*Yearbook of the Poznań Linguistic Meeting 8 (2025), pp. 67–93*
*DOI: 10.14746/yplm-2025-0004*

# Foreign lexical items in the lexicon of the Karaim language

Matyáš Foltýn
*Masaryk University, Brno*
*462357@mail.muni.cz*

## Abstract

The focus of this article is on a comprehensive etymological lexical analysis of the Karaim language as a whole and its dialects. While the practice of at least partially examining entire foreign elements within a specific language is relatively common nowadays (cf. Haspelmath & Tadmor 2009), Karaim has so far avoided this kind of comprehensive research with the literature focused around Zajączkowski (1959, 1961) and Jankowski (2013, 2015) providing papers focused either on diachronic development of specific loanwords or a basic general overview of the situation.

The article, despite presenting only preliminary results of the analysis, aims to provide a more in-depth examination of the historical development and synchronic composition of the Karaim lexicon. The results indicate a pattern of distribution of loanwords among semantic classes seen in other languages, confirm the expected distribution of borrowings among semantic fields, clearly point out the lexical differences between West and East Karaim, and also show that the composition follows a pattern concerning the adoption of Hebrew and Slavic loanwords similar to that observable in the lexicon of Yiddish (Foltýn 2022). Despite the preliminary character of the results, they offer direct insight into the inner workings of the language's lexicon.

**Keywords:** Karaim; lexicon; loanwords; language contact; lexicology.

## 1. Introduction

This paper presents preliminary results of a dictionary-data-based comprehensive lexical analysis of the Karaim language. The paper begins with an introductory section in which the reader gets familiarized with the geographical, genetical, and ethnical placement, brief historical context, internal divisions, and current synchronic situation of both the language and its speakers which

is then followed by a description of the utilized sources of data and methodology. After that, the various situations of linguistic contact that occurred during the development of Karaim are explained. Finally, the presentation of and elaboration on the results is provided in Section 5. This section includes the all-inclusive Karaim results as well as the figures for each dialect. The analysis is conducted in terms of lexemes, their distribution among semantic classes, and their integration in the phonological and morphological system of Karaim, and unique lexical roots. A brief overview of limitations is provided in the section right after.

## 2. The language and its speakers

Karaim is a member of the Turkic language family, within which it belongs to the Western subbranch of the Kipchak languages. The classification of the Turkic languages is shown in Figure 1.

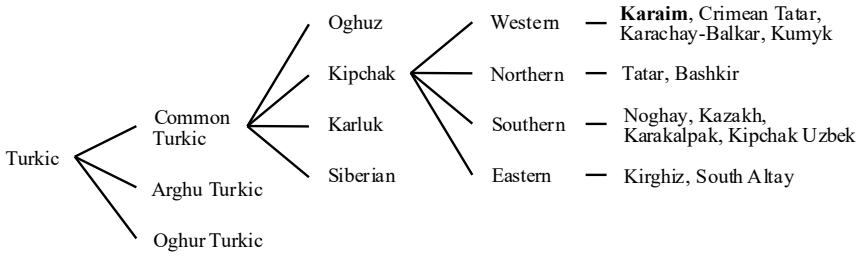


Figure 1. Position of Karaim within the classification of Turkic languages (based on Johanson & Csató 2022).

The Karaim language is considered a critically endangered one. It is spoken by about four dozen (Kocaoğlu 2006) out of the roughly one and a half thousand ethnic Karaims living in Eastern Europe according to national censuses.

The Karaims or Karays are an ethnoreligious community currently divided into three main Eastern European areas and several other locations outside Europe. Although speakers of a Turkic language, only a minimum of Karaims are ethnically Turkic, in fact, the genetic study conducted by Brook (2014) shows

that the vast majority is originally of Middle Eastern Jewish origin. That said, it is apt noting that the official Karaim narrative favors the Turkic Khazar hypothesis of their origin (see Section 2.2), as mentioned by Kizilov (2014).

For the sake of clarity, the followers of the Karaite Judaism who are not necessarily ethnic Karaims are referred to as Karaites. Karaite Judaism or Karaism is originally an old offshoot of Judaism, one of the smaller movements formed in opposition to the rabbinic Judaism in the second half of the 8th century that refuses to accept the oral tradition (see Section 2.2).

