Balcania et Slavia

Studi linguistici | Studies in Linguistics

Edizioni
Ca'Foscari

Balcania et Slavia
Studi linguistici | Studies in Linguistics

Editor-in-Chief
Iliyana Krapova

Balcania et Slavia
Studi linguistici | Studies in Linguistics

Semestral journal

Table of Contents

Introduction

Gjorgje Bozhoviq, Aleksandar Trifunović
Independent scholars

Iliyana Krapova
Università Ca' Foscari Venezia, Italy

Summary 1 Fundamentals and Advances in Balkan Linguistics (II). – 2 Independent paper.

1 Fundamentals and Advances in Balkan Linguistics (II)

This issue of *Balcania et Slavia*, intended as a continuation of Volume 4, Issue 1 (2024), brings four additional papers based on the presentations given at the first international meeting in the *Fundamentals and Advances in Balkan Linguistics* (FABL) series, held at the University of Belgrade on November 16–18, 2023.

As was the case with the previous selection, this time, too, all major branches of Balkan languages, as well as some peripheral members of this linguistic area, are included: viz. Balkan Slavic, Albanian, and Balkan Romance (namely, Judeo-Spanish), as well as Turkish. The four papers collected in this volume tackle the Balkan linguistic data in its wider context of language use and function, illuminating it from different methodological perspectives. Thematically, all four papers deal with semantics, or its interfaces with grammar or discourse and society, in four language- or dialect-specific situations and in four particular domains. Methodologically, they apply different frameworks and tools to analyse meaning construction and representation in the domains of evidentiality, spatial relations, language naming, and tense: i.e. genre and narrative analysis, contrastive and cognitive linguistics approach, sociolinguistics, and formal methods.

Check for
updates

Merima Krijezi and Predrag Mutavdžić (University of Belgrade) co-author a paper on the semantics of the Albanian accusative preposition *në*, mostly based on Krijezi (2023). As the authors point out, *në* is one of the most frequent prepositions in Modern Albanian, primarily expressing spatial relations, as well as having a number of secondary non-spatial meanings. In order to disentangle the primary and secondary meanings of *në*, its semantic scope, as well as its combinatory potential, the authors put forward an analysis rooted in the morpho-syntax, semiology, and cognitive linguistics. As such, the paper is indicative of the kind of linguistic research conducted currently under the auspices of the Belgrade cognitive linguistics school, deriving ultimately from late professor Ranko Bugarski's (1933–2024) oeuvre and his lasting influence (cf. Bugarski 1969).

Vesela Simeonova (University of Graz) and Seçkin Arslan (University of Côte d'Azur and CNRS) present an analysis of Turkish and Bulgarian evidentiality across seven written genres: news, history, police reports, fiction, fables, religious texts, and dreams. This interesting and acribic study, based on a substantial corpus of empirical data, shows that evidentials are restricted to different narrative modes. In four of the seven environments examined, evidentials are used identically in both languages. Their use in news and general knowledge and the report category differs, however, in Bulgarian and Turkish. The paper continues and complements the line of research on evidentiality and areal contact between Turkish and languages of the Balkan area, initiated and championed by Victor A. Friedman (e.g. Friedman 1999; 2008; cf. now also Friedman & Joseph 2025).

Ana Stulic's (University Bordeaux-Montaigne) paper on Judeo-Spanish language and its linguistic culture, based on an invited talk she delivered at the 2023 FABL meeting, is a rare contribution focusing on this minority Balkan language and the culture of its speakers. While discussing the three linguistic ideologies that historically have shaped the language use and multilingual practices among the Sephardim in the Balkans, this paper provides a framework for understanding the range of terms employed by both present-day speakers and scholars to designate the language of Sephardic Jews: viz. *Ladino*, *Judeo-Spanish*, and *Judezmo*. As argued by the author, these ideas are crucial for understanding both the external and internal histories of Judeo-Spanish, as well as its role in language contact settings in the Balkans. While looking at the intriguing case of Judeo-Spanish, one cannot but think of the nonminority, albeit not less intriguing case of Serbo-Croatian, and other languages of the Balkans, where various linguistic ideologies, national and religious identifications, and overall linguistic cultures of their speakers are intertwined with language naming issues. Although deeply rooted in

the Judeo-Spanish case study, Ana Stulic's paper, therefore, has the potential for outlining a wider Balkan research paradigm.

Last but not least, Jelena Živojinović (University of Graz) and Natalia Jardón Pérez (Central European University) co-author a paper on the expression of past events in western South Slavic. Investigating how past events are expressed in the balkanised Torlak, as opposed to the standard variety, this paper identifies semantic and morpho-syntactic differences between the aorist, the imperfect, and the perfect in these closely related varieties. Analysing their distribution and usage patterns, with a focus on the aorist/perfect divide, the paper shows how the expression of past tense in Torlak is related to the broader framework of Balkan-type evidentiality. Thus, in addition to the broad literature on the topic, the present study contributes both empirically and theoretically informed formal account for the already observed fact in traditional dialectology, that the aorist and the imperfect are better preserved and more frequently used in the South-East than in the North-West (e.g. Ivić 1956; 1957).

We extend our gratitude to the editorial team of *Balcania et Slavia* and Edizioni Ca' Foscari for hosting this special issue. As it comes out of press, the second FABL event will have taken place at the University of Belgrade, organized by the Institute for the Study of Antiquity and Classical Heritage on October 31, 2025 (cf. Božović & Ligorio 2025). We are looking forward to another Balkan linguistics meeting in the series and, hopefully, future events.

The year 2025 has been one of the hardest in the history of higher education in Serbia. Amidst unprecedented oppression by the authorities, many of us have been deprived of our institutional affiliations, employment, and salaries. In the streets, if not at our workplaces, we remain in full solidarity with the Serbian university students, who struggle for a fairer and better society.

2 Independent paper

In addition to the four contributions presented in the section prepared by Gjorgje Božović and Aleksandar Trifunović, which derive from the second edition of the international conference *Formal Approaches to Balkan Linguistics (FABL II)* held in Venice in November 2023, the present issue also includes an independent paper by Maksym O. Vakulenko (Lviv).

Vakulenko's study, *Linguistic and Didactic Aspects of Overcoming the Ukrainian Phonetic Crisis*, addresses the current state and challenges of Ukrainian phonetic research, proposing a modernized transcription system based on the International Phonetic Alphabet (IPA). By combining linguistic and didactic perspectives, the author underscores the need to align Ukrainian phonetic scholarship

with contemporary international standards. The paper addresses the decline of experimental phonetic research in Ukraine and the persistence of outdated Soviet-era transcription systems. Vakulenko proposes a modernized conversion of the Cyrillic-based phonetic symbols into the International Phonetic Alphabet (IPA) and provides updated experimental data on Ukrainian vowels, affricates, and the voiced/voiceless consonants [ɣ], [ʕ], and [h]. The study includes detailed tables for transliteration and recommends adopting IPA-based transcription in teaching and research. It also argues that English and German [h] should be represented in Ukrainian loanwords with the letter *x* rather than *г*. The author concludes that revitalizing Ukrainian phonetics requires the systematic integration of contemporary international standards into linguistic education and codification practice.

Together, the papers included in the present volume illustrate the thematic diversity and vitality of the current issue, which bridges Balkan and neighbouring linguistic traditions while continuing the comparative and interdisciplinary spirit of the journal *Balcania et Slavia*.

We take this opportunity to draw our readers' attention to a particularly valuable recent contribution to the field — the comprehensive handbook by Victor A. Friedman and Brian D. Joseph, The Balkan Languages (Cambridge University Press, 2025). This volume represents a milestone in Balkan linguistics, offering the first systematic and accessible synthesis of the Balkan linguistic area written in English. Bringing together decades of scholarship by two of the most authoritative voices in the field, the book combines descriptive depth with theoretical insight and provides a comparative overview of the main structural features shared by the Balkan languages. Its open-access format makes it an indispensable reference for specialists and students alike, ensuring that the study of Balkan linguistics continues to evolve within an inclusive and internationally accessible framework.

The editorial team of *Balcania et Slavia* would like to thank all contributing authors and reviewers for their collaboration and support. This issue continues to affirm the mission of our journal as a space for dialogue among scholars, theoretical frameworks, and linguistic and cultural traditions of the Slavic, Balkans and neighboring regions.

References

Božović, Đ. & O. Ligorio (Eds.) (2025). *Fundamentals and Advances in Balkan Linguistics II: Folk, Areal and Balkan Linguistics: Book of Abstracts*. Belgrade: University of Belgrade. Available at https://reff.f.bg.ac.rs/handle/123456789/7603

Bugarski, R. (1969). *Predlozi* over, under, above, below *i* beneath *u savremenom engleskom jeziku* [*The Prepositions* over, under, above, below *and* beneath *in Contemporary English Language*]. Belgrade: University of Belgrade.

Friedman, V. A. (1999). Proverbial evidentiality: on the gnomic uses of the category of status in languages of the Balkans and the Caucasus. *Mediterranean Language Review* 11: 135–155.

Friedman, V. A. (2008). The typology of Balkan evidentiality and areal linguistics. In: O. M. Tomić (eds.), *Balkan Syntax and Semantics*. Amsterdam: John Benjamins, 101–134.

Friedman, V. A. & B. D. Joseph (2025). *The Balkan Languages*. Cambridge: Cambridge University Press.

Ivić, P. (1956). *Dijalektologija srpskohrvatskog jezika: Uvod i štokavsko narečje* [*The Dialectology of the Serbo-Croatian Language: Introduction and the Štokavian Dialect*]. Novi Sad: Matica Srpska.

Ivić, P. (1957). Značaj lingvističke geografije za uporedno i istorisko proučavanje južnoslovenskih jezika i njihovih odnosa prema ostalim slovenskim jezicima [The importance of linguistic geography for the comparative and historical study of South Slavic languages and their relationship with other Slavic languages]. *Južnoslovenski filolog* 22: 179–206.

Krijezi, M. (2023). *Semantika predloga u savremenom albanskom jeziku* [*The Semantics of Prepositions in Contemporary Albanian Language*]. Belgrade: University of Belgrade.

Мовознавчі та дидактичні аспекти виходу з української фонетичної кризи

Maksym O. Vakulenko
Institute of Problems of Artificial intelligence of MES and NAS of Ukraine;
State Scientific and Technical Library of Ukraine, Kyïv, Ukraine

Abstract   This paper proposes a system for converting transcription symbols from the Soviet-era Cyrillic-based notation into the International Phonetic Alphabet (IPA). It also presents up-to-date phonetic data on the labiovelar approximant [w], the affricates, and the voiceless vocoid [h], which refine and expand existing understandings of Ukrainian phonetics. The obtained results are essential for phonetic research and for teaching phonetics at all levels of Ukrainian language study.

Keywords   Ukrainian phonetics. Phonetic transcription. International Phonetic Alphabet. Latinization. Allophone.

Index   1 Вступ. – 2 Методологія дослідження. – 3 Результати дослідження. – 4 Висновки.

1   Вступ

Фонетичні дослідження та відповідно викладання фонетики в українських і закордонних навчальних закладах мають велике значення для української мови, оскільки її провідною засадою є фонетична (Шевельов 2012, 510). І таких досліджень сьогодні дуже бракує – адже справжній розквіт українських фонетичних досліджень другої половини XX ст. (Білодід 1969; Перебийніс 1970; Тоцька 1981) змінився фазою занепаду. Останні «канонічні» експериментальні фонетичні дані містяться у відомому підручнику Ніни Тоцької 1981 року (Тоцька 1981). У 2008 році було закрито відділ фонетики в Інституті української мови НАН України. Десь із того часу і лабораторія експериментальної фонетики Київського національного університету імені Тараса Шевченка науковою роботою не займається, а виключно обслуговує навчальний процес. Як наслідок, на 17-ій славістичній конференції в Саппоро (вересень 2022 року), куди запрошували всіх українських фонетистів, які мають хоч якісь власні експериментальні результати, було таких тільки двоє – професор Олена Стеріополо й автор цих рядків. У світі фонетичні експерименти проводять за допомогою спеціальних приладів – насамперед із застосуванням ультразвуку (Czaplicki, Ćavar 2024) та артикулографії (Rebernik et al. 2021), а в Україні цього досі немає.

Усе це вказує на те, що, на жаль, в українському мовознавстві так і не сформувалася своя потужна фонетична школа – тож невдовзі після розвалу Радянського Союзу українські фонетичні дослідження у фахових установах України зійшли практично нанівець. Тому фонетична частина сучасної навчально-наукової літератури з української мови й містить переважно матеріал 50-річної давнини, взятий із праць корифеїв – монографій (Білодід 1969; Перебийніс 1970) і навчальних посібників (Жовтобрюх, Кулик 1965; Тоцька 1981). Слід розуміти, що для фонетики це надто великий строк. Та й одержано ці дані від одного-єдиного мовця, за допомогою морально віджилого інструментарію та почасти з порушеннями загальноприйнятої в сучасній фонетиці методики досліджень (див. Vakulenko 2018; 2019). Тому суперечності між прийнятим у світі розумінням базових фонетичних понять і уявленнями, які досі панують в українському мовознавстві, доволі виразні. І ця надмірна залежність від фонетичного спадку радянського часу стала істотною перешкодою належному впровадженню здобутків світової фонетики.

Наприклад, у світі давно не є предметом дискусій той факт, що напружені голосні (англ. *tense vowels*) розташовані далі від центра діаграми голосних, ніж ненапружені (англ. *lax vowels*), оскільки вони потребують більших артикуляційних зусиль. Численні ілюстрації цього явища містяться у фаховій літературі,

зокрема у відомій монографії американського фонетиста Кеннета Стівенса (Stevens 1998, 296). На Мал. 1 подано розташування ненапружених англійських голосних відносно їхніх напружених відповідників, наведене в цій монографії. Оскільки українська мова є консонантною і не має такої розвиненої підсистеми голосних, як англійська (Исаченко 1963), для неї співвідношення між напруженими та ненапруженими голосними зводиться до відповідності між наголошеними та ненаголошеними. Відповідні ілюстрації містяться у відомих роботах (Білодід 1969, 123; Тоцька 1981, 101) і є подібними між собою. На Мал. 2 подано діаграму наголошених і ненаголошених українських голосних із монографії (Білодід 1969, 123).

Із порівняння Мал. 1 і Мал. 2 бачимо, що відповідають загальномовним фонетичним закономірностям тільки співвідношення між наголошеними і ненаголошеними голосними [а у і и]. З незрозумілих причин ненаголошений [е] опинився далі від центра, ніж наголошений, а ненаголошений [о] – вище від наголошеного.
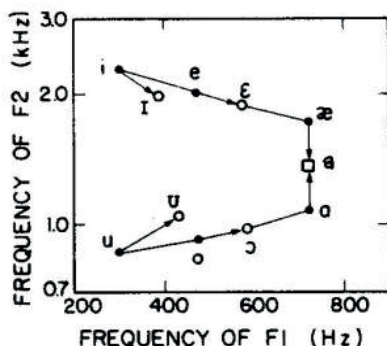
Мал. 1
Співвідношення між напруженими та ненапруженими голосними за Стівенсом (Stevens 1998, 296)
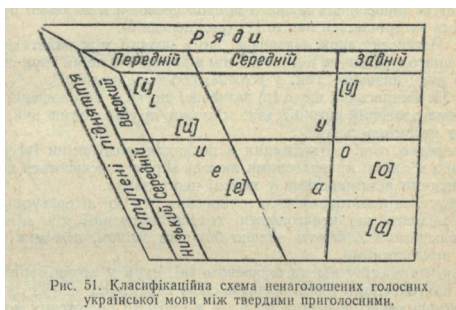
Мал. 2 Співвідношення між наголошеними та ненаголошеними голосними за Тоцькою (Білодід 1969, 123)

Ці помилки справляють відчутний негативний вплив на українську мову, оскільки призводять до хибних орфоепічних рекомендацій (Погрібний 1984, 10-11).

Тепер порівняймо означення форманти, прийняте у світі та в Україні.

Девід Крістал означує форманту так: «a formant is a concentration of acoustic energy, reflecting the way air from the lungs vibrates in the vocal tract, as it changes its shape» (форманта – це концентрація акустичної енергії, яка відображає те, як повітря з легенів вібрує в голосовому тракті, коли він змінює свою форму. – *Авторський переклад*) (Crystal 2008, 196). Це далеко не ідеальне тлумачення, але воно принаймні має сенс. А в енциклопедії «Українська мова» бачимо зовсім альтернативний підхід: «область посилення частот називається формантою» (Тоцька 2004, 105; див. також Тоцька 1981, 42). Це означення не має сенсу, адже посилюватися може те, що має силу. А частота сили не має.

Виразною ознакою невідповідності української фонетичної літератури сучасним вимогам є й те, що в ній досі застосовують віджилу радянську систему фонетичної транскрипції на базі кирилиці, що суперечить загальноприйнятій у світі практиці використання Міжнародного фонетичного алфавіту (МФА) (Vakulenko 2018; 2019). З огляду на те, що на заняттях з іноземних мов українські учні та студенти все одно знайомляться з фонетичною транскрипцією на базі МФА, ця радянська система транскрибування видається навряд чи потрібною. Крім того, така практика дуже ускладнює завдання практичної транскрипції, необхідні для кодифікації запозичень. Це перешкоджає взаєморозумінню з зарубіжними колегами й ускладнює процес викладання української мови як іноземної.

2   Методологія дослідження

Як уже було зазначено, фонетичні дані з монографії (Білодід 1969) та навчальних посібників (Жовтобрюх, Кулик 1965; Тоцька 1981) досі залишаються основою для написання фонетичних розділів сучасних підручників з української мови. У роботах (Vakulenko 2018; 2019) на основі всебічного аналізу цих та інших даних і з урахуванням результатів власних фонетичних експериментів було встановлено позначення МФА для українських фонем. Ці результати відкрили принципову можливість систематично використовувати для звуків української мови міжнародні позначення. На жаль, українські мовознавці – на відміну від зарубіжних – мало цікавляться англомовними публікаціями. Тому в Україні практично не знають, як правильно записати звуки української мови за міжнародними нормами, і продовжують

користуватися застарілими радянськими правилами.

Мета цієї статті полягає в тому, щоб сформулювати правила спрощеної транслітерації радянської системи фонетичної транскрипції звуків української мови знаками МФА, а також навести отримані останнім часом експериментальні факти, які істотно розвивають і уточнюють дотеперішні уявлення про українську фонетику.

Для переведення знаків радянської фонетичної транскрипції у знаки МФА застосовано *метод транслітерації* з урахуванням *методу порівняння* на основі опису відповідних звуків мовлення у роботах (Vakulenko 2018; 2019; 2023a). Крім того, для уточнення уявлень про українську фонетику використано результати, одержані *експериментальним методом* у роботах (Vakulenko 2023b; 2024; Czaplicki, Ćavar 2024).

Матеріалом дослідження є фонетичні дані щодо звуків української мови, а також англійської та німецької мов.

3 Результати дослідження

3.1 Латинізація фонетичних знаків радянської школи

Таблиця 1 містить правила латинізації фонетичних позначень для голосних звуків української мови.

Таблиця 1 Голосні звуки

Позначення радянської фонетичної школи	Позначення Міжнародного фонетичного алфавіту	Примітка
a	ɑ	Під наголосом після голосних і твердих приголосних і на початку слова
a	ɐ	У ненаголошеній позиції та після палаталізованих приголосних
aᵉ	з	Асимільований
o	o	Під наголосом після голосних і твердих приголосних і на початку слова
oᵉ	ɵ	Асимільований
oʸ	ọ	Асимільований
y	u	Під наголосом після голосних і твердих приголосних і на початку слова
y	ʊ	У ненаголошеній позиції та після палаталізованих приголосних
i	i	

и	ɨ+⊤	
иᵉ	ә⁺	У ненаголошеній позиції
е	ɛ	Під наголосом після голосних і твердих приголосних і на початку слова
еᴙ	ɛ̠⁻	У ненаголошеній позиції
еⁱ	е	Асимільований
е	е	Після паталізованих приголосних

У Таблиці 1 знак «⊤» показує знижену артикуляцію відповідного базового звука, що супроводжується зростанням частоти його першої форманти *F1*, а «⊥» означає піднесену артикуляцію базового звука, що супроводжується зменшенням частоти його першої форманти *F1*. Знак «+» показує більш передню артикуляцію відповідного базового звука, що супроводжується зростанням частоти його другої форманти *F2*, а «–» означає більш задню артикуляцію базового звука, що супроводжується зменшенням частоти його другої форманти *F2*.

У роботах (Vakulenko 2018; 2024) було двома різними методами встановлено напівзакритий (напівпіднесений) характер алофонів української фонеми /o/. Як відомо, друга форманта приголосного перед українськими [о у] понижується, що є наслідком огублення (лабіалізації) та велярізації (Тоцька 1981, 96; Вакуленко 2023, 240-1), а перед напіввідкритим звуком [ɔ] (як у слові 'boy') цього пониження немає (Vakulenko 2018). З іншого боку, частота першої форманти ненаголошених алофонів /o/ дещо зростає, а частота другої зростає істотно (Vakulenko 2024). Ці факти свідчать про те, що в українській мові фонема /o/ реалізується такими варіаціями звука [о], які (за відсутності акомодації та асиміляції) розташовуються на діаграмі голосних завжди вище середньої лінії.

Щодо позначення українського звука [и] було взято до уваги такі міркування:
- позначення ɪ («паличка») потрібне для відтягнутого пониженого алофона /i/;
- ненаголошені [и] та [е] подібні на слух;
- в англійській мові звук [ɪ] є ненапруженим відповідником напруженого звука [i], а в українській мові звук [и] не є ненапруженим;
- немає чітких критеріїв розмежування звуків середнього і переднього ряду.

