorodne vede, obenem pa dialektologom omogoča sledenje oz. opazovanje spreminjanja posebnosti narečja v času: tj. od najstarejšega verodostojnega zapisa do najnovejših. Starejše gradivo zato današnjim jezikoslovcem in dialektologom ponuja dragocene podatke za raznolike diahrone primerjave.

Pri analizi (nekdanjih) posebnosti istrskih govorov izhajamo iz gradiva, ki ga je za *Slovenski lingvistični atlas* (SLA) v Istri zapisal Tine Logar, in iz splošnih glasoslovnih značilnosti izbranih govorov, ki jih je po terenskih raziskavah vsakokrat zabeležil in povzel. V raziskavi nas zanima, katere glasovne značilnosti, ki jih je zapisal Logar, so govorci ohranili, zato starejše zapise (stare približno sedemdeset let) primerjamo z novejšimi, torej z gradivom, pridobljenim s terenskimi raziskavami, izvedenimi v istih krajih v letih 2013–2023 (*Slovenski istrski lingvistični atlas* – SiLA 1 in 2). V primerjalni analizi se osredotočamo predvsem na posebnosti dveh istrskih govorov, ki se umeščata v šavrinsko in rižansko podnarečje slovenskega istrskega narečja. Na več simbolnih kartah prikazujemo morebitno opuščanje, stapljanje ali ohranjanje izbranih glasovnih posebnosti.

Predvidevamo, da so izrazite glasovne posebnosti, po katerih ločimo rižanske in šavrinske govore, še vedno ohranjene, na primer diftongi v rižanskih in monoftongi v šavrinskih govorih, druge posebnosti pa govorci najverjetneje postopoma opuščajo, npr. razvoj *y* iz *u*, in sicer nekoč *f'ryška*, danes *f'ruška* ‘hruška’, na kar je v poročilu iz leta 1956 opozoril tudi Logar.

Fruzsina S. Vargha
HUN-REN Hungarian Research Centre for Linguistics, Hungary
fruzsa@gmail.com

FINDING GEOCENTERS OF LINGUISTIC SIMILARITY – A NEW APPROACH TO THE CLASSIFICATION OF HUNGARIAN DIALECTS

Dialectometry, the method dedicated to reveal structures hidden in linguistic atlas data, has become one of the most evolving fields of dialect geography in the last decades (Goebl 2023). Several computational methods have been developed for the automatic classification of dialects, based on linguistic similarity/distance matrices derived from atlas data, i.e. linguistic data linked to geographical places. Since the early days of computational dialectology, it has been claimed that both geographic and linguistic information are relevant to the classification of dialects or the delimitation of dialect areas (Inoue 1996). However, most techniques, such as cluster analysis or multidimensional scaling, do not take geographic data into account, yet the results of the purely linguistic analysis are mapped.

Geographical information is sometimes involved in the analyses, but to a limited extent, when the correlation between linguistic and geographical distances is calculated and mapped (e.g. Goebl 2005), or with techniques that compute only the linguistic similarity of neighboring investigation points. Examples are the honeycomb map and the beam map methods (Goebl 2010), but the selection of neighboring localities is itself sometimes problematic (Inoue 1996).

In the most recent dialectometric classifications of Hungarian dialects, which are based on linguistic or ethnographic atlas data, geographical information was not considered at all (Vargha 2024). This paper aims to present a computational (dialectometric) method to determine, based on both linguistic and geographical information, the “dialect home”, calculated as the geocenter of dialect similarity, for each location of a linguistic atlas; in our case the integrated database of the two largest Hungarian atlases, The Atlas of Hungarian Dialects and The Atlas of Hungarian Dialects in Romania, which together cover the whole language area.

A geocenter (weighted geographic center of mass) is expressed as a single point defined by geographic coordinates and is based on both geographic information and linguistic similarity values. The latter are derived from atlas data through the Levenshtein algorithm (cf. Heeringa 2004). The method presented here is based on objective linguistic data, but also inspired by the arrow method of perceptual dialectology (Preston 1999), and is intended to be used to identify core dialect areas as well as dividing lines that shape the dialect continuum. A definition of a dialect dividing line is presented.

Factors impacting the computation of geocenters (weighted or unweighted, number of similar dialects to be considered, the issue of dialect islands, and also linguistic aspects determining the similarity patterns) and the resulting classificatory consequences are discussed and illustrated. It is shown that the confrontation of neighboring locations is not always reliable for finding dialect boundaries. The method seems to be particularly efficient in proposing boundaries even where classical approaches have indicated large transitional areas.