## 2.1. Current distribution

Currently, the Karaims can be found in three regions of Eastern Europe – the area adjacent to the Lithuanian town of Trakai, the Polish-Ukrainian historical region of Galicia and the region of Volhynia, and the Crimean Peninsula – and in various other regions outside Europe, such as Israel or the United States. Historically, the main regional cultural centers were the fortress of Chufut-Kale and Bakhchysarai, Lutsk and Halych, and Trakai and Panevėžys in the Crimea, Galicia-Volhynia, and Lithuania respectively. Among other important cities where the Karaim community was present, one can find Kaffa, Lviv, Kaunas, Pskov, Saint Petersburg, or the capitals of Kyiv, Warsaw, Vilnius, and Moscow, within most of which a kenesa, the Karaite equivalent of a synagogue, is located.

The situation is quite different when it is looked at in terms of the Karaim speakers. Out of these regions, the region of Galicia can be considered almost void since Csató (2001) already reports about six elderly speakers. The region of Crimea is then rather difficult to determine. Whereas, in 2014, 26 people claimed to speak Karaim in the local census, the Crimean community has historically been under a very strong cultural and linguistic influence of the Crimean Tatar majority which led to a high degree of linguistic assimilation, so the census data may say more about a sense of cultural affiliation than a linguistic one. Lastly, the Trakai region is the only region with at least a somewhat active speaker community which consists of somewhere around 25 people (Csató 2012; Jankowski 2024). Nonetheless, the majority of the Lithuanian Karaims should have at least passive knowledge of the language for liturgical purposes.


Figure 2. Map of distribution of Karaims in Eastern Europe.

## 2.2. Historical background

Since the history of the Karaims as an ethnic group is long preceded by the history of Karaite Judaism, the author finds it apt to begin this brief historical overview at that time.

The beginnings of Karaite Judaism are usually put in the 8th century Baghdad, where Anan ben David as a leader unified several Jewish anti-Talmudic sects under the name of Ananism that later gradually morphed into Karaite Judaism (Pełczyński 2004). This offshoot of Judaism then relatively rapidly gained followers that could by the start of the 11th century be found throughout the Byzantine Empire but also in the neighboring Egypt and Persia. Such an intense spreading of it later led to an ideological confrontation with followers of the rabbinic Judaism. The general attitude towards each other in the following centuries has, however, varied.

It is not entirely clear whence the Karaites came to the Crimean Peninsula. Several hypotheses have been proposed including the arrival of Jewish Karaites from Constantinople or alongside the Turkic armies, Karaims being descendants of the Turkic Khazars or of the Jews deported to Mesopotamia during the early times of the Persian empire. The Khazar hypothesis is especially favored by the Karaim scholars and community. (Harviainen 2003a; Kizilov 2014) The ethnogenesis of the Karaims likely began no later than at that time, i.e. the 13th century, by which they already adopted a Turkic language. Only after the fall of Constantinople did a notable number of the Byzantine Karaites join the Crimean community.

Before the Crimean Khanate turned into a vassal state of the Ottoman Empire in the mid-15th century, part of the Karaims had already left for Trakai and Volhynia (Tjahlyj 2007). According to tradition, the relocation to the Trakai occurred based on the impulse of the ruler of the Grand Duchy of Lithuania Vytautas the Great who fought the Golden Horde in the area of today's western Ukraine in 1397, shortly after followed by the relocation of another few hundred Karaims to Volhynia. The Karaims that later moved to Galicia only did so after about a century of Ottoman rule when the Grand Duchy of Lithuania turned into the Polish-Lithuanian Commonwealth. During the Ottoman rule, the Crimean Peninsula underwent quite severe Islamization and Turcization (Harviainen 2003a).

Until the annexation of Crimea by the Russian Empire in 1783 which put the Ottoman rule there to an end, European Karaite Judaism experienced something of a golden age during which Trakai became the main religious center (Akhiezer 2010). The town managed to retain this position even after the remigration of an appreciable number of Karaims following the famine on the peninsula. By the end of the century, almost all European Karaim communities were once again united under Russian rule except for the Galician community located within the Habsburg monarchy. The following two centuries brought official legislative dejudaization in the eyes of the Russian administrative apparatus based mostly on the academically controversial efforts of and proofs provided by Abraham Firkovich (Kohler & de Harkavy 1906), and continuation of economic growth of the Karaim. On the other hand, the start of the 20th century marks the slow end of Karaim religious and scientific literature written in the Hebrew language which was replaced by the secular literature written in Karaim (Kotler 1994).