Тому український звук [и] позначено перекресленим просунутим пониженим ɨ⁺ (Vakulenko 2018).

Таблиця 2 містить правила спрощеної транслітерації фонетичних позначень для приголосних звуків української мови.

Таблиця 2 Приголосні звуки

Позначення радянської фонетичної школи	Позначення Міжнародного фонетичного алфавіту	Примітка
п	p	
п', пь	pʲ	
б	b	
б', бь	bʲ	
т	t̪	
т', ть	tʲ	
д	d̪	
д', дь	dʲ	
с	s̪	
с', сь	s̪ʲ	
з	z̪	
з', зь	z̪ʲ	
ш	ʃ	
ш', шь	ʃʲ	
ж	ʒ	
ж', жь	ʒʲ	
х	x	
х', хь	xʲ	
г	ʕ	Перед голосними а, о, у, и, е
г	ɣ	Перед дзвінкими приголосними та в кінці слова
г', гь	ɣʲ	
ф	f	
ф', фь	fj	
ц	t͡s̪	
ц', ць	t͡s̪ʲ	
ц':, ць:	t͡ɪt͡s̪ʲ, t͡ɪ:s̪ʲ	
д͡з	d͡z̪	
д͡з'	d͡z̪ʲ	
ч	t͡ʲʃ	
ч'	t͡ʲʃʲ	
ч':, чь:	tʲt͡ʲʃʲ, t͡ʲ:ʃʲ	
д͡ж	d͡ʲʒ	
д͡ж'	d͡ʲʒʲ	
v	v	
в	ʋ	
в', вь	ʋʲ	
w	β̞	

w', wь	β̞ʲ
ў	β̞
л	ɫ
л', ль	lʲ
м	m
м', мь	mʲ
н	n̪
н', нь	ɲ
р	r
р', рь	rʲ
й	j
ј	j
ї	j

Звук [г] традиційно класифікують як дзвінкий глотковий (фарингальний) фрикативний приголосний [ʕ] (Білодід 1969, 176). Експериментально підтверджено також постання задньоязикового алофона [χ] фонеми /г/ перед приголосними та наприкінці слова (Vakulenko 2019, 42; Czaplicki, Ćavar 2024). Можливий також проривний алофон [g]. Гортанний алофон [ɦ] досі не має переконливих експериментальних підтверджень, хоча він теж теоретично можливий.

У літературі радянського періоду (Білодід 1969; Тоцька 1981) губно-губну реалізацію /в/ позначали латинською літерою 'w', що є помилкою. У сучасній фонетичній класифікації це позначення зарезервоване для губно-задньоязикового апроксиманта (англ. *labiovelar approximant*), як у словах *what*, when тощо, який характеризується дуже низьким значенням другої форманти – близько 700 Гц (Stevens 1998, 517-32) – і сильно витягнутими губами (Vakulenko 2019, 40-1). Своєю чергою, частоти другої форманти українського [в] лежать у діапазоні 1100-500 Гц (Вакуленко 2023, 228), а під час вимови цього звука губи практично не витягуються (Vakulenko 2019, 40-1).

У наших експериментах за допомогою тривимірних спектрограм було уточнено також характер українських африкат. Зокрема, було показано, що перша фаза у вимові «дж» і «ч» являє собою не звичайний, а пом'якшений проривний звук (Вакуленко 2023, 235-6; Vakulenko 2019, 39). Крім того, було спостережено, що у вимові подвоєних пом'якшених африкат подовжується не вся африката, а тільки приступ першої частини, що є достатньою підставою відповідно подавати транскрипції цих звуків (Vakulenko 2019, 43).

Оскільки й інші експериментальні фонетичні дані в літературі радянського періоду можуть бути застарілими або некоректними, ці таблиці мають переважно рекомендаційний характер. Їх варто

застосовувати для наближеної передачі реального звучання українських слів знаками МФА на основі радянських транскрипцій.

Для точної передачі звучання українських слів потрібні додаткові фонетичні експерименти, провести які в Україні можливостей фактично немає. Утім, для приблизної, але коректної передачі звучання можна використовувати систему фонем української мови (Vakulenko 2018; 2019), до якої створено також інструмент автоматичної фонемізації українських текстів (Vakulenko 2022). Цей інструмент доступний на ресурсі GitHub.[1]

3.2 *Передача англійського та німецького звука [h] у нових запозиченнях*

Оскільки, як показано вище, алофони твердої української фонеми /г/ зображуються знаками МФА [ʕ ɣ g] і, можливо, [ɦ], постає закономірне питання: якою українською літерою варто передавати глухий звук [h], який існує в англійській, німецькій та інших мовах?

У дисертації «Сучасна українська термінологія: методологія, кодифікація, лексикографічна практика» (Вакуленко 2023), поданій на здобуття ступеня доктора філологічних наук зі спеціальності 10.02.01 – українська мова, було ґрунтовно досліджено звуки української мови, зокрема проведено порівняння англійського та німецького глухого вокоїда [h] з українськими [x] і [г]. У рамках загального дослідження цієї проблеми було проведено власні фонетичні експерименти з залученням українських і американських інформантів (проєкт за програмою Фулбрайта), досліджено отримані осцилограми та спектрограми, проаналізовано авторитетні джерела – зокрема відомі за кордоном роботи (Stevens 1998; Kawasaki 1982; Boase-Beier, Lodge 2003) та інші. Було проведено порівняння відповідних звуків за допомогою метричного методу вимірювання фонетичних відстаней (Vakulenko 2021). Результати цього дослідження дали змогу зробити науково обґрунтований висновок про фонетичну близькість англійського та німецького глухого вокоїда [h] до українського [x], а відповідне положення щодо англійського звука було винесено на захист. Тож після успішного захисту дисертації воно стало мовним фактом, визнаним в українській лінгвістиці.

Це дає достатні підстави для того, щоб передавати звук [h] у недавніх та нових запозиченнях з англійської та німецької мов за допомогою української «х»: *біхевіоризм*, *ноу-хау*, *хаб*, *хайп*, *хакер*, *хаскі*, *хеджування*, *хедрол*, *хелпер*, *хілер*, *хіпстер*, *хіп-хоп*, *хіт*, *ховер*, *хол*, *хорор*, *хот-дог*, *хуліган*, *хенде хох*, *Беверлі-Хіллз*,

---

1 https://github.com/Mova-2020/Phonemic-transcription-of-Ukrainian-texts.

Хаббл, *Хаймарс*, *Халк*, *Харріс*, *Хевісайд*, *Хелловін*, *Хіггс*, *Хірш*, *Хокінг*, *Хоффенхайм*, *Хюбшнер* тощо. Більшість цих форм уже закріпилися в українській мові. Це правило не поширюється на запозичення, які давно вживаються з «г»: *Гамільтон*, *Гарвард*, *Гаррі*, *Генрі*, *Ноттінгем*, *Саутгемптон*, *Гамбург*, *Ганновер* та інші. Список слів, які традиційно вживаються з «г», слід перевіряти за

Переліком нормативних словників української мови.[2]

Висновки

Українській мові дуже бракує надійних і сучасних фонетичних даних, отриманих із застосуванням новітніх технічних засобів. Це призводить до невиправданого поширення застарілої і почасти хибної інформації, джерела якої містяться у літературі радянського періоду. Головною проблемою тут є невміння коректно класифікувати звуки української мови, зокрема відобразити традиційні кириличні позначення цих звуків, уживаних у літературі радянського періоду, знаками Міжнародного фонетичного алфавіту. Запропоновані таблиці латинізації радянської фонетичної транскрипції та рекомендації щодо передачі англійського та німецького звука [h] українською «х» ставлять процес засвоєння іншомовних слів у міжмовних контактах української мови на наукову основу і роблять його значно ефективнішим і простішим. Їх слід використовувати в наукових дослідженнях і запровадити в навчальні програми з української мови.

---

2   Щодо визначення переліку нормативних видань та словників української мови, які рекомендуються для використання при підготовці та виданні колективних, індивідуальних праць, наукових збірників та періодики НАН України. *Розпорядження Президії НАН України № 161* від 12.03.2013 p. http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILA=&2_S21STR=rbndu_2014_19_15.

Список використаної літератури

Білодід, І.К. (ред.) (1969). *Сучасна українська літературна мова. Фонетика*. Київ: Наукова думка.

Вакуленко, М.О. (2023). *Сучасна українська термінологія: методологія, кодифікація, лексикографічна практика* [Дис. на здобуття наук. ступеня док. філол. Наук]. Київ: Київський національний університет імені Тараса Шевченка, Міністерство освіти і науки України.

Жовтобрюх, М.А.; Кулик, Б.М. (1965). *Курс сучасної української літературної мови*. Ч. 1. Київ: Радянська школа.

Исаченко, А. (1963). «Опыт типологического анализа славянских языков». *Новое в лингвистике III*, 106-21.

Перебийніс, В.С. (1970). *Кількісні та якісні характеристики системи фонем української літературної мови*. Київ: Наукова думка.

Погрібний, М. (1984). *Орфоепічний словник*. Київ: Радянська школа.

Тоцька, Н.І. (1981). *Сучасна українська літературна мова: Фонетика, орфоепія, графіка, орфографія*. Київ: Вища школа.

Тоцька, Н. (2004). «Голосні звуки». Русанівський, В.М.; Тараненко, О.О.; Зяблюк, М.П. та ін. (ред.), *Українська мова: енциклопедія*. [2-ге вид., випр. і доп.]. Київ: Вид-во "Укр. енцикл." ім. М.П. Бажана, 104-5.

Шевельов, Ю. (2012). *Нарис сучасної української літературної мови та інші лінгвістичні студії (1947-1953 рр.)*. Київ: Темпора.

Boase-Beier, J.; Lodge, K. (2003). *The German Language: A Linguistic Introduction*. Oxford: Blackwell Publishing.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. Malden, MA; Oxford, UK; Carlton, Victoria: Blackwell Publishing.

Czaplicki, B.; Ćavar, M. (2024). «Variation in the Realization of Ukrainian Back Fricatives as Onset Lenition and Non-Markedness Reducing Coda Neutralization: A 3D/4D Ultrasound Study». *Glossa: A Journal of General Linguistics*, 9(1), 1-34. https://doi.org/10.16995/glossa.9795

Kawasaki, H. (1982). *An Acoustical Basis for Universal Constraints on Sound Sequences* [PhD Dissertation]. Berkeley: University of California.

Rebernik, T.; Jacobi, J.; Jonkers, R.; Noiray, A.; Wieling, M. (2021). «A Review of Data Collection Practices Using Electromagnetic Articulography». *Laboratory Phonology* 12(1), 6. https://doi.org/10.5334/labphon.237

Stevens, K. (1998). *Acoustic Phonetics*. Mass.: MIT Press.

Vakulenko, M.O. (2018). «Ukrainian Vowel Phones in the IPA Context». *Govor,* 35(2), 189-214. https://www.hfiloloskod.hr/images/HFD/Govor/Govor-2-2018-web.pdf

Vakulenko, M.O. (2019). «Ukrainian Consonant Phones in the IPA Context with Special Reference to /v/ and /gh/». *Linguistica online* 22, 36-61. http://www.phil.muni.cz/linguistica/art/vakulenko/vak-001.pdf

Vakulenko, M. (2021). «Calculation of Phonetic Distances between Speech Sounds». *Journal of Quantitative Linguistics*, 28(3), 223-36. https://doi.org/10.1080/09296174.2019.1678709

Vakulenko, M.O. (2022). «A Tool for Scholarly Phonemic Transcription of Ukrainian Texts». *Proc. of the 2022 12th Intern. Conf. on ACIT* (26-28 September 2020). Spišska Kapitula, Slovakia: IEEE, 516-19. https://doi.org/10.1109/ACIT54803.2022.9913111

Vakulenko, M. (2023a). «Normalization of Ukrainian Letters, Numerals, and Measures for Natural Language Processing». *Digital Scholarship in the Humanities*, 38(3), 1307-21. https://doi.org/10.1093/llc/fqac090

Vakulenko, M.O. (2023b). «Unified Parametrization of Phonetic Features and Numerical Calculation of Phonetic Distances between Speech Sounds». *Journal of Quantitative Linguistics*, 30(1), 67-85. https://doi.org/10.1080/09296174.2022.2095751

Vakulenko, M. (2024). «Positional Variations of Ukrainian Back Vowel Formants». *Proceedings of the 8th International Conference on Modern Research in Social Sciences (ICMRSS)*, 1-12 (19-21 April 2024). Athens, Greece: Diamonds Science Publishing. https://doi.org/10.33422/icmrss.v1i1.301

The Expression of Past Events in South Slavic: A View from Torlak

Jelena Zivojinovic
University of Graz, Austria

Natalia Jardon
Universitat Autónoma de Barcelona, Spain

Abstract This paper investigates how past events are expressed in Torlak, a non-standardized variety spoken in southern Serbia, as opposed to the standard BCMS (Bosnian, Croatian, Montenegrin, and Serbian). Past events in Torlak are conveyed mainly through the use of the analytic perfect and the synthetic aorist and imperfect, just as in BCMS. However, unlike in BCMS where the perfect is argued to be the default marker of anteriority in lieu of the aorist, in Torlak the aorist is somewhat more productive, changing the overall distribution. Descriptively, we identify semantic and morphosyntactic differences between the aorist, the imperfect and the perfect in Torlak compared to standard BCMS. We examine the distribution and usage patterns within the Torlak paradigm, with a focus on the aorist/perfect divide: We also show how the expression of past tense in Torlak is related to the broader framework of 'Balkan' evidentiality. Overall, this paper contributes to our understanding of the mechanisms underlying tense expression in South Slavic varieties.

Keywords Torlak. BCMS. Perfect. Aorist. Imperfect. Evidentiality.

1   Introduction

This paper offers a preliminary investigation into the expression of past events in Torlak **[fig. 1]**, a non-standard South Slavic variety spoken in southern Serbia and in the bordering areas of Northern Macedonian, Bulgaria and Kosovo, with a comparative perspective on Bosnian-Croatian-Montenegrin-Serbian (BCMS).[1] Our focus lies on the interplay between analytic and synthetic strategies in marking past tense: the perfect (analytic), aorist (synthetic), and imperfect (synthetic). While BCMS exhibits a preference for the analytic perfect as the unmarked marker of anteriority (cf. Pušić 2013), independently of possible diastratic and diamesic microvariation, Torlak employs synthetic forms such as the aorist more productively in everyday speech.

In this study, we provide a descriptive account of the morphosyntactic and semantic characteristics of past-referring verb forms in Torlak, especially focusing on the use and functional distinctions between the aorist, imperfect and perfect, as compared to BCMS. A particular point of interest is the role of the aorist in Torlak, which, unlike in standard BCMS, is not obsolete or stylistically marked, but appears in a range of contexts, including colloquial narratives. We examine how this distribution may reflect deeper typological alignments of Torlak with Balkan Sprachbund features, including evidentiality-related meanings (cf. Friedman 1986; 2003; Izvorski 1997; Karagjosova 2021).

Torlak has long occupied a transitional position both linguistically and geographically, bridging Slavic and Balkan typological profiles (cf. Friedman, Joseph 2025, a.o.). In its southernmost varieties, spoken in and around Vranje and Trgovište, one finds numerous features commonly associated with the Balkan linguistic area. These include pervasive clitic doubling (attested across much of the Torlak area),[2] the presence of postposed definite articles (e.g., *knjiga-ta* 'the book'; cf. Vuković, Samardžić 2018; Azzolina et al. 2021), loss of the infinitive, and analytic future and evidential constructions. Taken together, these features support the view of Torlak as a tendentially Balkan variety, aligning it typologically with core Balkan languages (e.g., Bulgarian, Macedonian, Albanian, Romanian).

This contribution should be regarded as a first step in a broader research agenda. While we primarily focus on the distribution and functional load of past tenses here, especially the aorist and

---

[1] As Reviewer 1 pointed out, the comparative perspective provided in this article only focuses on Torlak vs. Neo-Štokavian BCMS. No Kajkavian or Čakavian varieties are included here. For simplicity, we will only refer to this variety as BCMS.

[2] See Runić 2014; Milosavljević 2017; 2018; Escher 2021a; Milosavljević, Živojinović, forthcoming; Živojinović 2021; forthcoming, a.o.

its interaction with the perfect in Torlak, we plan to extend this work by conducting a systematic, in-depth and crosslinguistic analysis of the perfect across contexts in future research. Given the theoretical complexity and diachronic instability of the perfect[3] such an investigation will be essential for a fuller understanding of the grammaticalization and semantic restructuring processes shaping tense in South Slavic.

The structure of this article is as follows. In Section 2 we provide a general overview of the expressions of the past. In Section 3, we review previous work on the past tense system in South Slavic, with particular attention to the analytic-synthetic opposition. Section 4 outlines the distribution of aorist, imperfect, and perfect forms in BCMS and Torlak. In Section 5, we focus specifically on the aorist-perfect divide in Torlak, interpreting our findings within the broader domain of evidentiality and tense in the Balkans. Section 6 concludes the paper and outlines directions for future research.

Figure 1
The overall geographical distribution of Torlak. The delimited area represents an approximation of the maximal geographical expansion of Torlak. It includes what is traditionally labeled as the Prizren-Timok dialect group, as well as the Kosovo-Resava variety. For the purpose of this study, however, the focus is exclusively on the area between Vranje and Trgovište. Base map publicly available; annotations by the authors.

---

3 See Comrie 1976; Lindstedt 2000; Iatridou et al. 2001; Plungian 2017, a.o.

2   Ways to Talk About the Past

When comparing the imperfect, the aorist, and the perfect, distinctions arise both in their structure and interpretation. Here we review, in a very concise manner, how these three forms have been defined in linguistic studies, mostly from the perspective of English, but also, in the case of the imperfect, with a reference to other languages.

The imperfect refers to a finite verb form, present in the paradigm of some languages, which expresses both past tense and imperfective aspect. Simplifying somehow, we can say that, with the imperfect, we refer to events that were ongoing or habitual in the past (e.g., Spanish *cantaban*, translated to English as 'They were singing' or 'They used to sing', depending on the context).

The aorist or simple past is a synthetic, finite verb form, that expresses past tense and perfective aspect, referring to concluded events in the past (e.g., *I ate*).

In contrast to both imperfect and aorist, perfect often takes the form of an analytic construction, involving auxiliary verbs alongside participles, (e.g., I have eaten). Semantically, the perfect conveys not only the reference to a past action, like in the case of the imperfect and the aorist, but also asserts a resultant state or relevance to the present. This distinction imbues the perfect with a deeper layer of meaning, suggesting an ongoing impact or connection between the past event and the present situation.

Generally speaking, the perfect has been (mainly) associated with three prototypical readings:[4]

  i.   The universal reading, describing events that started sometime in the past and run continuously to the present: e.g. *I have been sick since Tuesday*;
  ii.   The resultative reading, describing result states from previous events that hold at present: e.g. *I have broken my glasses*;
  iii. The experiential reading, referring to the subject's experience of some past event: e.g. *I have seen that movie (before)*.

Historically, the perfect is considered an "unstable" category (cf. Lindstedt 2000). Its evolution has been regarded as a dynamic process of grammaticalization, characterized by shifting semantic and syntactic patterns over time. Initially associated with stative possession, as in *I have my eyes closed*, it gradually extended its scope to encompass result states, as in *I have broken my glasses*. In other cases, the notion of result became less and less obvious, leaving only the trace of a state that exists by virtue of an event having taken place

---

4   cf. McCawley 1971; Comrie 1976; Parsons 1990; Iatridou et al. 2001, among others.

(what Parsons 1990 called *R-states*). An example of this is a sentence like *Anna has seen the Mona Lisa*, where there is no result of the seeing event as such, as the Mona Lisa painting remains unchanged. Describing this in the perfect, however, would amount to saying that Anna is in the state of having seen the Mona Lisa before, and this state would not be subject to change (Anna cannot "un-see" the Mona Lisa). Additionally, in French, German, and Northern varieties of Italian, among others, the perfect has come to be used as a marker of anteriority: to refer to a past event, and nothing else, just like the aorist in other languages.

The semantic variation on the use of the perfect over time revolves around two central features: one is the reference to a state of some sort, from a relation of stative possession to a result state, or a consequent R-state; the other is the inclusion of a past event, either in combination with a state (the prototypical perfect in English), or alone (acting like a past tense). All this variation maps in many cases onto a single analytic construction (such as *have* + participle in English).

Nevertheless, there are numerous attestations of languages where the perfect is associated with a synthetic form. Notably, in Portuguese, the aorist serves roles traditionally associated with the perfect in other Romance languages, as illustrated by the usage of *comi* to mean both 'I ate' and 'I have eaten.' This was also the case in Latin, before the development of the perfect in the Romance systems. In sum, across languages featuring an aorist and a perfect, there exists variability in how the semantics aligns with the morphosyntactic structures, reflecting an intricate interplay in the mapping between meaning and form.

3 The Expression of Past Events from Old Church Slavonic to BCMS

Focusing now on the South Slavic context, in Table 1 we summarize the temporal and aspectual information associated with the imperfect, the aorist and the perfect, in line with the general description above.