Philip Vergeiner
Ludwig Maximilian University of Munich, Germany
philip.vergeiner@germanistik.uni-muenchen.de

Philipp Stöckle
Austrian Academy of Sciences, ACDH-CH, Austria
philipp.stoeckle@oeaw.ac.at

GEOGRAPHICAL PATTERNS IN THE BAVARIAN DIALECTS OF AUSTRIA AND SOUTH TYROL: A REAL-TIME COMPARISON USING DIALECTOMETRIC METHODS

The Bavarian dialects of Austria and South Tyrol are among the best-documented varieties of German, benefiting from over a century of data collection. This wealth of data offers a unique opportunity to study dialect dynamics across time (see e.g., Fleischer 2017; Stöckle 2024). However, despite the richness of this data, real-time analyses directly comparing historical and contemporary data remain rare. Most studies of language change rely on apparent-time analyses or compare modern data with older qualitative descriptions (e.g., Bülow et al. 2019; Vergeiner et al. 2021). Even among studies employing real-time comparisons, the focus is typically restricted to individual linguistic features (e.g., Stöckle & Wittibschlager 2022; Vergeiner & Bülow 2024), limiting broader insights into the change of geolinguistic patterns. Dialectometric methods, which aggregate multiple linguistic features to uncover regional structures, are ideally suited for such analyses, yet they have not been applied to long-term comparisons of historical and contemporary dialect data in Austria (and even beyond).

Our study addresses these issues by investigating geographical patterns in phonological variation within the Bavarian dialects of Austria and South Tyrol over the past century, using a quantitative, dialectometric approach. Specifically, we aim to address the following research questions: (1) What geolinguistic patterns can be identified in historical and contemporary datasets? (2) How do these patterns correspond to traditional qualitative dialect classifications? (3) Can we detect evidence of language change by comparing historical and contemporary data?

To answer these questions, we integrate historical dialect data from the 'Dictionary of Bavarian Dialects in Austria' (*Wörterbuch der bairischen Mundarten in Österreich*; see Stöckle 2021) with contemporary data from the project 'Variation and Change of Dialect Varieties in Austria (in Real and Apparent Time)' (see e.g., Vergeiner et al. 2021; Vergeiner & Bülow 2024), supplemented by additional data from the project 'German Dialects in South Tyrol' (*Deutsche Dialekte in Südtirol*; see Scheutz 2016). We analyze a consistent set of 31

phonological variables across these datasets using aggregative dialectometric methods (e.g., Wieling & Nerbonne 2015), including multidimensional scaling and cluster analysis.

Our findings reveal a generally strong correspondence with traditional dialect classifications across both datasets (e.g., Kranzmayer 1953). Moreover, a comparison of historical and contemporary data demonstrates remarkable stability in the broad geolinguistic structures of Bavarian dialects, with enduring divisions such as those between (South) Central and South Bavarian. However, we also identify local changes in specific areas, including the retreat of South Bavarian in southeastern Austria and increasing regionalization within the dialect continuum. In our talk, we will further discuss potential reasons for these changes as well as the overall stability observed in the data.

Anna Westerberg
Institute for language and folklore (Isof), Uppsala, Sweden
anna.westerberg@isof.se

THE SCANDINAVIAN DIALECT CONTINUUM IN THEORY AND PRACTICE

The traditional Swedish dialects are commonly divided into six main groups, although there are no sharp boundaries between them: South Swedish dialects, Göta dialects, Svea dialects, Norrland dialects, East Swedish dialects, and Gotlandic dialects (Wessén 1935). By and large, the same classification can be applied to the regional varieties of Standard Swedish spoken by most people today (Elert 1994). As these regional varieties preserve some characteristics from the old dialects, especially in their phonology, they still form a continuum within the Swedish language area. However, it can be discussed whether this Swedish continuum of regional varieties also fits into a Scandinavian one, or whether there is such a thing at all. In this paper, some of the linguistic transition zones between the dialects and regional varieties within the current borders of Sweden and those outside of them are briefly described, with special attention to the southern end of the area. This leads to the issue of mutual intelligibility between Danish, Norwegian and Swedish: does the Scandinavian language community really exist? Historically, it has been assumed that most speakers of Danish, Norwegian and Swedish can relatively easily learn to communicate with one another in their respective native languages, but in recent years, this view has been called in question. In a survey commissioned by the Nordic Council of Ministers (Skjold Frøshaug & Stende 2021), approximately 1200 people aged 16-25 in Denmark, Norway and Sweden were asked about their views on language, culture and perceived mutual comprehension of Scandinavian languages. The results unambiguously indicate that comprehension of neighbouring languages is particularly challenging in the case of Danish and Swedish, whereas young Norwegians and Swedes understand one another fairly well. On the other hand, young Norwegians have less trouble than Swedes understanding Danish. The same relative differences in intercomprehension were observed in three earlier studies: Maurud 1976, Delsing & Lundin Åkesson 2005, and Brink 2016. Nevertheless, it's also evident that there has been an overall decline in comprehension over time, especially in Denmark and Sweden. In the 2020's, young Scandinavians clearly understand English better than the neighbouring languages.