After World War I, despite partial migration from Crimea to Central Europe (Kaleta 2015), the cultural life of Karaims bloomed again as new Karaim

literature was being published, both original and translated, and schools opened. This period of cultural boom ended with World War II during and after which the communities fell into steep decline. The Nazis occupied all of the main Karaim centers and although the occupying forces officially did not consider them to be Jews, also thanks to the huge effort of S. M. Shapshal, the Karaims in several cities were executed alongside the rabbinical Jews (Kizilov 2009). In addition to that, after the recapturing of Crimea by the Russian forces, some local Karaims got deported for collaboration with the Nazis to Central Asia along with the Crimean Tatars (Harviainen 2003b). The next half a century of Soviet rule which actively endorsed policies of cultural assimilation then dealt the final blow to the religious and cultural life of Karaims.

The dissolution of the Soviet Union at the end of 1991 came far too late for the Karaim language. Even though the kenesas in Trakai and Yevpatoria were renovated and restituted to the communities which led to a small religious reawakening, culturally, Karaims became mostly assimilated to the majority.

## 2.3. Dialectal differentiation

The formation of separate dialects began in the period of division of the original Crimean Karaim community in the 14th and 15th century and was later only strengthened due to geopolitical changes in the region.

The dialectal division is parallel to the geographical distribution of the Karaim presented in Section 2.1. There are three main dialects called the Trakai dialect, the Lutsk-Halych (L-H) dialect, and the Crimean dialect. The former two which are named after the regional cultural centers are usually clustered together under the label of West Karaim in order to put them into opposition to the Crimean dialect or East Karaim, which has been under a severe Oghuz and Crimean Tatar influence.

All three dialects should be fairly mutually intelligible with later borrowing-induced lexical differences most likely posing the biggest of problems. However, such problems could be, assuming the speaker has a good knowledge of the language, avoided by utilization of the original Turkic or common older borrowed synonyms that are still present in the language according to the dictionary data, e.g. the word ‘book’ is *ksiaşka* in Lutsk-Halych and *ksonşka* in Trakai (< Polish *książka*) but also *sefer* (< Hebrew סֵפֶר *séfer*) in Lutsk-Halych and most likely also other dialects, *kitab*, *kytab,* and *kitap* (< Arabic كِتَاب *kitāb*) present in all three as well as Turkic *bitik* (cf. Turkish *betik*).

The primary phonological differences between the dialects that are reflected in orthography are shown in (1).

| (1) | | Trakai | Lutsk-Halych | Crimea |
|---|---|---|---|---|
| | 'for life; forever' | *ömiurliuk* | *emirlik* | *ömürlik / ömirlik* |
| | 'writer' | *yazuvçu* | *yazuvtsu* | *yazıvçı* |
| | 'fast(ing)' | *oruç* | *oruts* | *orıç* |
| | 'bottle' | *şişe* | *sise* | *şişe* |
| | 'window' | - | *pendzere / pındzıra* | *pencere* |
| | " | *tereje / terjä / teräjä* | *tereze / teredze* | *terece / taraca* |
| | 'deer' | *olen' / olen* | *olen / oläп* | *ölen* |

The most phonologically divergent is the Lutsk-Halych dialect which shows a loss of front rounded vowels which are replaced by the front unrounded ones and a depalatalization of alveopalatal consonants in most positions as shown in the first five examples in (1) as well as by the example of *snur* 'rope' (< Polish *sznur*) but then also *şliaxta* 'nobility' (< Polish *szlachta*), *şem* 'name' (< Hebrew *שֵׁם shém*) among other exceptions (Németh 2015), The Trakai dialect noticeably differentiates between /e/ and /æ/ but the situation in Lutsk-Halych or the Crimean dialect is unknown to the author since the source does not distinguish between the two phonemes.

Phoneme /v/ is realized as a voiced labial fricative in West Karaim and as a voiced labial-velar approximant in East Karaim (Musayev 1964). The voiced velar nasal as well as the uvular pronunciation of unpalatalized voiceless velar occlusive are absent in West Karaim. The palatalization should affect a significantly wider range of consonants in the Trakai dialect than in the Crimean dialect. In contrast, the Lutsk-Halych dialect shows almost no palatalization except around the velar occlusives, as is shown in the Soviet Cyrillic orthography used in Baskakov et al. (1974), e.g. – for the sake of the illustrative nature, the scholarly transliteration of Cyrillic is utilized for these examples in lieu of the Common Turkic alphabet – Trakai *кеняшлянь- kjenjaşljan'-*, Crimean *кэнъэшлен- keñeşlen-*, and Lutsk-Halych *кенэслэн- kjeneslen-* 'to consult, confer' or Trakai *тильсизь til'siz'*, Crimean *тильсиз til'siz* and Lutsk-Halych *тилсиз tilsiz* 'mute' where the palatalization is indicated by the soft sign. Note that the <*е*>–<э> differentiation in Halych-Lutsk and Crimean

dialects does not indicate an identical situation, since the latter indicates the absence of palatalization in Halych-Lutsk but in the Crimean dialect it is to be assumed present as its absence would be indicated by a hard sign, cf. Lutsk-Halych *кылык kylyk* and Crimean *къылыкъ k"ylyk"* which would be rendered as *kılık* and *qılıq* 'character; manner' respectively in the Common Turkic alphabet.