Table 1 Temporal and aspectual values of past-referring forms in South Slavic

Tense	Morphology	Temporal semantics	Aspectual semantics
Imperfect	Synthetic	Past	Imperfective
Aorist			Perfective
Perfect	Analytic	Present~Past	Imperfective~Perfective

A preliminary observation from Table 1 – which we will return to later in the paper – is that the perfect typically displays a more complex

functional profile than the imperfect or the aorist. This aligns with the view that the perfect is a historically unstable category, characterized by a complex and often multifunctional meaning, which in South Slavic is also reflected in its morphosyntactic complexity. Indeed, this multifunctionality is attested in the Old Church Slavonic periphrastic perfect as well, which often diverges from strict syntactic and semantic equivalence with Greek models. Plungian, Urmanchieva (2017) argue that the OCS perfect resists a unified semantic description and displays highly text-sensitive distributions. They identify three primary uses:

i. Interpretative perfect: used to evaluate or comment on a previously mentioned situation;

(1) Съказалъ еси въ людехъ силѫ твоиѭ їзбавилъ еси
    declare.PstP AUX in people strength your deliver.PstP AUX

    мышьцеѭ твоеѭ люди твоꙗ сꙑны їꙗковлꙗ їосіфовы
    arm your people your sons Jacob Joseph

'You have declared your strength among the people; you have redeemed your people with your arm, the sons of Jacob and Joseph'
(Psalm 76 (77):15; retrieved from Plungian and Urmanchieva 2017, 29)

ii. Characterizing Perfect: denotes a property or identity by appealing to past action;

(2) Благо сътворилъ еси съ рабомъ твоимъ гӏ: По словеси
    good create.PstP AUX with servant your Lord By word

    твоемоу: Благости ї наказазанию ї разоумоу
    your to_goodness and instruction and understanding

    наоучимꙗ: ꙗко заповѣдемъ твоїмъ вѣрѫ ꙗсъ
    taught.1pl for commandements your believe.1sg be.1sg.pres

'You have done good to your servant, O Lord, according to your word: to goodness, correction, and understanding we are taught; for I have believed in your commandments.'
(Psalm 118:65-6; retrieved from Plungian, Urmanchieva 2017, 26)

iii. Existential/Experiential Perfect: signals that an event occurred at least once in the past, often without specifying exact temporal anchoring;

(3) показалъ еси чюда. ї истаилъ еси ꙗко паѫчинѫ
    show.PstP AUX wonder and make.PstP AUX as spider_web

    дꙗѫ его
    Lord.dat it

'You have shown mercy. And you have hidden it like a spider's web from the Lord'
(Psalm 38:11-12; retrieved from Plungian, Urmanchieva 2017, 30)

Plungian and Urmanchieva (2017) claim that the perfect also served disambiguation functions, especially in 2nd-person singular forms of *i*-conjugation verbs, to avoid homonymy with imperatives and aorists, as in (4).

(4)

Оуничьжилъ	еси	вьсіа	остжпаіжштіа	отъ	оправъданеі
destroy.PstP	AUX	all	those_who_depart	from	justification

твоіхъ:	Ѣко	неправедьно	помышление	їхъ
your	for	unjustly	thought	their

'You have destroyed all those who turn away from your justification, for their thoughts are unjust'

(Psalm 79:5-8; retrieved from Plungian, Urmanchieva 2017, 26)

Moreover, Plungian, Urmanchieva (2017) provide attestations of auxiliary omission which they intend as pure stylistic variation, for instance чоудо сътворилъ зрацимъ '(He) has performed a miracle before the onlookers'.

From a synchronic perspective, Pušić's (2013) survey of BCMS (with particular attention to the Serbian variant) demonstrates that the perfect has become the default past tense across registers, largely replacing the synthetic aorist and imperfect. The aorist persists only in restricted contexts, i.e. speakers use it almost exclusively for telic events, especially in narrative or stylistically marked contexts,[5] and seldom (if ever) for atelic or ongoing past events. Even in spoken language the aorist can still be used for certain immediate past or modal effects, indicating it is *not* entirely moribund in the standard(ized) variety. By contrast, the imperfect is virtually absent in BCMS. Indeed, Pušić's informants produced imperfect forms only rarely and many demonstrated an inability to conjugate the imperfect at all, signaling that this tense has all but vanished from *all* registers. In short, BCMS shows a strong shift toward a one-tense (perfect) past system, with the aorist retained only as a specialized narrative device and the imperfect in full decline. Pušić (2013) attributes the overall decline of the aorist and imperfect in the standard language to their functional overlap with the perfect and the consequent loss of their distinct semantic roles, compounded by decades of non-use that have led younger speakers to lose competence in these forms.

Similarly, Mišeska Tomić (2006, 631-5) discusses the perfect in BCMS and confirms a typological tendency shared with many

---

5 It has been reported at least since Žic Fuchs's (2002) that aorist forms are widely spread across instant messaging platforms, possibly (but not exclusively) thanks to their morphological conciseness, for instance *Vidjeh te pred faxom* 'I saw you in front of the faculty building' (cf. Žic Fuchs 2002, 605 on Kajkavian). We thank an anonymous reviewer for bringing this to our attention.

European languages: the analytic perfect is replacing the older synthetic past tenses (aorist and imperfect). In contemporary use, it functions as a general past tense, covering a wide range of contexts. Specifically, it is used to express (a) past events with clearly defined temporal boundaries, (b) past events that maintain relevance for the present (e.g., resultative or experiential readings) and (c) events of which the speaker lacks direct knowledge – an evidential nuance emergng particularly in the southeastern dialects (Torlak).

4 Bcms as Compared to Torlak: The Empirical Description

4.1 BCMS

In BCMS, past events are predominantly expressed through an analytic tense, descriptively known as *perfect tense* (cf. Pušić 2013). Structurally, this perfect consists of a finite auxiliary, the verb *biti* 'to be' which agrees with the subject, and an active participle, as in (5).

(5) *Prošle   zime   **je   bilo***   *jako   hladno*
    last   winter   aux   be.PstP   very   cold
    'It was really cold last winter'

Semantically, the perfect in BCMS denotes general anteriority, just as the past tense in English (e.g. *I ate*) and the perfect tense in Northern regional Italian (e.g. *Ho mangiato* 'I have eaten'), French (e.g. *J'ai mangé* 'I have eaten'), and Standard German (e.g. *Ich habe gegessen* 'I have eaten'). This tense structure encapsulates the notion of past events occurring prior to the time of speech. It covers a wide range of perfect values, such as the resultative or experiential perfect (cf. Stojanović 2017, 2018), but does not cover the so-called Universal Perfect where the described eventuality starts in the past and terminates at the reference time[6] (expressed by the present, e.g. *Živim u Beogradu 10 godina* 'I have lived in Belgrade for 10 years). In addition to the perfect, BCMS features two synthetic past tenses, the aorist and the imperfect, both of very limited use. The aorist consists of a finite verb form (6), and denotes an event that took place shortly before the time of speech, expressing perfective aspect (cf. Pušić 2013).

---

6   See Grønn, von Stechow 2020 for an overview.

(6) ***Rekoh*** *ja* *da* *deca* *ne* *treba* *da* *skaču* *u* *bazen* *posle* *ručka*
    say.AOR I that kids not need to jump in pool after lunch
    'I have said that kids shouldn't jump into the pool after lunch'

Although the aorist is rarely used in contemporary spoken language and is often considered as 'bookish' (cf. Jung 2025), it finds its place in literary works and exclamatory statements aimed at defining witnessed events occurring just before the speech time, for instance *Pade Sava!* 'Sava fell!', intended as the reporting of the witnessing of Sava falling shortly prior to uttering it. Such expressive meanings are exclusively attested for the VP-fronted aorist, as suggested by Halupka-Rešetar and Todorović (2015), which denotes a completed, factual event, making it compatible with exclamative factivity, as in *Pade Sava!* (lit. 'fell Sava') and (6). Arsenijević (2013) and Todorović (2016) further argue that BCMS aorists are purely aspectual, rather than temporal forms, expressing no inherent past interpretation.

Note that, as Reviewer 1 points out, the BCMS aorist is attested in contexts which are not witnessed *stricto sensu*, but are interpreted counterfactually, as in (7).[7]

(7) *Ako* *ne* *budemo* *odlučni*, ***propadoše*** *nam* *svi* *planovi*
    if not be decisive fall_through.AOR to_us all plans
    'If we are not decisive, all our plans will fall through'
    (cf. Riđanović 2012, 317; retrieved from Todorović 2016)

On the other hand, the imperfect is the imperfective counterpart of the aorist, as in (8), denoting an ongoing or continuous event. Despite being reported in descriptive grammars, the imperfect has fallen out of use in contemporary spoken language and is primarily attested in literary texts.[8]

(8) *Celu* *noć* ***padaše*** *kiša*
    entire night fall.IMP rain
    'It was raining all night'

---

7 This use of the aorist in hypothetical context is not the default reading for the BCMS aorist. However, there is a parallel in Romance, where the Latin pluperfect *amaveram*, expressing a past-oriented, temporal meaning, shifted into a subjunctive form expressing a hypothetical or non-factual meaning in languages like Spanish. Nevertheless, in certain varieties, *amaveram* continued to express the pluperfect, temporal meanings that it had in Latin (Jardón 2016), and its use as a subjunctive is reported but rare. These meanings are expressed through *amase*, from Lat. *amavissem*.

8 Reviewer 1, once again, points out that the only apparent productive imperfective form still currently used in spoken Štokavian is *be(ja)še* from *biti* 'to be': *Kako be(ja)še?* 'What was it again?'.

4.2   Torlak

Interestingly, the aorist is still a productive category in Torlak, in contrast to its very limited use in BCMS.[9] In Torlak, like in Portuguese, the aorist seems to be able to express anteriority (9), but also perfect meanings where a past event is linked to a present state of affairs (10).

(9)   *Jučer   gi   **obra***
      yesterday   them   pick.AOR
      'I picked them yesterday (the flowers)'

(10) *Ja   li   te   **probudi**?*
      I   q   you   wake_up.AOR
      'Have I woken you up?'

The imperfect in Torlak, on the other hand, marks imperfective aspect, leading to continuous (11) and habitual (12) readings.

(11) *Koji   su   oni   što   **svireševu** kad[t] **bemo**   u   grad[t]?*
      which   are   they   that   play.IMP   when   be.IMP   in   town
      'Who were those [people] who played/were playing when we were in town?'

(12) *Stanča   **čitaješe**   po   50   strane   svak   d'n.*
      Stanča   read.IMP   around   50   pages   every   day
      'Stanča used to read 50 pages per day'

In addition, the perfect is also attested in Torlak, as in (13-14). The auxiliary is generally omitted,[10] except in negated contexts (15).

(13) *Deda   **(je)**   **legnaja**   da   spava*
      granpa   aux   lie_down.PstP   to   sleep
      'Granpa has gone to sleep'

(14) *Vodu   gu   pa   **puštija***
      water   her   again   open.PstP
      'He has opened the water again'

---

9   For a broader overview of Torlak, its internal variation and contexualization within the Balkan area see Friedman 2007; Vuković 2021;Vuković et al. 2022; Mirić, Arsenijević 2024; Friedman, Joseph 2025; Milosavljević, Živojinović, forthcoming; Živojinović, forthcoming, a.m.o.

10   Also attested in the Timok variety (see Escher 2021b).

(15) *Doktorka   ne   me   je   **pitala**   za   tebe*
      doctor   not   me   aux   ask.PꜱᴛP   for   you
      'The doctor did not ask me about you'

A preliminary overview of the distributions in BCMS and Torlak is given in Table 2. As this summary reveals, there is very little descriptive information on the specific semantic contribution of the aorist and the perfect in Torlak. In what follows, we will develop a hypothesis about the semantic contribution of these two forms, based on evidentiality.

Table 2 The distribution of past event-reporting forms in BCMS and in Torlak (preliminary)

Tense	Morphology	Semantics BCMS	Semantics Torlak
Imperfect	Synthetic	Ongoing or habitual events in the past	
Aorist		Witnessed events occurring just before the speech time	?
Perfect	Analytic	General anteriority	?

5   Past Tense and Evidentiality

We began this work by defining Torlak as a tendentially Balkan variety. In this section, we argue that there are reasons to believe that, when it comes to meaning, the division of labour between the aorist and the perfect in Torlak can be better understood and systematized in the general context of 'Balkan' evidentiality (Friedman 1986, 2003; Karagjosova 2021; Izvorski 1997; among others).

Evidentiality in Balkan languages has been extensively studied (Friedman 1986, 2003; Izvorski 1997; Karagjosova 2001; 2021; Simeonova, Zareikar 2015, Simeonova 2020), with a consensus that the perfect often functions as a locus for evidential marking. In Bulgarian, for example, the present perfect can encode indirect evidentiality, such as inference or hearsay (Karagjosova 2001), while the aorist is generally reserved for expressing direct or witnessed events. Izvorski (1997) explicitly refers to the Bulgarian and Macedonian perfects as 'evidential perfects,' which can convey inference or hearsay depending on the context and the presence of the auxiliary verb.

Simeonova, Zareikar (2015) further explore this relationship, suggesting that the perfect in Bulgarian is closely tied to indirect evidential meanings and cannot be used with direct evidentiality. In their analysis, evidential marking is parasitic on the

temporal-aspectual-modal (TAM) system, meaning that the perfect tense, which encodes indirect evidentiality, cannot simultaneously convey direct evidence, which would require a different syntactic structure or verbal form. This observation extends naturally to Torlak, where a similar distinction between the perfect and aorist tense systems is found.

Torlak exhibits a comparable asymmetry. For example, consider the data in (16),[11] where the analytic perfect is infelicitous unless it is embedded in a reportative or memory-impaired context:

(16) *??Jučer* **sam** **pročitaja** *Rat* *i* *mir.*
      yesterday aux read.PstP War and peace
      'Yesterday I read War and Peace'

The perfect in (16) is infelicitous unless it is used in a specific context of memory loss, with the utterance being reported. This constraint mirrors what Simeonova describes for Bulgarian: the perfect resists direct evidential uses. A felicitous version of this sentence in Torlak would instead have the aorist, which, consistent with Friedman (2003), encodes direct, witnessed events. The distribution thus aligns with the Balkan-wide tendency to mark evidential distinctions through past tense morphology.

This evidential split becomes clearer in examples (17a-c):

(17) a. **Rodila** **sam** *ćerku* *1968e* non-evidential
         give-birth.PstP aux daughter 1968
         'I gave birth to my daughter in 1968'
     b. *Baba* **(je)** **rodila** *ćerku* *1968e* inferential evidence
         grandma aux give-birth.PstP daughter 1968
         'Grandma gave birth to a daughter in 1968'
     c. **Baba* **rodi** *ćerku* *1968e* direct evidence
         grandma give-birth.AOR daughter 1968
         'Grandma gave birth to a daughter in 1968'

(17a) is fully felicitous and does not imply indirect evidence, inference, or renarration. Rather, it simply reports a past event without highlighting the source of information. This type of perfect aligns with what Izvorski (1997) terms the neutral perfect, where the speaker makes no epistemic commitment to whether the information was personally witnessed or indirectly obtained. Such

---

11   The judgments for the examples discussed in this section are based on the intuitions of Author 1 who is a heritage speaker of Torlak.

uses are analogous to statements like *World War II began in 1939* – grammatically perfect, and evidentially neutral.[12]

In (17b), the optional auxiliary plays a key role in evidential interpretation. Following Karagjosova's (2001) analysis of Bulgarian, the absence of the auxiliary in third-person contexts pragmatically signals that the speaker did not witness the event but inferred it, marking an indirect evidential use of the perfect. When the auxiliary is overtly realized, by contrast, the evidential reading is not obligatory: in Torlak, third-person perfects may allow either a neutral (non-evidential) or an indirect evidential interpretation depending on context. This raises the possibility of a person split in evidential usage, where first- and second-person forms remain largely neutral, while third-person forms oscillate between neutral and evidential readings. However, this pattern requires further investigation; at present, the only robust clue is that auxiliary drop in third person systematically signals inferred evidence. This distribution nonetheless aligns with Izvorski's (1997) claim that evidential perfects lack temporal anchoring and rely on indirect epistemic authority.

Example (17c) is ungrammatical in most contexts because the aorist requires direct, perceptual evidence. Since the speaker could not have witnessed the grandmother's childbirth, the use of the aorist misfires. This supports the proposal that, in Torlak, the aorist encodes direct evidentiality – what Friedman (2003) terms "witnessed" – and that it is semantically incompatible with inferential or hearsay contexts.

Summing up, the distribution of the aorist and the perfect tenses in Torlak, as well as the possibility of omitting the auxiliary in the perfect in this variety, seem to both be related to evidentiality, as it has been described for other languages in the Balkan context. The main findings for Torlak complete the chart from Table 2, as shown below.

Table 3   The distribution of past event-reporting forms in BCMS and in Torlak

Tense	Morphology	Semantics BCMS	Semantics Torlak
Imperfect		Ongoing or habitual events in the past	
Aorist	Synthetic	Witnessed events occurring just before the speech time	Witnessed events occurring any time in the past

---

12   As a reviewer points out, in Izvorski's argument it is hard to keep epistemic modality and evidentiality separate, as they seem to blend in the discussion. We approach "neutrality" here from an evidential standpoint, that is, whether the speaker can rely on any circumstantial evidence supporting the evaluated proposition.

Perfect	Analytic	General anteriority	Inferred events in the past (auxiliary may be omitted in 3SG)
			Non-evidential events in the past (auxiliary can't be omitted)

6   Concluding Remarks

In this paper, we have provided a systematic overview of the factors that condition the distribution and use of past-referring tenses in Torlak, a tendentially Balkan variety of South Slavic spoken primarily in Southeastern Serbia.

While there is not much variation in the licensing of the imperfect past in Torlak with respect to BCMS, there are important differences in the conditions of use of the other two tenses: the perfect and the aorist. In BCMS, the perfect has come to be the default expression of general anteriority, and the aorist is restricted to literary contexts and specific sentences describing events directly witnessed by the speaker shortly before the utterance time, like *Pade Sava!* 'Sava fell (just now)!'. By contrast, in Torlak, the aorist is not so restricted: it can refer to any event in the recent or distant past. What determines the choice of tense in Torlak is not temporality (relative distance between the time of the event described and the utterance time) or pragmatics (a sense of relevance to the present). Rather, the distribution of aorist vs. perfect is conditioned by the type of evidentiality reported in the sentence (if any): the aorist is used when the speaker has witnessed the event himself (direct evidentiality); otherwise, Torlak resorts to the perfect. As it stands, the evidentiality in Torlak seems to be built upon 'temporality', i.e. abstract deictic schemata specifying the relative position of the event participants with respect to the spacio-temporal coordinates of the event itself (cf. Pancheva, Zubizarreta 2019 for a comparison with Paraguayan Guaraní). This view of evidentiality is compatible with formal accounts such as Migdalski (2023) or Todorović (2016), who argues against the existence of a Tense projection in South Slavic. What follows is that the involved temporality in the computation of evidentiality might be more primitive than 'tense'.

Furthermore, the perfect form itself is sensitive to whether there is still an indirect source of report (inference), and in those cases the auxiliary may be omitted, or whether evidentiality is gone altogether, in which case the speaker has to express the perfect in its full form (auxiliary + participle).

Our findings are in line with the general claims made in previous literature on Balkan evidentiality, with a systematic association

between direct evidentiality and the aorist. The link between evidential meaning and auxiliary omission, however, has only been briefly touched upon in the current study and deserves more attention. Overall, our study has examined morphosyntactic and semantic variation within South Slavic, providing a basis for more formal analyses of a range of linguistic phenomena.

Abbreviations

aor	aorist
aux	auxiliary
imp	imperfect
pres	present
PstP	past participle
q	question particle (clitic)

Acknowledgements

Bibliography

Arsenijević, B. (2013). "Vremensko i aspektualno značenje aorista (The tense and aspect components in the semantics of the /Serbian/aorist)". *Srpski jezik,* 18, 253-61.

Azzolina, B.; Girolami, V.; Živojinović, J. (2021). "Postponed Articles and DP Structures in Torlak". *Balcania et Slavia*, 1(1), 97-116.

Comrie, B. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*, vol. 2. Cambridge: Cambridge University Press.

Escher, A. (2021a). "Double Argument Marking in Timok Dialect Texts (in Balkan Slavic Context)". *Zeitschrift für Slawistik*, 66(1), 61-90. `https://doi.org/10.1515/slaw-2021-0004`

Escher, A. (2021b). "Auxiliary Omission in the Perfect Tense in Timok". *Balkanistica*, 34, 41-64.

Friedman, V.A. (1986). "Evidentiality in the Balkans: Bulgarian, Macedonian, and Albanian". Chafe, W.; Nichols, J. (eds), *Evidentiality: The Linguistic Coding of Epistemology*, 168-87. Norwood, NJ: Ablex.

Friedman, V. (2003). "Evidentiality in the Balkans with special attention to Macedonian and Albanian". Aikhenvald A.; Dixon, R.M.W. (eds), *Studies in evidentiality*, 189-218. Amsterdam: John Benjamins.

Friedman, V.A. (2007). "Balkanizing the Balkan Sprachbund". Aikhenvald, A.; Dixon, R.M.W. (eds), *Grammars in Contact: A Cross-Linguistic Typology*, 201-19. Oxford: OUP.

Friedman, V. (2011). "The Balkan Languages and Balkan Linguistics". *Annual Review of Anthropology 40*, 275-91.

Friedman, V.; Joseph, B.D. (2025). *The Balkan languages*. Cambridge: CUP.

Grønn, A.; von Stechow, A. (2020). "The Perfect". Gutzmann, D.; Matthewson, L.; Meier, C.; Rullmann, H.; Zimmerman, T.E. (eds), *The Wiley Blackwell Companion to Semantics*. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118788516.sem046

Halupka-Rešetar, S., Todorović, N. (2015). "Serbian Verbs Front to be Expressive". *Grazer Linguistische Studien*, 83, 139-48.

Iatridou, S.; Anagnostopoulou, E.; Izvorski, R. (2001). "Observations About the Form and Meaning of the Perfect". *Current Studies in Linguistics Series*, 36, 189-238.

Izvorski, R. (1997). "The Present Perfect as an Epistemic Modal". Lawson, A.; Cho, E. (eds), *Proceedings of SALT VII*, 222-39. Ithaca, NY: Cornell University.