Melita Zemljak Jontes
Univerza v Mariboru, Filozofska fakulteta, Slovenija
melita.zemljak@um.si

Mihaela Koletnik
Univerza v Mariboru, Filozofska fakulteta, Slovenija
mihaela.koletnik@um.si

INSTRUMENTALNO-SLUŠNA DOLOČITEV KVALITETE ODRAZOV ZA *$\bar{ě}$ IN *$\bar{e}$ V PRLEŠKEM NAREČJU

Prispevek na podlagi fonetično transkribiranega gradiva za SLA in dostopnih zvočnih posnetkov za izhodiščnoslovenska glasova stalno dolgi jat (*$\bar{ě}$) in stalno dolgi *e* (*$\bar{e}$) obravnava fonetično vrednost *e*-jevskih glasov kot odrazov zanju v prleškem narečju.
Pregled fonetične transkripcije prleškega gradiva za SLA za izhodiščno slovenska *$\bar{ě}$ in *$\bar{e}$ ter primerjava s fonetičnim zapisom v dialektološki literaturi kaže na nekatera odstopanja. Po Ramovšu (1935: 177) se je *$\bar{ě}$ v prleškem narečju razvil v $\hat{e}$, tj. »zelo ozek in napet *e*, a brez *i*-jevske barve«, *$\bar{e}$ pa v $\hat{ẹ}$, ob Ščavnici že tudi v $ẹ$ (prav tam: 178). Po Kolariču (1968: 633) je odraz za *$\bar{ě}$ v prleškem narečju dolgi ozki e ($ẹ$), za *$\bar{e}$ pa dolgi nevtralni e ($\hat{ẹ}$). Zorkova piše (2009: 242), da se »v večini govorov prleškega narečja pojavljata dva ozka e-ja: navadni ozki $ẹ$ je refleks za dolgi nosni *en*, za dolgi etimološki *e* in za dolgi polglasnik, medtem ko je zelo ozki *e* refleks za dolgi jat«. Gradivo za SLA (v mrežo krajev so vključene naslednje krajevne točke: T369 Brengova, T370 Sveti Jurij ob Ščavnici, T371 Veržej, T372 Križevci pri Ljutomeru, Gajševci, T373 Bučkovci, T374 Ljutomer, T375 Gibina, T376 Sveti Tomaž, T377 Miklavž pri Ormožu, T378 Juršinci, T379 Podvinci, T380 Cvetkovci, T381 Ormož in T382 Središče ob Dravi) pri evidentiranem besedju za *$\bar{ě}$ in *$\bar{e}$ izkazuje prleško enoglasniško naravo samoglasnikov, ob tem pa še: za *$\bar{ě}$ štiri kakovostne različice e-jev ($ẹː$, $ėː$, $\hat{ẹ}ː$ in $eː$) ter kolikostno opozicijo ($ẹː$: $ẹ$), za *$\bar{e}$ pa tri kakovostne različice e-jev ($eː$, $\hat{ẹ}ː$ in $ẹː$).

Na izbranih primerih bo opravljena instrumentalno-slušna analiza dostopnih zvočnih posnetkov. Pri tem bodo analizirani samoglasniki razmejeni od glasovnega okolja, s spektralno analizo formantov, po potrebi tudi v povezavi s kvantiteto, z naglašenostjo, zlogovno zgradbo besed in glasovnim okoljem, pa bodo določene njihove kvalitativne značilnosti. S pomočjo pridobljenih rezultatov analize glasov se bo tako skušalo pojasniti vzroke za razlike v kvaliteti ali vsaj utemeljiti njihov obstoj v obravnavanem narečnem govoru.