The two dialect groupings generally present a somewhat different approach to vowel harmony. While West Karaim applies roundness-based vowel harmony alongside the front/back one consistently, even though in the Lutsk-Halych dialect the front allomorphs are non-existent because of the abovementioned loss of front rounded vowels, East Karaim lost or were in the process of losing this type of vowel harmony during the compilation of this study's sources, see the first three examples in (1). This loss is most likely due to the influence of Crimean Tatar where the roundness-base vowel harmony is also generally absent beyond the stem, cf. Crimean Tatar *yazıcı* and *ömürlik* (Kavitskaya 2010). The Crimean dialect, however, is the only one to apply front/back vowel harmony within the stem of newer loanwords, see the last example in (1) and the word for "ruble" which is *ruble* but also *rüble*.

## 2.4. Orthographies

The Karaim language was up until the end of the 19th century generally written in the Hebrew script. Afterward, the notation was different for each dialect and sometimes even also synchronically within a single dialect. Since the beginning of the 20th century, publications in Cyrillic started to appear. The shift, including the mandated one to the Soviet Latin orthography and later voluntary one to the Lithuanian- and Polish-based Latin orthography for Trakai and Lutsk-Halych respectively, was, however, quite slow and some Karaims kept on using the Hebrew script up until mid-Soviet times, if only for the religious texts (Olach 2016; Kocaoğlu 2006). In 1928, during the first language policy changes, all three dialects were once again united in that they – as well as the vast majority of minority languages in the USSR – were officially assigned a Latin-based orthography. This Soviet Latinization was a rather short-lasting decision overridden by the 1940s' extensive switch to the Cyrillic alphabet that only did not affect the Karaims living in the Polish part of Galicia. The specifics of each dialect's Cyrillic orthography were, however, slightly different, thus the difference in rendering the word *küçlü* 'strong' in Trakai Cyrillic

*ключлю – kjučlju* and Crimean *кӱчлӱ – küčlü* (Baskakov et al. 1974) Following the disintegration of the Soviet Union, the Lithuanian Karaims adopted a new Latin Lithuanian-based orthography (Kocaoğlu 2006).

## 3. Sources of data and methodology

The data analyzed here is dictionary-based and consists of one thousand and five hundred entries from Baskakov et al. (1974), a comprehensive Karaim–Russian–Polish dictionary. Since the original entries are provided in Cyrillic, the CTA transliteration provided here is produced by the author. The entry is then included in an Excel database and supplied with additional information including etymology provided by the dictionary, verified etymology, dialect, and translation, and assigned values for root, word class, and a semantic area.

The morphological and etymological analysis is conducted via the author's knowledge of the Slavic languages in question and the consultation of other entries within the dictionary, Arabic, Hebrew, and Persian dictionaries and Turkic etymological dictionaries. With consideration for the following steps of the research, finally, the entries are then clustered into four groups based on their etymology: 1) entries of Turkic origin; 2) borrowings inherited into the Karaim language – note that 1) and 2) will further be collectively labeled as 'inherited (lexical) material' (IM); 3) borrowings from Jewish liturgical languages (JLL), i.e. Hebrew and Aramaic; 4) borrowings from other languages, generally major languages that influenced the already separate Karaim language, i.e. the Slavic languages, Lithuanian, Crimean Tatar or Turkish – note that the latter two groups will further be collectively labeled and referred to as 'loanwords proper' to indicate they were borrowed after the genesis of the Karaim language. The less clear cases in the morphological analysis were also consulted with Jankowski & Aqtay (2015). A corresponding analysis method is then applied to plain lexical roots in Section 5.1. The analysis only concerns matter replication, i.e. morpheme transfer, and does not include cases of pattern replication, i.e. model reproduction (Sakel 2007; Matras 2009; Weinreich 1979).

Since the source includes lexemes from all three dialects and handily indicates the lexeme's affiliation to them, the data from each dialect is analyzed separately. As a result, data for each dialect consists of roughly five hundred entries.