Jardón, N. (2016). "The Distribution and Licensing of -*ra* in Eonavian Spanish: A Nanosyntactic Analysis". *Isogloss*, 2(1), 21-41. https://doi.org/10.5565/rev/isogloss.32

Joseph, B.D. (2020). "Language Contact in the Balkans". Hickey, R. (ed.), *The Handbook of Language Contact*, 618-33. Hoboken, NJ: Wiley.

Jung, H. (2025). "What Happens When Person Marking Is Lost? Consequences of the Loss of Synthetic Past Tenses in Old Russian". Keynote presentation at FDSL 18 (Formal Descriptions of Slavic Languages) on 22 September 2025 in Wrocław, Poland.

Karagjosova, E. (2021). "Mirativity and the Bulgarian Evidential System". Blümel, A.; Gajić, J.; Geist, Lj.; Junghanns, U.; Pitsch, H. (eds), *Advances in Formal Slavic Linguistics 2018*, 133-67. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.5483102

Lindstedt, J. (2000). "The Perfect-Aspectual, Temporal and Evidential". *Empirical Approaches to Language Typology,* (6), 365-84.

McCawley, J.D. (1971). "Tense and Time Reference in English". Fillmore C.; Langendoen, T. (eds), *Studies in Linguistic Semantics.* New York: Holt, Rinehart, Winston.

Migdalski, K. (2023). "A Case of the Non-Directionality of Language Change: The Strength of Auxiliaries and Pronouns in South and West Slavic". *Slavia Meridionalis, 23.* https://doi.org/10.11649/sm.3147

Milosavljević, S. (2017). "Raspoređivanje enklitika u govoru jablaničkog kraja". *Svet reči*, 43/44, 64-70.

Milosavljević, S. (2018). "Osobine klitičkog udvajanja u govoru jablaničkog kraja". Kovačević, M. (ed.), *Savremena proučavanja jezika i književnosti: zbornik radova sa IX naučnog skupa mladofilologa Srbije, održanog 8. aprila 2017. Godine na Filološko-umetničkom fakultetu u Kragujevcu / knjiga 1*, 41-52. Kragujevac: Filološko-umetnički fakultet u Kragujevcu.

Milosavljević, S.; Živojinović, J. (forthcoming). "The Continuum of Pronominal Clitic Systems in Torlak as a Window in the Architecture of Clitics". To be published in the JSL Special Issue for the 25th anniversary of Franks & King (2000).

Mirić, M.; Arsenijević, B. (2024). "Complementizer Drop and Clitic Climbing in Torlakian: Evidence from the Timok Variety". *Balcania et Slavia*, 4(2), 175-200.

Pancheva, R.; Zubizarreta, M.L. (2019). "On the Role of Person Features in the Evidential-Temporal Connection". *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 64(4), 673-708.

Parsons, T. (1990). *Events in the Semantics of English: A Study in Subatomic Semantics.* Cambridge, MA: The MIT Press.

Plungian, V.A.; Urmanchieva, A. Yu. (2017). "Перфект в старославянском: был ли он результативным?" (The Perfect in Old Church Slavonic: Was it Resultative?). *Slovene*, 6(2), 13-56.

Pušić, D. (2013). "Past in the Serbian Variant of Serbo-Croatian: Distribution of Verbal Tenses in Various Registers". *Scando-Slavica*, 59(1), 108-37. `https://doi.org/10.1080/00806765.2013.800728`

Riđanović, M. (2012). *Bosnian for Foreigners – With a Comprehensive Grammar*. Rabic: Sarajevo.

Runić, J. (2014). *A New Look at Clitics, Clitic Doubling, and Argument Ellipsis: Evidence from Slavic* [PhD Dissertation]. Storrs: University of Connecticut.

Simeonova, V.; Zareikar G. (2015). "The Syntax of Evidentials in Azeri, Bulgarian and Persian". Vinerte, S. (ed.), *Proceedings of the Annual Meeting of the Canadian Linguistic Association.* Toronto: Canadian Linguistic Association.

Simeonova, V. (2020). *The Syntax and Semantics of Light Attitudes* [PhD Dissertation]. Ottawa: University of Ottawa.

Stojanović, S.B. (2017). *Funkcionalno-semantičko polje perfektnosti u ruskom i srpskom jeziku* [Doktorska disertacija]. Beograd: Univerzitet u Beogradu. `https://nardus.mpn.gov.rs/handle/123456789/9032?show=full`

Stojanović, S. (2018). "On Basic Meaning of the Active Perfect in Serbian and its Typological Classification". *Južnoslovenski filolog*, 74(2), 109-38.

Todorović, N. (2016). *On the Presence/Absence of TP: Syntactic Properties and Temporal Interpretation* [Phd Dissertation]. Storrs: University of Connecticut.

Tomić, O.M. (2003). *Balkan Syntax and Semantics.* Amsterdam: John Benjamins.

Tomić, O.M. (2006). *Balkan Sprachbund Morpho-Syntactic Features*. Berlin: Springer.

Vuković, T. (2021). "Representing Variation in a Spoken Corpus of an Endangered Dialect: The Case of Torlak". *Language Resources & Evaluation*, 55, 731-56. `https://doi.org/10.1007/s10579–020–09522–4`

Vuković, T.; Samardžić, T. (2018). "Prostorna raspodela frekvencije postpozitivnog člana u timočkom govoru". Ćirković, S. (ed.), *Timok: folkloristička i lingvistička terenska istraživanja 2015–2017*. Knjaževac: Narodna biblioteka "Njegoš".

Vuković, T.; Escher, A.; Sonnenhauser, B. (2022). "Degrees of Non-Standardness. Feature-Based Analysis of Variation in a Torlak Dialect Corpus". *International Journal of Corpus Linguistics*, 27(2), 220-47.

Žic Fuchs, M. (2002). "Communication Technologies and Their Influence on Language: An Example from Croatian". *Studia Romanica et Anglica Zagrabiensia*, 47-8, 597-608.

Živojinović, J. (2021). "Torlak Clitic Doubling: A Cross-Linguistic Comparison". Bluemel, A.; Gajić, J.; Geist, Lj.; Junghanns, U.; Pitsch, H. (eds), *Advances in formal Slavic linguistics 2018*, 423-41. Berlin: Language Science Press.

Živojinović, J. (forthcoming). "Tracing Language Contact: Parameter Hierarchies and Torlak Clitic Doubling". To be published in *Ricerche Slavistiche*.

What's in a Name (Again)? About Judeo-Spanish and Sephardic *Linguistic Culture*

Ana Stulic
Bordeaux Montaigne University, France

Abstract   This article examines the Judeo-Spanish language and the *linguistic culture* of its speakers. It puts forward three ideas that have shaped language use and multilingual practices among Sephardic Jews in the Balkans. The first concerns the centrality of Hebrew, the language of the Torah, which ceased to be spoken around 400 CE yet continued to hold supreme ontological and theological authority as the 'Holy Tongue'. The second, rooted in the experience of exile from the Iberian Peninsula, highlights the special status accorded to Sephardic Spanish as a specifically 'Jewish' language. The third points to the strong sense of loyalty to the spoken vernacular that emerged among Judeo-Spanish speakers in the late nineteenth century. The discussion of these three ideas provides a framework for understanding the range of terms employed by both present-day speakers and scholars to designate this language: Ladino, Judeo-Spanish, and Judezmo.

Keywords   Sephardic Jews. Linguistic culture. Ladino. Judeo-Spanish. Espanyol. Judezmo.

1 Introduction

Judeo-Spanish[1] is a conventional name for a language spoken by the descendants of Sephardic Jews who were expelled from Spain in 1492, the majority of whom settled in what was then the territory of the Ottoman Empire. The exiles, from different regions of the Iberian Peninsula, spoke different Iberian Romance varieties. Judeo-Spanish emerged as the result of a process of *koineization* in the new settlement (Minervini 2002). The core properties of this language are those of Castilian, but it also incorporated elements from Aragonese, Leonese, Portuguese, and Catalan. After the expulsion, the Sephardic Jews came into contact with numerous other languages.

The mass deportation and killings of Jews during World War II led to the disappearance of the Sephardic population in many localities in Greece, Romania, and ex-Yugoslavia (now Bosnia and Herzegovina, North Macedonia, and Serbia). The majority of Sephardic Jews in Bulgaria and Turkey survived WWII, but they faced strong antisemitism, and many subsequently emigrated to Israel, the USA, and other countries. Therefore, contemporary Judeo-Spanish is mostly 'displaced' from the Balkans, where it had been spoken for centuries, and is currently considered a "severely endangered" language (Mosely 2010, 24-5).

Although obviously a Balkan language, in the sense that it has evolved in the Balkan area, Judeo-Spanish has been studied in Balkan Linguistics only in a very marginal way. Recently, Friedman and Joseph (2014) brought this language – which, following Bunis (2018), they call *Judezmo* – into the focus of Balkan Linguistics and tried to systematically relate possible Balkanisms to the available Judeo-Spanish data (Friedman, Joseph, forthcoming). The same is attempted in Adamou and Sobolev (forthcoming).

Judeo-Spanish has a long tradition of literacy and is rather well documented. The correct interpretation of Judeo-Spanish sources – almost exclusively in Hebrew script until the beginning of the twentieth century, when Latin, Cyrillic or Greek script (depending on the place) entered into use – requires knowledge of other languages, as well as an understanding of the Jewish tradition and the specificities of Sephardic culture. Moreover, for understanding the Sephardic speaker's perspective (cf. Weinreich [1953] 1966, 72-4), we need to understand her/his *linguistic culture*, a notion defined as follows:

> the sum totality of ideas, values, beliefs, attitudes, prejudices, myths, religious structures, and all the other cultural 'baggage'

---

that speakers bring to their dealings with language from their culture. (Schiffman 1996, 5)

In situations of language contact, a multilingual speaker can choose – or refuse – to use a form that belongs to a repertoire they identify as another language (Matras 2009, 336, 341) because this form produces a certain effect in the utterance (cf. the *poetic* function of language, Jakobson 1960). Doing that can be conscious or unconscious, but these choices are not arbitrary. They are guided (without being determined) by the specific goals the speaker has in a concrete communication situation, by her/his cognitive limitations, as well as by the cultural representations s/he has about the language(s) s/he speaks. The representations that constitute the speakers' *linguistic culture* are transmitted along with the language. They can change over time and, as such, are also historical phenomena.

In this paper, I put forward three ideas held by Sephardic Jews in the Balkans that have had an impact on their language use and multilingual practices. The first idea is what will be called the 'Holy Tongue myth'. As it is a part of rabbinic Judaism, it precedes chronologically the emergence of Judeo-Spanish and is shared by other Jewish groups. The second one, deeply rooted in the experience of the exile from the Iberian Peninsula, is the special status given to Sephardic Spanish as a 'Jewish' language. The third idea involves the expression of loyalty towards the spoken language that emerged among Judeo-Spanish speakers in the late nineteenth century. It is one of the central components of a Sephardic cultural and political movement called *Sephardism*. The discussion of these three ideas will allow me to contextualize the different names given to this language in the academic literature today. My argumentation is based on the analysis of primary sources and on the reinterpretation of materials discussed by Şaul (2013; 2023) and Bunis (2011), who both thematized – adopting divergent points of view – the question of native designations of the language spoken by Sephardic Jews.

2 The Holy Tongue Myth and Ensuing Multilingual Practices

The 'Holy Tongue' (Hebr. *lashon hakodesh*) is a Jewish concept which consists in giving primacy to the language(s) of the sacred texts over vernacular languages. Hebrew, the language of the Torah, ceased to be a spoken by 400 CE, but the transmission of sacred texts and the language in which these texts were written continued to be the basis of Jewish education at all times. Therefore, in Jewish tradition, the Holy Tongue (which, besides Hebrew, also embraces Aramaic) is considered to be ontologically and theologically the most important

language. The vernacular language is called *la'az*, which in Hebrew means 'foreign language'. In this multilingual setting, even if not all speakers had a good command of Hebrew, this language was perceived as central and the spoken language was dependent on it in many ways. The borrowings from Hebrew could be and often were renewed. Hebrew was also the source of the scripts used for vernacular languages: their form and phonetic values were derived from the Hebrew writing system.

For didactic and liturgical reasons, Jews on the Iberian Peninsula developed a practice of translating Hebrew sacred texts into the vernacular Romance language already in the period prior to expulsion. As we can see in the first printed translations of biblical texts among the Sephardic Jews (sixteenth cent.), these translations kept traces of the Holy Tongue by preserving many features of the Hebrew language (syntax, word order, choice of lexical items according to phonetic similarity with Hebrew words, etc., see Sephiha 1973).

As Şaul pointed out (2013, 183), this hierarchical relation between two languages can be qualified as a case of *hieroglossia*, a term coined by Jean-Noël Robert and defined as follows:

> the sum of relations that develop between a language perceived as a central or founding element in a given culture area (this language being the *hierogloss*) and a language or languages that are perceived as being dependent, not historically or linguistically, but ontologically or theologically, on that hierogloss. Within a hieroglossic relationship, the language perceived as dependent, often called the "vulgar tongue" or "vernacular" [...] is clearly considered not to be self-sufficient. (Robert 2006, 26)

In the history of Jewish tradition, the status of Hebrew as hierogloss was stable and long-lasting, while spoken languages were subject to language shifts.

Putting and keeping Hebrew in the focus of education and cultural transmission as a hierogloss had important practical consequences in overall language use in the Jewish tradition. It led to the development of group-internal multilingual practices – more particularly, language learning, translation, and glossing practices – all directed, in the first place, to the understanding and transmission of sacred texts. What is significant for the history of Judeo-Spanish and its contact with other languages in the Balkans is that these multilingual practices were extended to these other languages without the limits imposed by the sacredness of the *hierogloss*. Let us give an example. In one of the sixteenth-century Sephardic sources, we can observe glosses used for introducing new words from Turkish. Moshe Almosnino, an erudite Salonica's rabbi wrote a text known today as *The Chronicles of the Ottoman Sultans* (c. 1567; see ed. by Romeu Ferré 1998). In this

text, I counted 124 lexical items of Turkish origin. More than half of these (78) are followed or introduced by a metalinguistic comment containing translation or explanation:[2]

> la escopeta, que ellos llaman **cundac** [...] (Romeu Ferré 1998, 95)
> 'gun, which they call **kundak'**

> cocinas, que ellos llaman **mutbac** [...] (Romeu Ferré 1998, 184)
> 'kitchen, which they call **mutbak'**

Progressively, many of these words entered everyday speech. Sephiha reports that 15% of entries of the Néhama's dictionary (1977) is of Turkish origin (Sephiha 2003, 68).

The etymology of one of the names currently used for the language of Sephardic Jews, *Ladino*, is directly related to the Holy Tongue myth. Derived from medieval *ladino* (>lat. *latīnus*), which referred to Latin or Castilian, in opposition to Arabic or another language (Révah 1970; Şaul 2013, 186), the noun *ladino* (mainly used in plural) was used to denote glosses for sacred texts in Hebrew and the verb *ladinar* meant 'translate from Hebrew to Spanish'. In the history of Judeo-Spanish, the word *ladino* shows semantic change, but two semantic components are constant: it designates texts in a language other than Hebrew and it refers specifically to written language (Şaul 2013, 191-5). By metonymy, it was used for liturgical text based on translations from Hebrew (although read orally, it was written text), for written text in the vernacular language and for printed letters of so-called Rashi type (*letras ladinas*). Later, the verb *ladinar* was also extended to simply mean 'to translate' as in *ladinadas en grego* 'translated into Greek' (attested in Salonica, late nineteenth century; Şaul 2013, 187).

The term *Ladino* started to denote the spoken language only in the twentieth century. It gained popularity among speakers in the post-Shoah period, especially in the USA and Israel, from where this new meaning was exported to the Balkans. One of the motivations underlying the use of the word *Ladino* for both written and spoken language was probably the intention not to separate the spoken language from the written language of rabbinical tradition (Şaul 2013, 205). However, present-day speakers familiar with the historical meaning of the word may feel reluctant to use it, because for them, *Ladino* cannot refer to spoken language.

---

2   The complete list of these metalinguistic comments can be found in Romeu Ferré (1989, 95-100).

3   *Espanyol* Meaning Jewish

What did the language of the Sephardic exiles look like? It is commonly accepted that, before the expulsion, Jews in the Iberian Peninsula spoke more or less the same language variety as their Christian neighbors (Minervini 2006, 18). The differences we know about are lexical (see Minervini 2006). Among the Sephardic exiles in the Ottoman empire, native speakers of Castilian were probably a minority, but at the time of the expulsion, this language enjoyed the status of a translocal prestigious language and many Jews from other Iberian regions probably had some command of it (Révah 1965). Şaul suggests that the emergence of the Judeo-Spanish *koine* was a result of a deliberate language shift towards Castilian instead of a spontaneous, slow-paced process of leveling of different language varieties. Castilian was chosen for use in writing and community affairs and there is evidence that the children of some of the Iberian Jewish exiles were intentionally trained to speak Castilian (for a detailed discussion, see Şaul 2023). This language shift, as Şaul underlines, may appear difficult to conceptualize because it is atypical in our modern world, where language shift generally affects speakers of a minority language abandoning it to adopt the official language or the language of the politically dominant majority. Şaul comes to the conclusion that it is the experience of language shift and the "resulting sense of constructedness" that gave rise to the propensity to adopt new linguistic forms, which he (rightly, I think) characterizes as an enduring feature of Sephardic linguistic culture (Şaul 2023, 882). In my opinion, both phenomena – the shift to Castilian and the propensity to adopt new linguistic forms – are favored by the belief that vernacular language is of secondary importance. The shift to Castilian was not mirrored in the name of the language: early sources use *la'az* ('foreign') or *romanse* for any Romance variety, while *castellano* ('Castilian'), common at the time of the expulsion, is absent from Sephardic sources.

It may seem surprising that the most common denomination for the vernacular language of Sephardic Jews through history is *Espanyol* ('Spanish')[3] (Şaul 2013, 211). One may ask why this name was not abandoned by the Jews, given the injustice they must have felt regarding the expulsion. The short answer would be that it was because *Espanyol* came to mean 'Jewish'.

The term *Espanyol* appears for the first time in Sephardic sources in the sixteenth century. This is also a period when this word became

---

3 Friedman also reports having heard "Spanyol" during his field work in what is now North Macedonia (Friedman, Joseph, forthcoming, 37, footnote 69). The variants "Eshpanyol" or "Shpanyol" can also be heard.

popular both in Spain and outside of it (Alonso 1968, 20-30) and its use was probably connected with the role the language it denotes played in the rejudaization of baptized Jews, who left the Iberian Peninsula long after the 1492 expulsion. Most of the baptized Jews coming back to Judaism were native speakers of Portuguese and were familiar with Castilian/Spanish, since it enjoyed great prestige in Portugal from the fifteenth to the eighteenth centuries and was used as a second language among writers (Roth 2025). The designation *Espanyol* appears in texts used for didactic purposes in Jewish instruction such as the Ferrara Bible prologue (1553) and *Hesheq Shelomo*, a biblical glossary (Venice, 1588) (Şaul 2013, 211). We can assume that the name *Espanyol* was reinforced, if not downright introduced, through contacts with these Portuguese-speaking baptized Jews, who joined existing Sephardic centers in the Ottoman empire, Italy, and Western Europe. In Western Europe, especially in Amsterdam, the baptized Jews learnt about Judaism through Bible translations and liturgical and theological texts written in Spanish. The need for instruction in basic Jewish observance led the emigres from the Iberian Peninsula to address themselves to trained rabbis from established Sephardic centers such as Venice, Livorno, and Salonica (Bodian 1997, 148). The appointment of Joseph Pardo, a rabbi born in Salonica, to Bet Jacob's congregation in Amsterdam, where he served from 1608 to 1619, is a good example of East-West contacts through Sephardic networks (Bodian 1997, 92, 95-6, 103, 165).

In the first centuries after the expulsion, both in the West and in the East, several factors may have contributed to reinforcing the identification of *Espanyol* as a Jewish language:

a. Spanish was a language of Jewish instruction for baptized Jews;
b. in the countries of their new settlement, Spanish was mostly spoken by Jews, and it was seen as 'Jewish' both by the speakers themselves and outsiders;
c. Spain was seen as a place where many baptized Jews still lived and to whom Sephardic Jews living outside the Iberian Peninsula were related by kinship.[4]

A good illustration of this is the attested utterance that can still be heard in Turkey "*Djudyo es? Avla espanyol?*" ('Is he Jewish? Does he speak Spanish?', Şaul 2013, 228).

---

[4] For Western European Sephardic communities, Bodian stresses the importance of kin-based networks in trade and community matters (Bodian 1997, 132-46).