Prispevek je nastal v okviru raziskovalnega projekta *Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik* (J7-4642), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

Danila Zuljan Kumar
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
Univerza v Novi Gorici, Fakulteta za humanistiko, Slovenija
DZuljan@zrc-sazu.si

Klara Šumenjak
Univerza na Primorskem, Fakulteta za humanistične študije, Slovenija
klara.sumenjak@fhs.upr.si

NEKAJ BESEDNOREDNIH ZNAČILNOSTI V KRAŠKEM GOVORU KOPRIVE

V prispevku bo obravnavan govor Koprive na Krasu, ki spada v kraško narečje primorske narečne skupine. To se govori po zahodnem delu Krasa in v spodnji Vipavski dolini. Na vzhodu meji z notranjskim narečjem, zaradi česar se v koprivskem govoru pojavljajo tudi notranjske glasoslovne značilnosti (Šumenjak 2013).

Glasoslovne in oblikoslovne značilnosti govora so bile natančneje predstavljene v Šumenjak (2013, 2016 in 2019), v pričujočem prispevku pa se bomo posvetili problematiki besednega reda kot prvini skladenjskega opisa, za katero je dolgo veljalo, da je slabo raziskano področje v slovnici. Na to je prvi opozoril Anton Breznik v svoji razpravi Besedni red v govoru leta 1908 (Breznik 1982: 233). Pred njim pa velja omeniti še Matijo Murka, ki se je s svojima razpravama o enklitikah v slovenščini (1891, 1892) dotaknil tudi besednorednih zakonitosti slovenskega jezika. Najnatančneje je vprašanje besednega reda v slovenskem pisnem jeziku obdelal Jože Toporišič, posebej v zadnji izdaji svoje *Slovenske slovnice* (2000).

Tudi v slovenski dialektologiji področje besednega reda ni pritegnilo širše pozornosti. Do danes so se tematiki natančneje posvetili Zorko (1994) v severovzhodnih slovenskih narečjih, Zuljan Kumar v nekaterih primorskih narečjih (2003, 2008, 2019, 2022) ter Šekli (2025).

V prispevku bodo natančneje analizirane naslednje besednoredne značilnosti govora Koprive:

1) stava ujemalnega pridevniškega prilastka v samostalniški besedni zvezi,
2) izpostavljeni stavčni člen,
3) besedni red v povedi, ki se začne s prislovnim določilom ali členkom,
4) besedni red v večstavčni podredno zloženi povedi,
5) stava prostih naslonk in njihovo zaporedje v naslonskem nizu,
6) stava naslonk v nikalnih velelnih stavkih.

Andrejka Žejn
ZRC SAZU, Inštitut za slovensko literaturo in literarne vede, Slovenija
andrejka.zejn@zrc-sazu.si

LUCIEN TESNIÈRE IN KOROŠKA NAREČJA

Francoz Lucien Tesnière, po Sreznjevskem prvi tuji raziskovalec slovenskih koroških narečij, je študiral na več evropskih univerzah, nekaj mesecev tudi slavistiko na dunajski univerzi, in bil v letih 1920–1923 lektor za francoski jezik na Univerzi v Ljubljani, kjer je takrat deloval Fran Ramovš. Tesnière je prišel v stik s slovenščino in koroškimi narečij v burnem času koroškega plebiscita ter deloval na ljubljanski univerzi v obdobju, ko se je na Slovenskem vzpostavljala slovenistika kot znanost. K tej je Tesnière prispeval z disertacijo o dvojini v slovenskih narečjih, v katero je vključil tudi gradivo z današnje avstrijske Koroške in podal prodoren vpogled v tamkajšnjo jezikovno situacijo. Gradivo za Koroško je zaradi poplebiscitnega vzdušja na Koroškem zbral večinoma pri intelektualcih, ki so izhajali iz teh območij in so takrat bivali v Ljubljani. Koroško gradivo je upošteval tako v doktoratu, ki ga je zagovarjal leta 1925 v Strasbourgu in je bil istega leta objavljen v knjižni obliki (monografija o dvojini v slovenščini in jezikovni atlas), kot v svojih drugih zgodovinskojezikoslovnih in dialektoloških razpravah.

V prispevku bo podrobneje analizirano Tesnièrjevo gradivo za koroška narečja z dveh vidikov. Z vidika jezikovne geografije bo analizirana mreža krajev na Koroškem (z izbiro informantov in načinom zbiranja gradiva) ter primerjana z mrežami za kasnejše sorodne jezikovnogeografske projekte. Z vidika gradiva pa bodo primerjani zapisi sorodnih pojavov v koroških narečjih, ki so nastajali v okviru drugih vsebinsko in metodološko raznovrstnih raziskav. Na ta način bo Tesnièrjev prispevek prikazan v kontekstu drugih raziskav slovenskih narečij na avstrijskem Koroškem.