4   Loyalty towards *Espanyol* and its New Names

From the beginning of the nineteenth century, nationalist states
– Bulgaria, Greece, Serbia, Romania, and Modern Turkey –
progressively replaced Ottoman rule. This had important
consequences for the language used by Sephardic Jews. Although
every country needs to be studied separately, with its own chronology
in linguistic matters,[5] we can make some generalizations that apply
to all. With the constitution of national states, Sephardic Jews faced
the necessity to learn the national language of the countries they
lived in. Inspired by nationalist ideas that were widely spread both
in Europe and in the Balkans in the nineteenth century, Sephardic
Jews developed their own version of nationalism based on language
identity. This can be seen in an effort to maintain the use of Judeo-
Spanish as a translocal language among Sephardic Jews in the
Balkans and in promoting 'authentic' Judeo-Spanish forms instead
of Balkan borrowings. The main vehicle for these new ideas was
the Sephardic press in Judeo-Spanish, which was prolific in the
second part of the nineteenth century (see Martínez Gálvez 2012).
A good example is a language debate in the journal *El Amigo del
Puevlo* (1894, Sofia), which provides a unique insight into what
some Sephardic Jews in the Balkans thought about the language
they spoke and how they felt about it. From a reader's letter (S.R.,
Amigo, 12, 15th of February 1894, 181-3), we learn that a group of
Sephardic Jews in Belgrade founded a society whose purpose was, as
the author puts it, "to fight against the use of the Spanish language",
which pushed him to react "before it's too late". The founders of
this society are referred to as "young compatriots" and the society
is probably *Srpsko-Jevrejska Omladinska Zajednica* 'Serbian-Jewish
youth community', whose declared aim was to promote the Serbian
language and patriotism among Jews.[6] S.R.'s letter triggered strong
reactions from other readers. Five other letters on the same topic
were published subsequently and, according to the editors, many
more were received. Only one of the readers pleaded for the adoption
of the national language instead of the "language of our enemies",
viz. Spanish (R.F., *Amigo*, 13, 1st of March 1894, 200-2). Three
other readers, as well as the editors of the journal, were against
abandoning Spanish. Their major arguments in favor of keeping
Spanish are consistent with the idea of preserving Jewish identity
and the Jewish way of life. For them, using Spanish was a way to

---

5   See Şaul (1983) for adopting Turkish in Turkey and Vučina-Simović (2016) for
adopting Serbian in Serbia and ex-Yugoslavia.

6   This society is mentioned in Šlang (1926, 97), but we know little about their activities
prior to 1900.

communicate with Sephardic Jews scattered "all over the world". The authors warn that abandoning Spanish would facilitate religious assimilation; they also argue that speaking Spanish preserves the memory of the illustrious Jewish Iberian past. Beyond the Jewish context, other advantages are seen in speaking Spanish: it facilitates learning other Romance languages, and it is useful for commercial activities. They also emphasize that it is a prestigious language with a very old and rich literature. One of the participants in the debate, D.Y. Mefano, even states that Spanish is his "mother tongue" (*Amigo*, 14, 15th of March 1894, 219-20), which represents a completely new way of thinking about the vernacular language among Sephardic Jews.

One of the arguments of the reader who promoted the use of the national language(s) instead of Spanish was that Sephardic Spanish was "corrupted" and different from "real" Spanish. The participants in the debate in *El Amigo del Puevlo* appear to be aware of differences between the peninsular Spanish and their variety, but judged them similar enough to ensure mutual communication and practical advantages. They were, however, convinced that their vernacular language should be refined more. On the pages of the same journal, even prior to this debate, we can observe how the glossing technique is used to promote Spanish lexical items at the expense of well-entrenched borrowings from Turkish, Greek, and Serbian, given in parentheses as translations: *puentes* (*kyupres*) 'bridges', *ruido* (*shemata*) 'noise', *kampo* (*lagir*) 'field', *muraya* (*kale*) 'fortress', *aguelo* (*papu*) 'grandfather', *akto* (*uverenya*) 'record'.[7] It is even more striking to observe that in some cases a Spanish lexeme was preferred to a well-known Hebrew word: *pretes* (*galhim*) 'priests'. The ideas expressed by the 'defenders' of Spanish in the debate in *El Amigo del Puevlo* foreshadow the emergence of *Sephardism*, a national Sephardic movement in which Spanish language and identity were one of the central themes. One of the participants in the debate, a 'Student' from Vienna, likely belonged to the Balkan Sephardic students who founded *La Esperansa* in 1896, a society which spread *Sephardism* ideas and was active until (at least) 1928. Even when the members of this society embraced Zionism in 1900, they continued defending the use of the Spanish language and Sephardic cultural identity (Vučina Simović 2013, 341). The dominance of Yiddish speakers within the Zionist movement was perceived as a cultural threat by some Sephardic Zionists, leading to a reaction centered on the role of language. On the margins of the 14th Zionist Congress in 1925 (Vienna), a group of Sephardic delegates founded a political organization named *The World Sephardi Federation*, whose aim was

---

7 The examples given here are found in the first six issues of *El Amigo del Puevlo*, 1888-89 (Stulic 2014, 158-61).

to provide a world-wide network for the protection of Sephardi Jews' interests and their identity, and which continued its activities even until after the WWII (Fabre 2023).

Sephardic intellectuals in the late nineteenth and early twentieth centuries helped develop scholarship into Sephardic 'national' history, language, and literary tradition, establishing a dialogue with the Occidental research tradition in straight line with the *Haskala* ('Jewish Enlightenment') movement. In their first academic publications about their language or its literary tradition, they adopted the term *Judeo-Spanish* (and its equivalents) as a name of the language. *Jüdisch-Spanisch* was first used by German-speaking scholars: Kayserling (1857), Grünwald (1882), and Grünbaum (1896). From there, it entered the French language: *judéo-espagnol* is first found in a review of an article by M. Grünwald in the French *Revue des Études Juives* (1881, 323) and was subsequently adopted by French-speaking Sephardic intellectuals, who wrote about Judeo-Spanish and its literary tradition in the same journal (Danon 1896) or elsewhere (Arditti 1904; Danon 1904). This paved the way for translating it into other languages, including Judeo-Spanish itself. The equivalents of "Judeo-Spanish" in different languages were used in the first descriptions of this language based on the fieldwork done by philologists, both native speakers (Baruch 1923; 1930; Luria 1930; Romano 1933)[8] and non-native speakers of the language (Subak 1905; 1906a; 1906b; Wagner 1914; 1923; 1930; Crews 1935). Unlike Bunis (2011, 69), I believe that the prestige accorded to peninsular Spanish played only a minor role in why native-speaker linguists adopted this term for the vernacular language. This name offered them the possibility of 'objectivizing' the common paraphrase *espanyol muestro* ('our Spanish') used to denote the language they spoke, just as *religyon* or *relijyon judayka* started to be used in the nineteenth century instead of *muestra ley* ('our law'), the traditional denomination for Judaism (Stulic 2014).

In the late nineteenth century, another term, *Judezmo* (>Sp. *judaísmo* 'Judaism') was sometimes used to refer to the vernacular language. This word, which means 'Jewish beliefs and practices', was extended to mean 'religious enthusiasm' (*ǧuđésmo*, Nehama 1977, 238) and – as Bunis shows very well (Bunis 2011, 50-4) – the 'Jewish way of speaking' or 'Jewish language', almost always in opposition to a non-Jewish language. From all the examples of the use of this term given in Bunis (2011), the earliest occurrence of *Judezmo* to unambiguously refer to the Sephardic vernacular is found in A. Benghiatt's journal *El Meseret* (1897, Izmir) (Bunis 2011, 59-61). According to Bunis,

---

8 Kalmi Baruch and Samuel Romano (both died in the Shoah) were ideologically related to *Sephardism*.

the editor's substitution of *Espanyol* by *Judezmo* was due to the limited success of the first issue of his journal. The name change was accompanied by a change in register towards a more popular one in the following issues. Benghiatt's metalinguistic comments quoted by Bunis (2011, 60-1) largely confirm this interpretation. As can be inferred, *Judezmo*, as a name of the spoken language, is contemporary with the appearance of the term *Judeo-Spanish*. It also represents an expression of loyalty towards the Sephardic vernacular language, similar to that we see in *El Amigo del Puevlo,* but it does so with different means, through the covert prestige of the colloquial style. *Judezmo* appeared only occasionally as a name of the language (for the decline of the term, see Bunis 2011, 67-9) and was not accepted by Sephardic intellectuals,[9] who considered it inadequate as a name for a language for general purposes.

5　　Conclusions

In this article, I have presented three important ideas held by Sephardic Jews in the history of Sephardic *linguistic culture*. These ideas are crucial for understanding both the external and internal histories of Judeo-Spanish, particularly in the context of language contact in the Balkans.

The primacy of the Holy Tongue and its coexistence with the vernacular language led to the development of multilingual practices: the learning of Hebrew (and Aramaic) and the translation and glossing of sacred texts into the spoken vernacular – which was considered ontologically secondary and thus dependent on the *hierogloss*. These multilingual practices also made it possible to introduce forms from other languages without being constrained by the religious and symbolic authority of the sacred language.

In the sixteenth century, the language resulting from the shift from a variety of Iberian Romance languages to Castilian played a key role in the *rejudaization* of baptized Jews who had emigrated from Spain and Portugal. It was referred to as *Espanyol*, a name which acquired a specifically Jewish meaning, used alongside terms such as *Djudyo* or *Djudio* ('Jewish').

9　Note that A. Danon qualifies the word *Judezmo* by the past participle of the depreciative Fr. verb *affubler* 'dress ridiculously': "Cette invasion de l'osmanli dans la littérature gnomique du Mahalé, nous amène à parler de l'intrusion, dans le judéo-espagnol, des vocables turcs, qui conjointement avec les mots grecs, hébraïques, bulgares, etc. constitue le jargon ou Ladino, affublé du titre honorifique de Judesmo par les descendants des exilés de l'Espagne et du Portugal, qui l'ont ainsi entouré d'une auréole sacrée" (Danon 1904, 6).

The adoption of national languages and the spread of nationalist ideologies in the nineteenth century gave rise to a feeling of loyalty to *Espanyol*, which led speakers in some contexts to avoid borrowings from Balkan languages (and Hebrew), and also stimulated academic interest among Sephardic scholars in the study of the language.

As we have seen, speakers used different names for their language at different times and in different contexts: *La'az* ('foreign') and *Romanse* for the Romance varieties spoken around the time of the 1492 expulsion; *Espanyol* – with the specific meaning of 'Jewish' – along with *Djudio* or *Djudyo*, from the sixteenth century onwards. In the late nineteenth century, two new terms appeared: *Judeo-Spanish* (and its translations), mainly used in academic and multilingual contexts, and *Judezmo*, often with a connotation of self-deprecation. In the post-Shoah period, *Ladino* gained prominence in an attempt to keep the spoken language aligned with the written rabbinic tradition. The question of how to name the language of Sephardic Jews became one of the main debates among scholars in Sephardic studies during the second half of the twentieth century (see Şaul 2013, 196-205). The severely endangered status of the language, however, has moved the focus of researchers towards more pressing concerns: advancing its linguistic description, making textual materials accessible through editions and electronic corpora, and engaging in revitalization efforts. Both researchers and speakers have since adopted a more conciliatory stance regarding the name of the language, and it is now common in academic works to mention multiple names alongside the one preferred by the author.

Bibliography

Adamou, E.; Sobolev, A.N. (forthcoming) (ed.). *The Balkan Linguistic Area. An Atlas of Variation and Contact*. Berlin: De Gruyter Brill.

Arditti, E.S. (1904). "La lingua giudeo-spagnuola e la sua evoluzione". *Corriere Israelítico*, 43, 131-3, 157-9.

Alonso, A. (1968). *Castellano, español, idioma nacional; historia espiritual de tres nombres*. Buenos Aires: Losada, 20-30.

Bodian, M. (1997). *Hebrews of the Portuguese Nation: Conversos and Community in Early Modern Amsterdam*. Bloomington: Indiana University Press.

Bunis, D.M. (2011). "Native Designations of Judezmo as a 'Jewish Language'". Tobi, Y.; Beinart, H.; Kurzon, D. (eds), *Studies in Language, Literature and History Presented to Joseph Chetrit*. Jerusalem: Carmel, 41-81.

Bunis, D.M. (2018). "Judezmo (Ladino/Judeo-Spanish): A Historical and Sociolinguistic Portrait". Hary, B.; Bunin Benor, S. (eds), *Languages in Jewish Communities, Past and Present*. Berlin: De Gruyter Mouton, 185-238.

Crews, C.M. (1935). *Recherches sur le judéo-espagnol dans les pays balkaniques*. Paris: Droz.

Danon, A. (1896). "Recueil de romances judéo-espagnoles chantées en Turquie". *Revue des études juives*, 32(63), 102-23.

Danon, A. (1904). "Essai sur les vocables turcs dans le judéo-espagnol". *A Magyar neprajzi tarsasag keleti szakosztalyanak kiadvanyai / Publications de la Section Orientale de Société Ethnographique Hongroise*, vol. 3.

Fabre, D. (2023). *Les recompositions de l'identité sépharade dans les colonnes du Judaïsme Sépharadi (1932-1966). Une revue sépharadiste plurilingue entre Paris et Londres* [Master Dissertation]. Bordeaux: Bordeaux Montaigne University.

Friedman, V.A.; Joseph, B.D. (2014). "Lessons from Judezmo about the Balkan Sprachbund and Contact Linguistics". *International Journal of the Sociology of Language*, 226, 3-23.

Friedman, V.A.; Joseph, B.D. (forthcoming). *The Balkan Languages*. Cambridge: Cambridge University Press.

Grünbaum, M. (1882). *Jüdischdeutsche Chrestomathie: zugleich ein Beitrag zur Kunde der hebräischen Literatur*. Leipzig: Brockhaus.

Grünbaum, M. (1896). *Jüdisch-Spanische Chrestomatie*. Frankfurt am Main: J. Kauffmann.

Grünwald, M. (1882). *Zur Romanischen Dialektologie*. Heft 1, *Über den jüdisch-spanischen Dialekt als Beitrag zur Aufhellung der Aussprache im Altspanischen*. Belovar: Fleischmann.

Jakobson, R. (1960). "Closing Statements: Linguistics and Poetics". Sebeok, T. (ed.), *Style in Language*. Cambridge: MIT Press, 350-77.

Kayserling, M. (1857). "Jüdisch-spanische Gedichte", *Monatsschrift für Geschichte und Wissenschaft des Judentums*, 6(12), 459-62.

Luria, M.A. (1930). "A Study of the Monastir Dialect of Judeo-Spanish based on Oral Material Collected in Monastir, Yugo-Slavia". *Revue hispanique*, 79(176), 323-583.

Matras, Y. (2009). *Language Contact*. Cambridge: Cambridge University Press.

Martínez Gálvez, C. (2012). *El periodismo sefardí ante su público* [PhD Dissertation]. Valencia: University of Valencia.

Minervini, L. (2002). "La formación de la koiné judeo-española en el siglo XVI". *Revue de linguistique romane*, 66, 498-512.

Minervini, L. (2006). "El desarrollo histórico del judeoespañol". *Revista Internacional de Lingüística Iberoamericana*, vol. 4, 2(8), 13-34.

Moseley, C. (ed.) (2010). *Atlas of the World's Languages in Danger*. UNESCO. `https://unesdoc.unesco.org/ark:/48223/pf0000187026.locale=en`

Révah, I.S. (1970). "Hispanisme et judaïsme des langues parlées et écrites par les Sefardim". Hassán, I.M. (ed.), *Actas del Primer Simposio de Estudios Sefardíes.* Madrid: CSIC, 1970, 233-42.

Révah, I.S. (1965). "Formation et évolution des parlers judéo-espagnols des Balkans". Straka, G. (ed.), *Actes du Xe Congrès International de Linguistique et Philologie Romanes (Université de Strasbourg, avril 1962),* vol. 3. Paris: Librairie Klincksieck, 1351-71.

"Revue bibliographique" (1881). *Revue des Études Juives*, 2(4), 306-33.

Robert, J.-N. (2006). "Hieroglossia: A Proposal". *Nanzan Institute for Religion and Culture*, Bulletin, 30, 25-48.

Romeu Ferré, P. (ed.) (1998). *Crónica de los reyes otomanos*. Barcelona: Tirocinio.

Roth, C. (2025). "The Role of Spanish in the Marrano Diaspora". Pierce F. (ed), *Hispanic Studies in Honour of I. González Llubera*. Oxford: The Dolphin Book Co. Ltd., 299-308.

Şaul, M. (1983). "The Mother Tongue of the Polyglot: Cosmopolism and Nationalism among the Sepharadim of Istanbul". *Anthropological Linguistics*, 25(3), 326-58.

Şaul, M. (2013). "What's In a Name? Ladino, Espanyol, Djudyo, Judeoespañol, Sefardi". Şaul, M. (ed.), *Judeo-Spanish in the Time of Clamoring Nationalisms*. Istanbul: Libra Kitapçılık ve Yayıncılık.

Şaul, M. (2023). "The Language Shift Origins of Judeo-Spanish". *Anuario del Seminario de Filología Vasca*, 57(1-2), 875-84.

Sephiha, H.V. (1973). *Le ladino. Judéo-espagnol Calque. Deutéronome. Versions de Constantinople (1547) et de Ferrare (1553)*. Édition, étude linguistique et lexique. Paris: Centre de recherches hispaniques.

Sephiha, H.V. (2003). "Mouvances judéo-espagnoles: impacts linguistiques". *Diasporas. Histoire et sociétés. Langues dépaysées*, 2, 65-74.

Šlang, I. (1926). *Jevreji u Beogradu*. Belgrade: Štamparija M. Karića.

Stulic, A. (2014). "L'emprunt ostensible face à l'emprunt invisible dans la création du style journalistique judéo-espagnol". *ReCherches*, 12, 147-64.

Subak, J. (1905). "Das Verbum im Judenspanischen". *Bausteine zur Romanischen Philologie, Festgabe für Adolfo Mussafia, zum 15. Februar 1905*. Halle: Max Niemeyer, 321-31.

Subak, J. (1906a). "Zum Judenspanischen". *Zeitschrift für Romanische Philologie*, 30, 129-85.

Subak, J. (1906b). *Judenspanisches aus Salonikki, mit einem Anhange: Judenspanisches aus Ragusa*. Trieste: Im Selbstverlage der Handelssektion der K.K. Handels- und nautischen Akademie

Vučina Simović, I. (2013), "Los sefardíes ante su lengua: los esperancistas de Sarajevo". Studemund-Halévy, M.; Liebl, Chr.; Vučina Simović, I. (eds), *Sefarad an der Donau. Lengua y literatura de los sefardíes en tierras de los Habsburgo*. Barcelona: Tirocinio, 341-60.

Vučina Simović, I. (2016). *Jevrejsko-španski jezik na Balkanu. Prilozi istorijskoj sociolingvistici*. Kragujevac: FILUM.

Wagner, M.L. (1914). *Beitrage zur Kenntis des Judenspanischen von Konstantinopel*. Vienna: Alfred Hölder.

Wagner, M.L. (1923). "Algunas observaciones generales sobre el judeo-español de Oriente". *Revista de filología española*, t. 10, 225-44.

Wagner, M.L. (1930). *Caracteres generales del judeo-español de Oriente. Revista de filología española–Anejo XII*. Madrid: Imprenta de la Librería y Casa Editorial Hernando

Weinreich, U. [1953] (1966). *Languages in Contact; Findings and Problems*. Pref. by A. Martinet. The Hague: Mouton.

On the Semantics of the Preposition *në* – A General Overview of its Use and Realization within Prepositional Phrases in Contemporary Albanian

Merima Krijezi
Universtiy of Belgrade, Faculty of Philology, Serbia

Predrag Mutavdzic
University of Belgrade, Faculty of Philology, Serbia

Abstract   Within the system of prepositions in Contemporary Albanian, the preposition *në* is classified in the group comprising the most representative Accusative prepositions expressing a variety of predominantly spatial relations. Our paper tackles the analysis of said preposition within prepositional phrases, its semantic meanings, scope, as well as its combinatory potential with other sentence constituents, from the aspect of morpho-syntax, semiology and cognitive linguistics. The analysis is directed towards identifying basic (primary) meanings which can be attributed to this preposition, in addition to the path of its secondary meanings.

Keywords   Semantics. Albanian language. Preposition në. Prepositional phrases. Lexical meaning. Cognitive Linguistics.

1 Introduction

The meaning of prepositions is one of the most complex issues in syntax, semantics, pragmatics and of grammar in general. Despite numerous interpretations and well-established linguistic concepts related to prepositions, this issue remains unresolved, giving the impression that there is always room for new interpretations of their meaning and functions.[1]

According to O'Grady (1997, 166), languages can be distinguished based on the way they categorise words according to the type of meaning they encode. While many linguists highlight the importance of formal criteria – such as the compatibility of a word with specific inflectional patterns, or its distributional behaviour, i.e., its ability to co-occur with particular types of words – they often treat meaning as just one of several relevant factors in determining word categories (cf. O'Grady 1997). Nevertheless, meaning/semantics play a crucial role in both the classification of words and in specifying their syntactic and functional properties within the language system.

Klikovac (2006) states that when defining the prepositions, at least two major difficulties arise, both of which are indicative of the nature of language meaning in general. The first one pertains to the inability of encompassing all the meanings of a single preposition in one general definition "regardless of how much the definition is "stretched" to encompass as many uses as possible" [...], there are still many remaining ones – very different ones, while the second one indicates that "a preposition must be defined by itself" (Klikovac 2006, 7), i.e., any definition of a preposition essentially includes using the preposition itself in defining its own meaning. Although prepositions are defined in many languages, including Albanian, as dependent and unchangeable part of speech, as well as a closed word group, it is a fact that they are such linguistic units (lexemes) that acquire their meaning within a broader structure, phrase/syntagm/sentence, and that, on the one hand, they are not entirely devoid of lexical meaning, while on the other, they also have a certain frequency of use. In many languages, Albanian included, they represent the most frequent category within the group of invariable parts of speech (Samara 1999), while in modern linguistics they are described as relation words that always occur with the linguistic unit they combine

---

1 The work is based on the results of the research conducted within the monograph entitled *Semantics of Prepositions in the Modern Albanian Language*, published by the first author of the paper, Merima Krijezi, in 2023.

with and are referred to as adpositions[2] (Bortone 2010; Matovac 2013; Šekrst 2013). They are defined as polysemic, non-parsable or parsable language units "which have a function to establish semantic and grammatical relation between the language unit with which they form an adposition phrase, and the other language unit in the phrase" (Hagège 2010, 8 in Matovac 2014, 7). A preposition itself, without a noun, can very rarely be used as an independent word in the subject [e.g. *On* is a preposition] or as an object [Write *across* properly] (Пипер, Клајн 2014, 205), nor is it typically found following a linguistic unit it combines with (in Serbian, preposition *radi* 'for the sake of', sometimes *uprkos* 'despite', *sebe radi* 'for one's own sake', *radi sebe* 'for the own sake', (Пипер, Клајн 2014, 205). Although an invariable part of speech, prepositions are not neutral in terms of grammatical categories such as number and form. Topalli (2011, 1111) emphasises that Albanian prepositions *ndër* 'among'/'between', *midis* 'between' and *ndërmjet* 'between'/'among' can be used only with a plural noun (e.g. *midis nesh* 'between us', but not *midis mua* 'between me'), as well as that use of some prepositions is possible only with nouns in an indefinite form (*Dola në oborr* 'I went out into the yard', but in no case *Dola në oborrin* 'I went out into the yard (of)' (Topalli 2011, 1120). In the Albanian language, prepositions cannot appear before a verb in the indicative mood, but may constitute a part of impersonal verb forms, such as participles (***për*** të punuar 'to work', ***për*** *të kërkuar* 'to look for'/ geg. ***me*** punue 'to work', ***me*** kërkue 'to look for').

In addition to its prepositional function and meaning, *në* in Albanian can also function as a conjunction – at the beginning of dependent conditional sentences (***në*** *ardhsha vonë, mos më prit* 'If I am late, don't wait for me'), in conjunction phrases (***në*** *qoftë se* 'if '/'in case', ***në*** *rast se* 'in case'/'if' etc.), and in sentences that have an indirect question form (*nuk e di* ***në*** *vjen apo jo* 'I don't know whether (s)he's coming or not'). Based on the research conducted by Spahiu (2010), who analysed the corpus of modern Albanian language on a sample consisting of literal and non-literal texts, or more precisely, on a sample of more than 35 million words (with the support of IT experts and the Albanian Academy of Sciences), one hundred most frequent words in Albanian language were identified. The preposition *në* ranked third on the list, while every identified word was marked by the ordinal number it appears under on the list, and with a multi-digit number that indicated its frequency. The preposition *në*, as the

---

2 "An unanalysable or analysable grammatical word constituting an adpositional phrase with a term that it puts in relationship, like case affixes, with another linguistic unit, by marking the grammatical and semantic links between them" (Hagège 2010 in Matovac 2013, 6).

third most frequent word in the corpus, had 1,199,646 occurrences (Spahiu 2010, 3) and was identified as the most frequently used preposition in the Albanian language (Çeliku 2019, 94). According to the British National Corpus (BNC; Burnard 2000), the most frequent words in the English language include prepositions "of, to, in and for" (cf. Baldwin, Kordoni, Villavicencio 2009), where preposition 'in', as one of the equivalents of the preposition *në*, occupies an important position and is ranked among top ten.

1.1   Aim and Scope

Our analysis is focused on the semantics of the preposition *në* in the Albanian language, and is motivated by the following reasons:
  a.   it is the most typical and most frequent preposition governing the accusative case, with a broad range of spatial and temporal meanings – 24 in total (according to FGJSSH 2006);
  b.   within its semantic range, *në* also expresses ablative-like meanings, indicating that an entity may be located either *inside* or *outside* – *on* another entity (*in* / *at* / *on*);
  c.   it denotes time and location in structures, indicates spatial relations, and marks the aspect of dynamics (movement) – a point towards which an entity moves, and final point of destination.

The objective of our research is to:
  a.   present the semantic range of the preposition *në*, including a description of its meaning structure and use;
  b.   determine the semantic network of the preposition *në*;
  c.   determine the types of structures in which it is realised; which parts of speech it most commonly co-occurs with, and which verbs affect its rection and semantics.

2   On the Semantics of Prepositions – A General Overview

In linguistics, prepositions are described as highly polysemic words. According to Lakoff (1987, 13), this polysemy stems from the fact that there are systematic relations between different cognitive models and elements of the same model. Croft and Cruse (2004, 111) argue that "polysemic units originate from the same lexical source, as a result of the extension process, such as of metaphor and metonymy", and that their polysemic property primarily depends on whether there is a perceived semantic relationship between two interpretations of a word or not, i.e., whether one interpretation is a convincing semantic extension of the other. Polysemy should, according to these authors,

enable variations in comprehending a single word across its different uses and realisations, as well as in singling out a specific meaning from a group of realistically plausible ones and simultaneously all possible meanings of the word within a structure. An additional observation of these authors is that a limited sense of a word is not a property of a lexical unit as such, but that it is rather constructed at the moment of actual use.

The appropriate understanding and acquisition of the meanings and semantic nuances of a preposition are not the same for all prepositions. While discussing prepositions in English, Lindstromberg (2010) emphasised that prepositions unrelated to space or time were observed as the least challenging ones, along with some prepositions of verbal origin. We would argue that this is also the case in the Albanian language, while noting that this group could also be extended so as to include prepositional phrases as well. Being complex prepositions, prepositional phrases generally do not denote time and place in Albanian. One of their key features is that their final constituent defines the rection of the following sentence constituent (Samara 1999; Agalliu et al. 2005), such as *me anë të* 'through'/'by means', *për shkak të* 'due to', *me ndihmën e* 'with the help of ', *falë ndihmës* 'thanks to the help', etc. This is clearly visible in the following instances: *me anë **të** studimit* (*të* + Ngen *studimit*) 'through the study'/'by means of studying', *për shkak **të** moskuptimit* (*të* + Ngen *moskuptimit*) 'due to misunderstanding', *me ndihmën **e** pajisjeve ortopedike* (*e* + Ngen *pajisjeve ortopedike*) 'with the help of orthopedic devices' and in the latter, *falë ndihmës **së** prindërve* (*së* + Ndat *prindërve*).

In Albanian, as in many other Indo-European languages, simple prepositions are polysemic, i.e., they have more than one meaning; complex ones most commonly have dual meanings, while prepositional phrases tend to have a single meaning. The lowest degree of lexical meaning clarity is expressed by simple prepositions, such as *në*, *me* 'with', *për* 'for', *nga* 'from', etc., whereas derived prepositions, such as *buzë* 'in front of', *mes* 'between', *afër* 'near', *kundër* 'against', "together with prepositional phrases, have the highest degree of lexical meaning clarity, whereat their semantic meaning clearly matches their formation productive base" (Çeliku 2019, 16). Topalli (2011, 1114) states that both regular prepositions, as well as those formed later through the internal development of the Albanian language, are associated with two cases: the accusative and the ablative. Tracing the diachronic development of the Albanian language through written documents and preserved historical materials, we can say that the accusative is one of the primary cases used with prepositions from the early stages, as well as that it aligns semantically and syntactically with a large number of prepositions, primarily *në*, *më* 'with', *për* 'for' and *nëpër* 'through', but also with several others

that are specific to this case. Texts from the sixteenth century indicate that all prepositions in the Albanian language were used with only one case, the so-called *kallëzore parafjalore* 'prepositional accusative', while in its pre-written phase, Albanian had a single case for the formation of prepositional syntagms (Topalli 2011, 1115). The entire group of prepositions used with this case in modern Albanian corresponds to the accusative both semantically and syntactically (Topalli 2011; Demiraj 1989), with the preposition *në* occupying the primary position in this group. The only preposition that represents an exception to this rule is the preposition *n* (Buzuku's *en = ṇ*), which corresponds semantically and syntactically to the ablative (Topalli 2011, 1117), although several examples of its use with the accusative are found in Buzuku's texts. The emergence of the preposition had significant consequences for the case system, where "in time, the preposition acquired a leading role in prepositional structures, while the case inflection of a noun became almost a redundant means, i.e., it turned into an additional means serving the preposition" (Demiraj 1989, 628). In a broader sense, prepositions took over the function previously performed by case inflections.

2.1   On the Semantics of the Preposition *në* in Albanian

When it comes to the meaning of a preposition, it has to be observed from a broader perspective, which includes all of its possible meanings as defined in grammar books and examples in the target language, meanings conditioned by social-cultural contexts, etymology, newly developed meanings etc., as well as its occurrence in different structures (syntagms, phrases, clauses) as language patterns. One of the properties of all natural languages is that they combine two or more words into units that are grammatically and semantically correct and comprehensible to all speakers of a given language. The number of such groups of words or structures is large in all languages, and according to Lindstromberg (2010, 5) "Some combinations of words have their own particular meaning to such an extent that if any of the words is replaced – even by a near synonym – that meaning disappears and/or the new combination seems odd". For instance, in English, the prepositions 'at' and 'in' can be synonyms used to denote where an entity is located. We can say: 'at school', 'at the faculty', but not 'at classroom'; we must always say 'in the classroom'.

In Albanian, we can say: *jam në shtëpi* 'I am at home', but not *jam mbi shtëpi* 'I'm above the house/on the house' (the second example would mean that we are above the house/on the house), although these are prepositions with synonymous meaning.

Grammar books and dictionaries of the Albanian language (a full list is provided at the end of the paper) list the following meanings of the preposition *në* (when used with a noun in the aaccusative case):

1. a place where someone or something is located, where an action is carried out;
2. destination of an entity's movement, as well as a place where an entity is placed or closed in;
3. final destination of a movement;
4. an entity (object, living creature, abstract item) which we come across, or we encounter during a movement, when we search for something and the like;
5. an object through which something runs, or by means of which some action is carried out;
6. time (when an action is carried out, or within which an action is carried out);
7. state (condition of a location where an entity is positioned, circumstances in which an action is carried out);
8. manner;
9. shapeshift (change of form), or the state of an entity;
10. business, activity, group, organisation, or similar, where someone works or participates in operation;
11. base, foundation, or site of mental or spiritual activity, or various forms of such activities;
12. goal, intention, reason or benefit for which something is done;
13. a whole, composed of parts (countable), which can be divided into parts based on different criteria;
14. a thing or phenomenon which serves to more closely define entities, characteristics, properties, abilities thereof, and the like;
15. relationship, bond with something or another entity;
16. entity under supervision, custody, monitoring, care, and the like;
17. event in which an entity participates;
18. colour or shade of an entity;
19. garment or object in which an entity is wrapped or dressed;
20. a member of a group;
21. with an accusative noun, it denotes a constituent of a word group, expression, phrase with verbal meaning;
22. in syntagms with prepositional meanings;
23. in phrases with a prepositional function: *në kundërshtim me* 'contrary to', *në bashkëpunim me* 'in collaboration with', *në krahasim me* 'in comparison to', *në emër të* 'on behalf of', *deri në* 'until', *në bazë të* 'on the basis of', etc.;
24. point of an action closure, (in conjunction with the preposition *nga* 'from'), extension or continual movement in space and time, or repetition of an action or state.

All listed meanings can be divided into two groups – a) with spatial and b) with non-spatial uses of the preposition *në*.

a. Spatial uses of the preposition *në*:

1. A place where someone or something is located, or where an action is carried out:
   Jemi në pastiçeri 'We are *at the pastry shop*';

1.a A place within an entity, observed as an enclosed space:
   Në shtëpi / shkollë / kinema… 'In the house / at school / in the cinema…'
   Në shpirt / errësirë / borxhe 'In the soul / in the darkness / in debts'

1.b A place on the surface and/or upper side of an entity:
   Në tryezë 'On the table' *Në shesh / rrugë / majë* 'On the square / in the street [on the road] / on the top' *Në letër* 'On paper'

Në / On

Në   **Në** / In

Illustration 1   Illustration 2

2. Destination of an entity's movement:
   Shkoj në treg 'I am going **to the market'**

3. Final point of movement:
   U ngjitën në çati 'They climbed **onto the roof'**
   Më në fund erdhën në fshat 'Finally, they arrived **at the village**'

b. Non-spatial uses:

1. Time (when an action is carried out or within which it is carried out):
   Në pranverë 'In spring'; *Në fëmijëri* 'In childhood'

2. The entity being acted upon:
   Foli në cellular 'He spoke **on the cellphone'** *Shkoj perin në gjilpërë* 'I thread (the thread **in) the needle**'

3. State (where an entity is positioned, circumstances in which an action is carried out):
   Është në humor të mire 'He is **in good mood'**; *Në ankth* **'In agony'**

4. Manner:
   Rri në këmbë 'Standing tall'(literally 'being **on legs**'); *Bie në gjunjë* 'Fall **on knees**'

5. Goal, intention, reason or benefit:
   Në luftë kundër korrupsionit '**In combat** against corruption'; *Në dobi* të komunitetit local '**In favour of** the local community'

6. A whole composed of (countable) parts, which can be divided based on different criteria:
   *Drama **në katër akte*** 'A drama **in four acts**'

7. Providing further information about an entity:
   *Punim **në gur*** 'Relief **in stone'**; *Kampion **në kërcim së larti*** 'Champion **in high jumping**'

8. Relationship, bond (with an entity):
   Në lidhje me përdorimin e dokumentacionit tatimor '**In relation to** the use of tax documents'; *Në konflikt me ligjin* **'In conflict** with law'; *Ata janë **në lidhje*** 'They are **in a relationship**'

9. Providing additional information about an action:
   *Isha **në një festë*** 'I was **at a party**'; *Mora pjesë **në një garë**.* 'I took part **in a competition**'

10. Affiliation within a group / group member:
    Në më të varfërit '**Among the poorest ones**'; *Në të prekurit* nga tërmeti '**Among those** hit by the earthquake'; *Në të mbijetuarit* '**Among the survivors'**

Synonymy includes both vertical and horizontal branching of meanings of prepositions. The authors of *A Comprehensive Grammar of the English Language* (Quirk et al. 1985) state that it is very difficult to classify the areas of meanings of prepositions, and that in some cases, it is better to think about the scope or spectrum of meanings, first as one separate category, and then also as a category divided into separate, yet mutually concordant segments. In the case of a large number of prepositions, we can talk about synonymous semantic branching in two planes – a vertical one, which pertains to all possible meanings that are synonymous to the meaning of another/other preposition(s), and a horizontal one, which pertains to semantic concordance with a high degree of equivalence (we can never talk about absolute or one hundred percent synonymy) and its usage value.

The horizontal plane should demonstrate which prepositions have a high degree of semantic equivalence and usage value, and in a large number of cases can be used in a sentence to replace the other ones, carrying the same semantic content and not altering the meaning of the sentence, as well as which preposition is the first choice as a possible synonym for one preposition.

In the case of the preposition *në*, the synonymy with preposition *mbi* was analysed, referring to only one aspect of their meaning – locative meaning. There are two planes in focus:
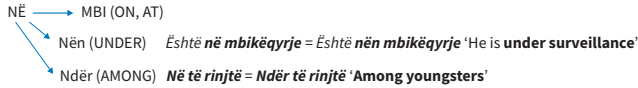
NË ⟶ MBI (ON, AT)

    Nën (UNDER)   *Është **në mbikëqyrje** = Është **nën mbikëqyrje*** 'He is **under surveillance**'

    Ndër (AMONG)   ***Në të rinjtë** = **Ndër të rinjtë*** '**Among youngsters**'

Illustration 3   A schematic overview of the synonymous meaning and its vertical branching on the example of prepositions në and mbi

The prepositions *në* and *mbi* are synonymous, and their synonymy is reflected in expressing a place where an entity is located, an action is carried out, and the like:

*E vuri librin **në tryezë** = E vuri librin **mbi tryezë*** 'He put the book **on the table**'

*Ai u ngrit dhe vuri kapelën **në kokë** = Ai u ngrit dhe vuri kapelën **mbi kokë*** 'He stood up and put the hat **on**'

Analysing the meanings listed in FGJSSH (2006) for the marker – preposition *në*, we noticed that this preposition can have a synonymous meaning with two more prepositions:

    a. with the preposition *nën* when denoting an entity under our surveillance, control:

        *Është **në mbikëqyrje** = Është **nën mbikëqyrje*** 'He is **under surveillance'**

        *E ka **në kujdestari** = E ka **nën kujdestari*** 'They are **under his/her custody'**

    b. and with the preposition *ndër* when denoting affiliation within a group:

        Në të rinjtë** = **Ndër të rinjtë '**Among youngsters**'

        Në të sëmurët** = **Ndër të sëmurët '**Among patients**'

Preposition *mbi*, in addition to the synonymous meaning with the above mentioned preposition *në*, can also have a synonymous meaning with the prepositions *përmbi*, *sipër* and *për*:

MBI ⟶ NË (ON)

    Përmbi (ABOVE)   *Një pëllëmbë **mbi kokë** = Një pëllëmbë **përmbi kokë***
                      'An inch **above the head**'

    Sipër (ABOVE, ON)   ***Mbi mesataren** = **Sipër mesatares** '**Above average**'
                     *U shtri **mbi rërë** = U shtri **sipër rërës*** 'He lay down **on the sand**'
                     (http://www.fjalori.shkenca.org/)

    Për (ON)   ***Mbi përdorimin** e bimëve mjekësore = **Për përdorimin** e bimëve mjekësore*
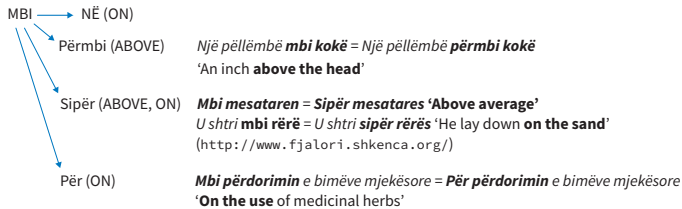   '**On the use** of medicinal herbs'

Illustration 4

1. *përmbi* to denote a place, or the position of an entity above another entity:

   *Një pëllëmbë **mbi kokë*** = *Një pëllëmbë **përmbi kokë*** 'An inch **above the head'**

   *Urë **mbi lumë*** = *Urë **përmbi lumë*** 'A bridge **on the river**'

2. The preposition *sipër* denotes states or quantities greater than usual, or the one taken as a basis, starting point:

   Mbi mesataren = ***Sipër mesatares*** **'Above average'**

2.a to denote a place, or the position of an entity above another entity:

   Mbi tokë = ***Sipër tokës*** '**Above the ground**'

2.b places where an entity moves around and an action is carried out:

   *Rrinin **mbi kreavate*** = *Rrinin **sipër krevateve*** 'They were sitting **on the beds'**

3. the preposition *për* (in scientific, publishing and journalistic style, in titles of books or articles):

   Mbi përdorimin *e bimëve mjekësore* = ***Për përdorimin*** *e bimëve mjekërsore* '**On the use of** medicinal herbs'

   Mbi gramatikën *e persishtes* = ***Për gramatikën*** *e persishtes* **'On the grammar of** the Persian language'

Note that use of the preposition *për* in such structures is more accurate and is the recommended orthographical norm.

NË ——→ MBI ·······▸ SIPËR

U shtri NË / MBI / SIPËR rërë/ rërës 'He lay down **on the sand**'

Illustration 5   Schematic overview of horizontal branching of synonymous meanings of the prepositions *në* and *mbi*

The analysis of the synonymous semantic branching of the preposition *në* indicated that vertical branching connects a greater number of prepositions with synonymous meanings (the prepositions *nën* and *ndër*) compared to horizontal branching, which implies synonymy with only one preposition (the preposition *mbi*). Preposition *mbi* has a high degree of synonymic equivalence with the preposition *në*, including only one aspect of their meaning – locativeness. Synonymous prepositions along a vertical branching line have a more limited degree of equivalence of synonymy with the proposition *në*.

2.1.1 Types of Structures Within which the Preposition *Në* is Realised (which Verbs Affect Its Rection and Semantics)

Applying a contrasting methodology to the basic and primary meaning, as well as the secondary, derived and metaphorical meanings of prepositions, we analysed the relations they express in Albanian language structures. In this analysis, we used and referenced all available and relevant monolingual and bilingual dictionaries of modern Albanian in both electronic and hard copies, as well as internet sources as auxiliary or secondary references. More than 30,000 pages were analysed, covering literal and non-literal content, parts of discourses, spoken language, and the like. The analysis demonstrated that nouns are the part of speech the prepositions are most commonly used with in a structure, but they can also be found with pronouns, adjectives and other noun words (like denominal adjectives) or adverbs.

The verbs found before the prepositional structure, which affect its rection and semantics, are state verbs: *jam* 'to be', *gjendem* 'to be located', *ndodhem* 'to be located', *rri* 'to sit/to socialise/to stand', *qëndroj* 'to stand', *shikoj* 'to look', *shtrihem* 'to spread', *banoj* 'to live' etc.; activity verbs, verbs of movement and the like: *nisem* 'to set off/ to leave/to go', *shkoj* 'to go', *eci* 'to go/to walk', *iki* 'to leave/to escape', *arrij* 'to arrive', *vij* 'to come', *dal* 'to go out', *kthehem* 'to return', *zbres* 'to descend', *hyj* 'to enter', *vrapoj* 'to run', *marr* 'to take/to gain', *lë* 'to leave', *jetoj* 'to live', *shëtit* 'to walk', *flas* 'to talk/to speak/to tell', etc.

3 Conclusion

The conducted analysis referring to the meaning of this preposition in a structure demonstrated that the primary meaning of the preposition *në* is spatial. With a noun in the accusative case, it has the meaning "which is close to the basic accusative meaning, which is border directiveness, reaching a goal or realizing some kind of activity related to a certain place (in a broader sense)" (Silić, Pranjković 2005, 113). It indicates a place where someone is located, a place where an action is carried out (locative meaning), and the place towards which an entity moves (adlative meaning), marking two positions:

1. a point within an entity observed as an enclosed space, and b) a point on the surface and/or upper side of an entity;
2. a point towards an entity moves, and b) final destination of the movement.

There are ten non-spatial semantic meanings, which refer to the following: time, the entity being acted upon, state (in which an entity

is, or circumstances in which an action is carried out), manner, goal, intention, reason or benefit; a whole composed of (countable) parts which can be divided based on different criteria; further information about the entity; relation/bond with an entity; further information about an action; and affiliation within a group (group membership).

The analysis of the synonymous semantic branching of the preposition *në* indicated that, compared to horizontal branching, which implies synonymy with only one preposition (the preposition *mbi*), vertical branching connects a greater number of prepositions with synonymous meanings. It also indicated that synonymous prepositions along a vertical branching line have a more limited degree of equivalence in terms of synonymy with the proposition *në*, and that the preposition *mbi* has a high degree of synonymy equivalence with the preposition *në*, including only one aspect of their meaning – locativeness.

The analysis also demonstrated that nouns are the part of speech the prepositions are most commonly used with in a structure, but that they can also be found with pronouns, adjectives and other noun words (such as denominal adjectives), or adverbs. Verbs that can affect its rection and semantics (found before the prepositional structure) include state verbs such as *jam* 'to be', *gjendem* 'to be located', *ndodhem* 'to be located', *rri* 'to sit/to socialise/to stand', *qëndroj* 'to stand' etc.; activity verbs, verbs of movement, and the like, such as *nisem* 'to set off/to leave/to go', *shkoj* 'to go', *eci* 'to go/to walk', *iki* 'to leave/to escape', *arrij* 'to arrive', *vij* 'to come', *dal* 'to go out', *kthehem* 'to return' etc.

Bibliography

Agalliu, F.; Andoni, E.; Demiraj, S.; Dhrimo, A.; Hysa, E.; Lafe, E. (2005). *Gramatika e gjuhës shqipe. Vëllimi I. Morfologjia*. Tiranë: Akademia e Shkencave e Shqipërisë / Instituti i Gjuhësisë dhe i Letërsisë.

Baldwin, T.; Kordini, V.; Villaiencio, A. (2009). "Preposition in Applications: A Survey and Introduction to the Special Issue". *Computational Linguistics*, 35(2), 119-49.

Bortone, P. (2010). *Greek Preposition from Antiquity to the Present*. United Kingdom: Oxford University Press.

Burnard, L. (2000). *Reference Guide for the British National Corpus*. Oxford, UK: Oxford University Computing Services.

Collins COBUILD (2011). *English Grammar (digital edition).* 3rd ed. Great Britain: HarperCollins Publishers. `https://akugue.com/docs/cobuild-english-grammar-4746565`

Croft, W.; Cruse, D.A. (2004). *Cognitive Linguistics*. UK: Cambridge University Press.

Çeliku, M. (2019). *Parafjalët e shqipes standarde*. Tiranë: Shtëpia botuese/shtypshkronja ILAR.

Demiraj, Sh. (1989). *Gramatikë historike e gjuhës shqipe*. Tiranë: Universiteti i Tiranës "Enver Hoxha", Fakulteti i Historisë dhe i Filologjisë, Shtëpia botuese "8 Nëntori".

Klikovac, D. (2006). *Semantika predloga*. Filološki fakultet. Beograd: Čigoja.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. (*What Categories Reveal about the Mind).* Chicago; London: The University of Chicago Press.

Lindstromberg, S. (2010). *English Preposition Explained*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Matovac, D. (2013). *Semantika hrvatskih prijedloga* [doktorski rad]. Osjek, Hrvatska: Sveučilište Josipa Jurja Strossmayera, Filozofski fakultet. `http://web.ffos.hr/download/Darko%20Matovac,%20Semantika%20hrvatskih%20prijedloga.%2011.%20listopada%202013..pdf`

Matovac, D. (2014). "The Semantics of the Preposition *po* in Croatian Language". *Croatica et Slavica Iadertina*, Zadar. `https://hrcak.srce.hr/file/201088`

O'Grady, W. (1997). *Syntactic Development*. 1st ed. Chicago: University of Chicago Press.

Пипер, П.; Клајн, И. (2014). *Нормативна граматика српског језика*. Друго измењено и допуњено издање. Нови Сад: Матица српска.

Silić, J.; Pranjković, I. (2005). *Gramatika hrvatskog jezika za gimnazije i visoka učilišta. Sintaksa, morflogija, fonologija*. Zagreb: Školska knjiga.

Samara, M. (1999). *Parafjalët në shqipen e sotme* (*Vështrim leksiko-semantik*). Tiranë: Panteon.

Spahiu, A. (2010). *100 fjalët më të shpeshta në gjuhën shqipe*. `http://www.shkenca.org/pdf/gjuhe/100_fjale.pdf` Šekrst, K. (2013). *Prijedlozi kao riječi uza sve leksičke kategorije s posebnim naglaskom na udvanje prijedloga u južnoslavenskim jezicima*. `https://www.researchgate.net/publication/357168301_Prijedlozi_kao_rijeci_uza_sve_leksicke_kategorije_s_posebnim_naglaskom_na_udvajanje_prijedloga_u_juznoslovenskim_jezicima_Prepositions_as_a_loose_lexical_category_with_special_emphasis_on_prepositiona#fullTextFileContent`

Topalli, K. (2011). *Gramatikë historike e gjuhës shqipe*. Tiranë: Qendra e Studimeve Albanologjike / Instituti i Gjuhësisë dhe i Letërsisë.

Quirk, R.; Greenbaum, S.; Leech, G.; Svartfik, J. (1985). *A Comprehensive Grammar of The English Language*. London; New York: Logman.

Dictionaries

Çuçka, Sh. (2006). *Fjalor anglisht-shqip*. Tiranë: Botimet Toena.

FGJSSH (1981). *Fjalor i gjuhës së sotme shqipe A-M*. Sipas botimit të Akademisë së Shkencave të RPSSH / Instituti i Gjuhësisë dhe i Letërsisë, Tiranë, 1980. Prishtinë: Rilindja.

FGJSSH (1981). *Fjalor i gjuhës së sotme shqipe N-ZH*. Sipas botimit të Akademisë së Shkencave të RPSSH / Instituti i Gjuhësisë dhe i Letërsisë, Tiranë, 1980. Prishtinë: Rilindja.

FGJSH (2006). *Fjalor i gjuhës shqipe*. Tiranë: Akademia e Shkencave e Shqipërisë/ Instituti i Gjuhësisë dhe i Letërsisë.

Elezi, M. (2023). *Fjalor i gjuhës shqipe*. Botimi i pestë. Tiranë: ALSAR.

Jashari, A. (2020). *Fjalor i homonimeve në gjuhën shqipe*. Tiranë: Akademia e Shkencave e Shqipërisë.

Qesku, P. (1999). *Fjalor Shqip-Anglisht / Albanian – English Dictionary*. Tiranë: EDFA.

Qesku, P. (2008). *Fjalor themelor i gjuhës shqipe*. Tiranë: EDFA.

Qesku, Pavli (2012). *Fjalor Anglisht-Shqip / English – Albanian Dictionary*. Tiranë: EDFA.

Xhuvani, A. (2020). *Fjalor i gjuhës shqipe*. Tiranë: Akademia e Shkencave e Shqipërisë/ Akademia e Studimeve Albanologjike.

From Fables to Newspapers: Turkish and Bulgarian Evidentiality in Written Genres

Vesela Simeonova
University of Graz, Austria

Seckin Arslan
CNRS, France

Abstract   We examine evidentiality in Turkish and Bulgarian across seven written genres (news, history, police reports, fiction, fables, religious texts, and dreams). The results reveal genre-specific patterns and show that similar markers are used differently in each language, highlighting cross-linguistic contrasts in evidential interpretation.

Keywords   Evidentiality. Bulgarian. Turkish. Narrative genres. Dream language.

1   Introduction

Evidentiality is the linguistic encoding of the source of information.[1] While all languages can express evidential meanings lexically (e.g. "reportedly" in English), about a quarter have specialized grammatical evidential paradigms (Aikhenvald 2004).

One of the major theoretical debates is whether evidentiality constitutes an independent grammatical category (see Aikhenvald 2004; Plungian 2002; De Haan 1999), or it is a part of the epistemic modality system.[2] While previous studies have focused on conversational data, written genres, due to their inherent limitations to certain types of evidence and degrees of certainty, provide a new testing ground for the two hypotheses. For example, reputable news reports and historical and encyclopedic texts convey a high degree of certainty (they communicate facts) but are based on indirect evidence. Conversely, narrating dreams describes events of low certainty but for which the speaker has, as it were, first-hand experience.

This paper documents evidential uses across seven narrative genres: traditional fables, modern fiction, religious texts (e.g. the Bible), news (journalism), police reports, texts about historical events, and dreams. The empirical focus of the study is on Bulgarian and Turkish, two languages that have long been in areal contact within the Balkan language area (Friedman 2008) with comparable yet different evidential systems. This allows us, additionally, to reveal finer-grained differences between them that may not manifest in prototypical conversational use.

We found that (i) evidential marking interacts with certainty in genre-specific ways; (ii) Bulgarian and Turkish show similar patterns in some genres, but differ in others; (iii) non-evidential forms interact with the use of evidential markers in complex ways.

2   Evidentiality in Bulgarian and Turkish

This section presents an overview on the evidential systems of Bulgarian and Turkish informed with data from constructed conversational scenarios.

---

[1] See Aikhenvald 2004; Chafe, Nichols 1986; Johanson, Utas 2000; Plungian 2001; Willett 1988.

[2] See Palmer 2001; Chafe, Nichols 1986; Givón 1982; Willett 1988; Izvorski 1997; Smirnova 2013; 2021.

2.1　Evidentiality in Bulgarian

Evidentiality in Bulgarian has been studied from multiple theoretical and methodological perspectives.[3] Below is a brief overview of the main properties relevant for this paper.

Bulgarian evidentiality is expressed by morphology intertwined with tense and aspect on finite verbs. In the past tense, which is the target data of this paper, evidential marking is obligatory and comes in three options illustrated below: direct (1), indirect (2), and reportative (3). The sentences are temporally and aspectually identical – they all use the verb 'to give birth' in the preterite ('aorist') tense (and perfective aspect) but are evidentially distinct. We propose that evidentials in Bulgarian in their prototypical use encode information source, as follows.

The direct evidential is used when the speaker witnessed or, more abstractly, consciously participated in a situation (cf. Korotkova 2020; Simeonova 2021). The sentences "My grandmother gave birth to my father in 1956" (1a) and "I gave birth to my son in 1956" (1b) both feature a very high degree of certainty, but they are evidentially distinct: In (1a) it is objectively impossible for the speaker to have direct evidence/experience of the event because it happened before she was born, and the direct form is infelicitous. In (1b), the direct form is acceptable as the speaker describes her own experience.

(1)　Direct evidential
　　a.　#Baba　　　mi　rodi　　　　　　　bašta　mi　prez　1956.
　　　　grandmother　my　gave.birth.3SG.AOR.DIR　father　my　in　　1956
　　　　'My grandmother gave birth to my father in 1956, <*direct*>.'
　　b.　Rodih　　　　　　sina　　si　　　　prez　1956
　　　　gave.birth.1SG.AOR.DIR　son.DEF　refl.gen　in　　1956
　　　　'I gave birth to my son in 1956, <*direct*>.'

The indirect marker signals indirect evidence of a general nature, underspecified for the exact type of indirect source. It is the form that the granddaughter would use instead of the infelicitous (1a), cf. (2a). It is sometimes suggested that the indirect is just an evidentially neutral general past (e.g. Alexander, Zhobov 2009). If this were so, it would be expected to be possible in an indirect context simply because it would be possible in any context. This would predict it to be possible also in direct evidential contexts, which is not the case, as (2b) shows.

3　See Friedman 2003; 2004; Fielder 1995; Izvorski 1997; Fitneva 2001; Alexander, Zhobov; 2009; Sonnenhauser 2013; Smirnova 2013; Rivero, Slavkov 2014; Korotkova 2016; Koev 2017; Arregui et al. 2017; Simeonova 2020; Tarpomanova 2021 a.o.

(2) Indirect evidential

  a. Baba   mi   e=rodila   bašta mi prez 1956.
      grandmother   my   gave.birth.3SG.AOR.IND   father my in   1956
      'My grandmother gave birth to my father in 1956, <*indirect*>.'

  b. #Rodila=sŭm   sina   si   prez 1956.
      gave.birth.1SG.AOR.IND   son.DEF   refl.gen   in   1956
      'I gave birth to my son in 1956, <*indirect*>.'

The reportative evidential marker is specialized for information whose source is specifically a report. It can be uttered, for example, after returning from having run into the neighbors, as a response to a question about what's new with them.

(3) Reportative evidential
     Kotkata   im   rodila.
     cat.DEF   their   gave.birth.3SG.AOR.REP
     'Their cat gave birth, <*reportative*>.'

As is clear from this context above, the reportative does not conventionally convey a reduced level of certainty (see also Izvorski 1997).

To summarize, Bulgarian has three morphologically and semantically different evidential markers that encode direct, indirect, and reportative information source, and in the past tense, there is no evidentially unmarked form, i.e. a choice between them is obligatory.

2.2   Evidentiality in Turkish

Turkish distinguishes between two evidential forms, direct and indirect, which are marked on finite verbs and morphologically related to tense and aspect (Şener 2011) and are obligatory in the past tense. The direct evidential, cf. (4a), marked by the suffix *–DI*, signals that the speaker personally witnessed the event or has forms of direct evidence through participation. The indirect evidential, as in (4b), formed with the suffix *–(I)mIş*, indicates that the speaker did not directly witness the event but instead obtained the information through report or inference (see Aksu-Koç, Slobin 1986; Slobin, Aksu 1982; Yavaş 1980).

(4) Turkish direct and indirect evidential

  a. Nurhayat çok   güzel   bir   şarkı söyle-di   (Arslan 2020, 3)
      Nurhayat very   beautiful   one   song   say.DIR
      'Nurhayat sang a very beautiful song.' <*direct*>

b. Nurhayat çok güzel bir şarkı söyle-miş
   Nurhayat very beautiful one song say.$_{IND}$
   'Nurhayat sang a very beautiful song.' *<indirect>*

It is debated whether Turkish evidentials are part of epistemic modality or a distinct grammatical category in their own right (see Aksu-Koç, Arslan 2024 for an overview). Another debate is whether Turkish marks both direct and indirect evidentiality, or whether the direct evidential functions as a neutral past tense marker (Johanson 2003). We follow the perspective that Turkish distinguishes between direct (firsthand) and indirect (non-firsthand) information source as illustrated above.[4]

2.3 Available Comparative Studies on Evidentiality in the Balkans

There are few comparative studies on evidentiality paradigms in Balkan languages, particularly regarding Bulgarian and Turkish. Makartsev (2014) conducted a comparative analysis of how grammatical (indirect) evidential markers appear in different genres of text across Bulgarian, Macedonian and Albanian. He analyzed excerpts from diverse text genres including folkloric texts (songs and ballads) shared across languages in the Balkans, as well as newspaper reports and literary texts. Makartsev concludes that evidential forms go beyond referring to information sources and, in these genres, they build a narrative frame and establish an epistemic author authority.

Bosnali and Kahraman (2015) conducted a study on the use of evidentials in retelling dreams in Japanese and Turkish. They recorded 12 dream narrations from 7 Turkish speakers, which had a total of 234 inflected verbs, of which 147 referred to events that happened within the dream. They counted both *–DI* and the present progressive form *–I(y)or* to be direct evidentials. In our study, we consider only *–DI* to be direct evidential. With that in mind, they report that in 6 of the 12 narrations there were no past forms whatsoever, 66% of the verbs were in the present tense. In the past tense, they report 23 uses of *–(I)mIş* and 30 uses of *–DI* (out of 170 uses).

Regarding Bulgarian and Turkish in particular, one comparative analysis is Grannes' (1996) volume, which delves into outcomes of language contact and briefly contemplates on the semantic relationship between the Bulgarian reportative and the Turkish

---

4 See Aksu-Koç, Slobin 1986; Arslan 2016; Kornfilt 1997; Lewis 1967; Sezer 2001.

indirect evidential. However, Grannes mainly focused on lexical borrowings in Bulgarian from Turkish.

Friedman (1999), to our knowledge, is the first direct comparative analysis of Bulgarian and Turkish specifically (along with more languages) focused on how genre affects the use of evidentiality. He analyzed a corpus of proverbs, arguing that this genre is interesting for evidentiality since it conveys "objective, universal truths" (135). He found that Bulgarian predominantly uses the reportative (117 out of 159 cases), while in Turkish, both direct and indirect evidentials are used (44 and 32 cases, respectively). He concludes that this reveals substantial differences in the evidential systems in the two languages despite their descriptively similar labels such as direct/confirmative or indirect/non-confirmative and links their use to their historical development.

However, proverbs as a genre are limited by their very nature as they are generic, abstract, and timeless. For that reason, proverbs rarely use a past tense to begin with, as Friedman (1999) confirms; the cases he found were among thousands of proverbs, which is a limitation since in both Bulgarian and Turkish, evidentials are only obligatory in the past (see above). Finally, proverbs are not a productive genre: the speaker merely quotes their content.

Our study is based on data from genres that communicate specific past events, both fictional and real, as described next.

2.4 The Present Study

In this study, we analyzed evidential markers in Bulgarian and Turkish across the following seven narrative genres: (i) general knowledge, (ii) police reports and/or minutes, (iii) Biblical/Quranic texts, (iv) modern fiction, (v) news reports, (vi) fables, and (vii) written recounts of dreams. All texts were extracted from contemporary media sources, for the Biblical/Quranic text we included their modern interpretations in Turkish and Bulgarian, and hence this study is purely synchronic.

3 Methods

From each genre in each language, we extracted 20 or more declarative sentences with finite verbs (see the previous section). An anonymized version of the data collected is publicly available at OSF with the following link: https://osf.io/ueynw/.

3.1 Bulgarian Data

The Bulgarian data collected across the seven categories is summarized below in Table 1 and discussed in more detail below. Special or shifted environments like the contents of letters, quotes, and direct speech were excluded from the data collection since they do not represent the main narrative device but are governed by their own internal rules, which are uniform across genres. They were most relevant in the categories Modern Fiction, Fables, and Bible.

Table 1 Details of text data collected for Bulgarian

Genre	Word count	Number of Sentences
General knowledge	934	52
Reports and minute records	469	32
Bible / Quran	314	19
Modern fiction, short story	224	21
News reports	584	27
Fables / folktale	566	158
Recounts of dreams	645	56

For general knowledge category, we collected six text samples ranging over the following different time periods: (i) Cold war (XX c.),[5] (ii) Roman empire (V c.),[6] (iii) Steve Jobs' biography (1955-2011),[7] and following excerpts taken from the Bulgarian Ministry of Education 9th and 10th grades history textbooks, *History and civilization for 9th grade* and *History and civilizations for 10th grade*:[8] (iv) Political systems in Europe (XX c.), (v) First settlements in Bulgaria (10,000-5,000 BCE), and (vi) Middle Ages (X-XVII c.).

For police reports, we used the official website of the Bulgarian Ministry of Internal Affairs,[9] which reports the daily activity of the

---

5 Accessible in electronic format at https://bg.wikipedia.org/wiki/%D0%A1%D1%82%D1%83%D0%B4%D0%B5%D0%BD%D0%B0_%D0%B2%D0%BE%D0%B9%D0%BD%D0%B0.

6 Accessible in electronic format at https://bg.wikipedia.org/wiki/%D0%A0%D0%B8%D0%BC%D1%81%D0%BA%D0%B0_%D0%B8%D0%BC%D0%BF%D0%B5%D1%80%D0%B8%D1%8F.

7 Accessible in electronic format at https://bg.wikipedia.org/wiki/%D0%A1%D1%82%D0%B8%D0%B2_%D0%94%D0%B6%D0%BE%D0%B1%D1%81.

8 Prosveta Sofia AD, authors Mihail Gruev, Boris Stoyanov, Kiril Slavchev, Yavor Siderov, ISBNs:978-954-01-4420-7and978-954-01-4278-4,availableinelectronicformatathttps://www.e-uchebnik.bg/product/istoriya-i-civilizacii-za-9-klas-2#&gid=1&pid=7 and https://www.e-uchebnik.bg/product/istoriya-i-civilizacii-za-10-klas.

9 Accessible in electronic format at https://www.mvr.bg/sdvr/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B5%D0%BD-%D1%86%D0%B5%D0%BD%D1%82%D1%8A%D1%80.

police, e.g. criminal actions like robberies, attacks, or accidents such as road accidents, fires, and missing people. The sample contained 469 words in 27 sentences, of which there were 32 finite verbs marking past events and 4 marking real present (e.g. "The investigation *is* still ongoing"). The Bible sample was taken from the Modern Bulgarian translation.[10] We sampled various parts of both the Old and the New Testament, documented individually in the data repository. For modern fiction category, we collected data from various novels in the category Modern Bulgarian novel on the free public library.[11] We sampled the most recent novels, published from 2007 to 2021. We chose works originally written in Bulgarian in order to eliminate any effects of translation. The list of all novels – both excluded and analyzed – is found in the supplementary material in the data repository.

News data were collected from three leading and reputable news websites.[12] Due to the nature of the evidential system described in the introduction, we focused on news about past events and excluded news about future events, such as about upcoming elections, planned events etc. We analyzed 6 news items with a total of 31 sentences. We only analyzed the main text of the news, noting that titles and subtitles form their own sub-genres in journalistic writing, which would need a larger sample to be analyzed separately; this goes beyond the scope of the present paper. We have included these in the raw data for completeness, but not in the analysis. For fables and/or folk tales, we analyzed data from five traditional Bulgarian fables collected and edited by Nikolay Raynov (adapted from Zahari Stoyanov 2005. ISBN 954-739-618-8), with a total of 158 sentences. For dreams, we collected data from a Facebook discussion group dedicated to interpreting people's dreams.[13] Authors post describing their dream and others comment their interpretation. We collected 56 sentences from 10 dream recounts by different authors.

3.2 Turkish Data

For Turkish data, text excerpts are from diverse sources described below, for a total of 189 sentences (2694 words), broken down in Table 2. General knowledge data were excerpted from three sources for a

---

10 Accessible in electronic format at `https://www.bible.com/`.

11 Accessible in electronic format at `chitanka.info`.

12 Accessible in electronic format at `https://novini.bg/`, `https://news.bg/`, and `https://www.vesti.bg/`.

13 Accessible in electronic format at `https://www.facebook.com/groups/sanovnika`.

total of 30 sentences: the Wikipedia entries for Nikia Riots,[14] and for the Halley Comet,[15] and a passage regarding the Renaissance period from the history school textbook published by the Turkish National Ministry for high school grade 11 (*Ortaöğretim Tarih 11 ders kitabı* ISBN 978-975-11-4963-3: MEB Yayınları).

Table 2   Overview of the composition of the Turkish data

Genre	Word count	Number of Sentences
General knowledge / history	472	30
Reports and minute records	429	20
Bible / Quran	473	40
Modern fiction, short story	337	25
News reports	423	23
Fables / folktale	264	22
Recounts of dreams	296	29

Police and official reports data were from the publicly available template reporting forms available online used for police reports for traffic and workplace accidents[16] and meeting minute reports from the Bursa Uludağ University Automotive Working Group meetings available online.[17] This data included a total of 20 sentences.

For religious texts, we extracted 20 sentences from The Quran's Chapter 8/75, *Surah Al-Anfal* 'The Spoils of War' (Diyanet İşleri Başkanlığı, 2011) and 20 sentences from The New Testament Bible's Chapter Luke 8/1–3 (New International Version, 2011).

For modern fiction, we extracted 25 sentences from Sait Faik Abasiyanik's short story *Karanfiller ve Domates Suyu* 'Carnations and Tomato Juice' (Abasiyanik 2021). For news reports, we extracted 23 sentences from two newspaper articles published in Oksijen, a national weekly news magazine in Turkey, including news about global cities index and Istanbul[18] and New York Times report on the aftermath of the 6th February Earthquakes.[19] For folklore tales,

---

14 Accessible in electronic format at https://tr.wikipedia.org/wiki/I._Justinianus#Nika_Ayaklanmas%C4%B1.

15 Accessible in electronic format at https://tr.wikipedia.org/wiki/Halley_kuyruklu_y%C4%B1ld%C4%B1z%C4%B1.

16 Accessible in electronic format at https://onikisubat.meb.gov.tr/www/tutanaklar/icerik/677.

17 Accessible in electronic format at https://uludag.edu.tr/ocg/tutanaklar-66745.

18 Accessible in electronic format at https://gazeteoksijen.com/turkiye/istanbul-kuresel-sehirler-endeksinde-25inci-siraya-yukseldi-194365.

19 Accessible in electronic format at https://gazeteoksijen.com/new-york-times/nyt-depremden-sonrasini-yazdi-bir-ailenin-mucadelesi-202095.

we took 22 sentences from two tales about Nasraddin Hodja, a folk character in Turkish and broader neighboring cultures.[20] And finally, written recounts of dreams comprised of 29 sentences from different sources including social media posts and discussion blogs including the threads under the *KizlarSoruyor* online anonymous blog (n.d.).

3.3   Procedure and Data Analysis

After extracting the data as described above, the authors annotated the evidential markers on every finite/matrix verb. We tallied the total number of direct and indirect evidential forms, as well as other relevant forms, such as the assumption and epistemic morpheme -DIR in Turkish and the historical present in Bulgarian, per genre, and calculated their percentage within the total number of finite verbs inflected for one of the evidential forms. The authors held meetings to compare annotation and data coding, so as to be able to follow a comparable strategy. For statistical analysis, we used Chi-square and Fisher's Exact tests to compare condition differences across each genre in each language.

4   Results

4.1   Results from Bulgarian Data

Table 3 presents a summary of the findings in Bulgarian in absolute counts and percentage.

In three of the seven categories, the results showed that one single form was categorically used. For Modern fiction and the Bible, this is the direct evidential, for Historical facts, it was the historical present, i.e. in that category, no evidential forms were used whatsoever. In Fables, the most dominant marker was reportative. An exact binomial test determined whether the proportion of REP occurrences (86 out of 90 tokens) was significantly greater than chance level ($p =.50$). The results indicated a statistically significant preference for REP over other categories, $p <.001$, 95% CIs [0.90, 1.00], with a sample proportion of 0.96.

---

20 Accessible in electronic format at https://www.milliyet.com.tr/cocuk/ eglence/nasrettin-hoca-fikralari-komik-fikralar-denildiginde-akla-gelen-nasreddin-hoca-fikralari-6214835.

Table 3   Summary of results for Bulgarian

	Bible	Modern fiction	Fables	Dreams	Wikipedia/ history textbook	Police report	News texts
words	314	224	566	645	934	469	584
sentences	19	21	158	56	52	27	31
Total number of evidentially marked forms	46	39	90	95	76	32	38
Direct	46 (100%)	39 (100%)	0	84 (88.4%)	0	1 (3.1%)	0
Reportative	0	0	86 (95.5%)	0	0	2 (6.2%)	11 (28.9%)
Indirect	0	0	0	0	0	12 (37.5%)	12 (31.5%)
Hist.pres.	0	0	4 (4.4%)	11 (12.2%)	76 (100%)	17 (53.1%)	15 (39.5%)

For Police reports, a Chi-square test revealed that the four categories were not equally distributed across the compiled text, $\chi^2(3, N = 32) = 31.50$, $p < .001$. The majority of tokens were HIST.PRES (n = 17) and IND (n = 12), while DIR (n = 1) and REP (n = 2) were rare. A Fisher's Exact Test revealed that the counts of IND and HIST.PRES conditions were not statistically different (p =.75, Odds Ratio = 1.40, 95% CIs [0.33, 6.64]). This suggests that police reports used both forms equally likely.

For News, the observed counts were: DIR (0), REP (11), IND (12), and HIST.PRES (15), totaling 38 observations. A Chi-square test was statistically significant ($\chi^2(3, N = 38) = 13.58$, $p =.004$), indicating that the categories were not equally distributed. Specifically, DIR had no occurrences, whereas REP, IND, and HIST.PRES were observed more frequently. Posthoc pairwise comparisons using the Fisher's exact tests with Bonferroni correction revealed that DIR differed significantly from all other conditions of REP, IND, and HIST.PRES frequencies (adjusted ps =.003,.001, and <.001, respectively). There were no significant differences between REP, IND and HIST.PRES (all adjusted $p = 1.00$).

4.2   Results from Turkish data

The frequency and percent occurrences of evidential forms found in the Turkish data is given Table 4. Based on the data extracted from general knowledge (historical and scientific texts), we found high use of the assumption/epistemic marker *–mIs+DIR* 17 (48.6%) and direct evidential *–DI* 11 (31.4%) comparative to indirect evidential *–mIs* and habitual *–DIR* markers. A chi-square goodness-of-fit test was significant $\chi^2(3, 35) = 15.17$, $p =.001$, indicating that occurrences of evidential markers were not equally distributed. We further

compared the counts between the assumption/epistemic and direct evidential markers (17 vs. 11). A Fisher's Exact Test revealed no significant difference (p = 0.22, Odds Ratio = 0.49, 95% CIs = [0.16, 1.43]). Therefore, we found that the general knowledge genre showed dominant uses of assumption/epistemic and direct evidential markers without a difference between the two.

In police/minute reports, generic –*DIR* 2 (13.3%) and –*mIs+DIR* 13 (86.6%) markers emerged only, and no other evidential forms were observed. A chi-square showed that those observed frequencies differed significantly from an equal distribution ($\chi^2(1)$ = 8.07, p = 0.004), indicating that generic assumed/epistemic form is used more often.

Table 4 Frequency and percent evidential forms appraised in different text genres in the Turkish data

	General knowledge	Police / minute reports	Religious texts	Modern fiction	News reports	Folktales	Dreams
Nr. Sentences	30	20	40	25	23	22	29
Total finite verbs	37	22	63	40	29	45	42
Finite verbs with an evidential form	35	15	44	16	25	30	39
Direct evidential -DI	**11 (31.4%)**	0 (0%)	**27 (61.3%)**	**13 (81.2%)**	**25 (100%)**	2 (6.6%)	**35 (89.7%)**
Indirect Evidential -mIs	2 (5.7%)	0 (0%)	2 (4.5%)	0 (0%)	0 (0%)	**28 (93.3%)**	4 (10.3%)
Habitual, Generic -DIR	5 (12.2%)	2 (13.3%)	15 (34.1%)	3 (18.7%)	0 (0%)	0 (0%)	0 (0%)
Assumption, Epistemic -mIs-DIR	17 (48.6%)	13 (86.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

In religious texts (Bible and Quran), direct evidential –*DI* 27 (61.3%) and habitual/generic –*DIR* form 15 (34.1%) were the most dominant markers. The use of indirect evidentials was merely minimal and restricted in the context where the information was quoted. A chi-square test showed a significant difference in frequencies, $\chi^2(3, 44)$ = 43.09, p <.001. A Fisher's Exact Test showed that difference between the evidential and the generic form was statistically significant (p =.015, Odds Ratio = 3.19, 95% CIs = [1.22, 8.72]). Thus, the direct evidential form was the most dominant in religious contexts. We should note that we counted a number of non-evidential forms, mostly observed under quotes and finite complements, as well, those included aorist (n=7), optative (n=3), imperative (n=6), progressive present (n=3).

In modern fiction, the direct evidential 13 (81.2%) was used more often than the other forms, including generic DIR 3 (18.7%), evidenced by a chi-square test ($\chi^2(1)$ = 6.25, p = 0.01). In fiction, several uses

of the aorist (n=14) and some present progressive forms (n=3) were found as serial verb constructions under a final direct evidential matrix verb.

In news reports the only form in the sample was the direct evidential. Folktale data showed the reverse, exclusively using indirect evidential form 28 of 30 finite verbs analysed (93.3%). The direct evidential forms found here (n=2) were in fact used on quoted complements.

In dreams, direct evidential –*DI* was used in 35 out of 39 analyzed finite verbs (89.7%); the other 4 were the indirect evidential. The frequency of these observations was significantly different from being equally distributed ($\chi^2(1) = 24.64$, $p < 0.001$). The indirect evidential –*(I)mIş* was used in context of mirativity or lack of consciousness. This is in contrast with Bosnali and Kahraman (2015), who found that about half of the dream narrations they collected were narrated entirely in the present tense (see above).

5 Discussion

Our aim was to investigate how Bulgarian and Turkish use evidentiality across seven text genres: general knowledge (textbooks, Wikipedia), police reports (and minutes), Biblical/Quranic texts, modern fiction, news reports, fables, and recounts of dreams. We found genre-determined evidential use in most categories. Table 5 summarizes the most dominant form that emerged per genre.

Table 5 Dominant evidential form emerged across different genres

Genre	Bulgarian	Turkish
Bible and/or Quran	Direct	Direct
Modern Fiction	Direct	Direct
Fables, Folktales	Reportative	Indirect
Dreams	Direct	Direct
General knowledge	Historical Present	Assumption / Direct
Police reports	Indirect / Historical Present	Assumption / Generic
News	Indirect / Reportative / Historical Present	Direct

In both languages, the direct evidential was predominantly used in the modern fiction, religious, and dream categories, while the indirect (reportative in Bulgarian) was used in traditional fables. It is important to note that fables and religious texts are somewhat archaic, they do not necessarily point towards synchronous use of evidential forms but were important in understanding how direct

evidential in religious contexts, and indirect evidential in fables are traditionally used. The two languages differed strikingly in the News category, where Turkish used the direct evidential form virtually exclusively, while Bulgarian showed a relatively even distribution among all forms except the direct, namely: reportative, indirect, and historical present. This might be because epistemic uncertainty associates with Turkish indirect evidential (see Arslan 2020) possibly more strongly than in Bulgarian, and as a result, direct evidential is preferred to maintain an authority that the information provided is certain or truthful, while in Bulgarian this might not be necessary.

In the general knowledge category, Bulgarian uses solely historical present and does not use any evidentials whatsoever, while Turkish uses assumption/epistemic *–mIş+DIR* (48.6%) but also the direct evidential (31.4%). The two languages also differed in the reports category, where Bulgarian used both the indirect evidential (37.5%) and the historical present (53.1%), while Turkish used the epistemic form *–mIş+DIR*.

Several patterns emerge from these results. One is the strict ban of the direct evidential in Bulgarian in realistic genres like historical/ science facts and news, while the Turkish direct evidential was used as the main form in news and about a third of the historical category. This suggests a deeper semantic difference between the two direct markers despite their similar use in prototypical conversational scenarios.

The second pattern is a competition of non-evidential forms with evidential ones, such as the Bulgarian historical present form and the Turkish generic assumption marker *–mIş+DIR*): both languages used these forms in genres that by design represent information as factual and neutral, such as textbooks/general knowledge and police reports (though notably not in news). However, a limitation must be acknowledged here, namely the close relationship between evidential markers and modal meanings in both languages, particularly with respect to mirativity and epistemic meanings. In both the languages, indirect evidential forms can be used in mirative contexts or used to indicate uncertain information; however, based on written genres, analyzing such contexts pose a limitation. In certain genres, such as news reports, especially in Turkish, information seems to have been manipulated in a way that direct evidentials were exclusively used to remove any forms of epistemic uncertainty, an observation that fits in with Makartsev (2014). This observation merits more detailed exploration in future work.

6 Conclusions

This study shows that evidentials are restricted to different narrative modes with strikingly little optionality. We find that Bulgarian and Turkish pattern identically in four of the seven environments. The biggest differences were (i) in the use of the direct evidential in news and general knowledge in Turkish that was not attested at all in Bulgarian; and (ii) the lack of the indirect in Turkish in the report category compared to Bulgarian, where they were used on a par with the historical present.

To conclude, our findings highlight the need for more nuanced accounts of the relationship between evidentiality, epistemic modality, and discourse genre, as well as the competition between evidential and other narrative forms, such as the historical present and the assumption marker.

Acknowledgments

Bibliography

Abasıyanık, S.F. (2021). *Karanfiller ve domates suyu*. İstanbul: İş Bankası Kültür Yayınları.

Aikhenvald, A.Y. (2004). *Evidentiality*. Oxford, UK: Oxford University Press.

Aksu-Koç, A.; Slobin, D.I. (1986). "A Psychological Account of the Development and Use of Evidentials in Turkish". Wallace, C.; Nichols, J. (eds), *Evidentiality. The Linguistic Coding of Epistemiology*, 159-67. Norwood, NJ: Ablex Publishing.

Aksu-Koç, A.; Arslan, S. (2024). "Tense, Aspect, Modality and Evidentiality". Johanson, L. (eds), *The Encyclopedia of Turkic Languages and Linguistics Online, Section Psycholinguistics and Turkish*, 1-12. Leiden: Brill.

Alexander, R.; Zhobov, V. (2009). "New Conclusions on the Conclusive". *Journal of Slavic Linguistics*, 17(1-2), 61-85.

Arregui, A.; Rivero, M.L.; Salanova, A.P. (2017). "Aspect and Tense in Evidentials". Arregui, A.; Rivero, M.L.; Salanova, A.P. (eds), *Modality Across Syntactic Categories*, 211-34. Oxford: Oxford University Press.

Arslan, S. (2020). "When the Owner of Information is Unsure: Epistemic Uncertainty Influences Evidentiality Processing in Turkish". *Lingua*, 247, 102989.

Arslan, S. (2016). *Neurolinguistic and Psycholinguistic Investigations on Evidentiality in Turkish*. Groningen: University of Groningen.

Bosnalı, S.; Kahraman, C. (2015). "Moods and Modality in the Turkish and Japanese Dream Narration". *LILA´15 / Linguistics and Language Conference Proceedings*, 169-85. İstanbul: DAKAM.

Chafe, W.L.; Nichols, J. (1986). *Evidentiality: The Linguistic Coding of Epistemology.* Norwood, NJ: Ablex Publishing.

Fielder, G.E. (1995). "Narrative Perspective and the Bulgarian l-Participle". *The Slavic and East European Journal*, 39(4), 585-600.

Fitneva, S.A. (2001). "Epistemic Marking and Reliability Judgments: Evidence from Bulgarian". *Journal of Pragmatics*, 33(3), 401-20.

Friedman, V.A. (1999). "Proverbial Evidentiality: On the Gnomic Uses of the Category of Status in Languages of the Balkans and the Caucasus". *Mediterranean Language Review*, 11, 135-55.

Friedman, V.A. (2003). *Turkish in Macedonia and Beyond: Studies in Contact, Typology, and Other Phenomena in the Balkans and the Caucasus*. Wiesbaden: Harrassowitz Verlag.

Friedman, V.A. (2004). "Evidentiality, Modality, and Narrative in Macedonian and Other Balkan Languages". Friedman, V.A. (ed), *Macedonian Studies: Papers from the Fifth International Macedonian–North American Conference on Macedonian Studies*, 85-102. Columbus, OH: The Ohio State University.

Friedman, V.A. (2008). "The Typology of Balkan Evidentiality and Areal Linguistics". Tomić, O.M. (ed), *Balkan Syntax and Semantics*, 101-34. Amsterdam: John Benjamins.

Givón, T. (1982). "Logic vs. Pragmatics, with Human Language as the Referee: Toward an Empirically Viable Epistemology". *Journal of Pragmatics*, 6(2), 81-133.

Grannes, A. (1996). *Turco-Bulgarica: Articles in English and French Concerning Turkish Influence on Bulgarian*. Wiesbaden: Otto Harrassowitz Verlag.

Izvorski, R. (1997). "The Present Perfect as an Epistemic Modal". Lawson, A.; Cho, E. (eds), *Proceedings of SALT 7*, 222-39. Ithaca, NY: CLC Publications.

Johanson, L.; Utas, B. (2000). *Evidentials: Turkic, Iranian and Neighbouring Languages*. Berlin: Walter de Gruyter.

Johanson, L. (2003). "Evidentiality in Turkic". Aikhenvald, A.Y.; Dixon, R.M.W. (eds), *Studies in Evidentiality*, 273-90. Amsterdam: John Benjamins.

Koev, T. (2017). "Evidentiality, Learning Events and Spatiotemporal Distance: The View from Bulgarian". *Journal of Semantics*, 34(1), 1-41.

Kornfilt, J. (1997). *Turkish*. London: Routledge.

Korotkova, N. (2016). *Heterogeneity and Uniformity in the Evidential Domain* [Doctoral Dissertation]. Los Angeles, CA: University of California.

Korotkova, N. (2020). "The Subjective Heart of Evidentiality". Paper presented at *GLOW 43: Semantic Universals in the Modal and Attitudinal Domain*, Humboldt University & ZAS, Berlin, Germany.

Lewis, G.L. (1967). *Turkish Grammar*. Oxford: Clarendon Press.

Makartsev, M.M. (2014). *Эвиденциальность в пространстве балканского текста* (Evidentiality in the Balkan text space). Moscow: Russian Academy of Sciences.

Palmer, F.R. (2001). *Mood and modality*. Cambridge: Cambridge University Press.

Plungian, V.A. (2001). "The Place of Evidentiality within the Universal Grammatical Space". *Journal of Pragmatics*, 33(3), 349-57.

Şener, N. (2011). *Semantics and Pragmatics of Evidentials in Turkish* [Doctoral Dissertation]. Storrs, CT: University of Connecticut.

Sezer, E. (2001). "Finite Inflection in Turkish". Erguvanlı-Taylan, E. (ed.), *The Verb in Turkish*, 1-46. Amsterdam: John Benjamins.

Simeonova, V. (2020). *The Syntax and semantics of light attitudes* [Doctoral Dissertation]. Ottawa: University of Ottawa.

Simeonova, V. (2021). "Challenging Evidential Non-Challengeability". Paper presented at SemDial 25, University of Tübingen, Germany.

Slobin, D.I.; Aksu, A.A. (1982). "Tense, Aspect and Modality in the Use of the Turkish Evidential". P.J. Hopper (ed.), *Tense-Aspect: Between semantics and pragmatics,* 185-200. Amsterdam: John Benjamins.

Smirnova, A. (2011). *Evidentiality and Mood: Grammatical Expressions of Epistemic Modality in Bulgarian* [Doctoral Dissertation]. Columbus, OH: The Ohio State University.

Smirnova, A. (2013a). "Evidentiality in Bulgarian: Temporality, epistemic modality, and information source". *Journal of Semantics*, 30(4), 479-532.

Smirnova, A. (2013b). "The Meaning of the Bulgarian and Turkish evidentials". *Съпоставително езикознание / Сопоставительное языкознание*, 2-3, 205-23.

Smirnova, A. (2021). "Evidentiality in abductive reasoning: Experimental support for a modal analysis of evidentials". *Journal of Semantics*, 38(4), 531-70.

Sonnenhauser, B. (2013). "'Evidentiality' and point of view in Bulgarian." *Съпоставително езикознание / Сопоставительное языкознание*, 2-3, 110-30.

Tarpomanova, E. (2021). "Semantic readings of the indirect evidentials in Bulgarian: knowledge vs. reliability." *Papers of BAS. Humanities and Social Sciences*, 8(2), 125-33.

Willett, T. (1988). "A Cross-Linguistic Survey of the Grammaticization of Evidentiality". *Studies in Language*, 12(1), 51-97.

Yavaş, F. (1980). *On the Meaning of the Tense and Aspect Markers in Turkish* [Doctoral Dissertation]. Lawrence, KS: University of Kansas.

Università
Ca'Foscari
Venezia