### 3.1. Common Turkic alphabet

Given both the diachronic and synchronic heterogeneity of the Karaim orthographies, the Common Turkic alphabet (CTA) as established by the Turkic Council in Devlet (1992) is utilized throughout this study for all the Karaim examples. Examples from other Turkic languages are written in their respective official writing systems and orthographies.

## 4. Contact situations

Karaims have, as stated in the brief historical overview in Section 2.2, come in contact with several nations most of which were of the Turkic or Slavic genetic and linguistic background, and so, understandably, has their language that also already carried signs of previous foreign influence when they adopted it. Their long-lasting position as a minority has only added to the strength of the majority languages' influence. In addition to that, the religiocultural aspect of their lives also leads, likely during the genesis of the language, to an intragroup influence of the Jewish liturgical languages. Lastly, most of the Turkic world has been subjected to the Islamic expansion and even though Karaims never converted to Islam, the language has still been significantly influenced by both Arabic and Persian. This process occurred via another Turkic language, be it Ottoman Turkish or Crimean Tatar.

### 4.1. Pre-Karaim Turkic contact situations

The only language the data source indicated that could have entered Turkic before the formation of a separate Karaim language would be the Mongolian language. Given the historical closeness of the Turkic and the Mongolic nations and the debates on their linguistic genetic kinship within the proposed Altaic macrofamily which would put the contact between the 4th millennium BCE and the 9th century CE with later resumption in the earlier centuries of the 2nd millennium CE, it would be rather difficult to determine whether the specific lexeme is indeed a borrowing or a respective reflection of a common predecessor. However, the number of entries marked as and found to be of Mongolian origin is in the lower single digits throughout the source. Therefore,

this issue will not be dealt with in the course of this paper, since whatever the case may be, the results will not be noticeably affected.

Two of the concerned entries are verbs related to social relations and warfare, the rest are nouns. The entry *alanğasar* 'giant' provided in several variants in all three dialects seems to be of Udmurt origin. For a more in-depth analysis of specific borrowings, see Zajączkowski (1959).

Any contact with other nations in the area is not apparent from the lexical material.

### 4.2. Impact of Hebrew and Aramaic

As followers of Judaism, Karaims understandably knew, if not used in their daily lives, the Hebrew language. Upon the ultimate shifting to the language of the Turkic, or more specifically Kipchak, majority, Hebrew, at least for some time, continued to be used for liturgical purposes and so the Karaim still maintained some passive knowledge of the language. During both the process of shifting and subsequent early development of new ethno- and religiolect it influenced the new language rather significantly. A similar process can be seen in the development of Yiddish (see Jacobs 1994; Foltýn 2022), Judeo-Spanish (see Zucker 2001; Bunis 2011) and other Jewish diaspora languages (Kahn & Rubin 2016). Similarly to the other Jewish diaspora languages, the involvement of Aramaic in the process is considerably smaller than that of Hebrew. An overview of the Hebrew impact on Karaim can be found in Jankowski (2013).

While the loanwords from Hebrew generally relate to religion, education, intellect, social relations, and, lastly, crafts and trade which the Jews usually worked in, there is also a notable number of them in the semantic field of the physical world including stellar constellations and cardinal points, e.g. *adonay* 'Lord', *amen* 'amen', *avan* 'sin', *aron* 'Ark of the Covenant' (< Hebrew *אהדונהי* '*adonay*, *אמן 'amén*, *עוון avón*, *ארון 'arón* respectively). The impact of Aramaic is represented in the source only by a single entry related to religion, i.e. Trakai and the Lutsk-Halych *girsa* "Biblical text in the original language".

### 4.3. Mediated contact with Arabic and Persian

The contact situation involving mediation of contact with the Arabic and Persian languages can be divided into two phases.

The Islamization of the Turkic nations in Central Asia as well as the Caucasus took place in the 9th century CE which would mark the beginning of the first phase. At that time, it is not entirely clear where exactly the Karaims were situated. Nonetheless, it can be safely assumed that they were within the reach of this process, even though they have not submitted to it.

During and following this process the Turks adopted a plethora of Arabic and later in the time also Persian words related to various semantic fields. The Karaims who were living among them then accepted some of the loanwords related to the religious, social, scientific, political, administrative, and worldview concepts of the majority, e.g. *ayip* 'shame, guilt', *adet* 'custom, tradition', *aqıl* 'reason', *amanat* 'savings' (< Arabic عيب *ʕayb,* عادة *ʕāda,* عقل *ʕaql,* أمانة *ʔamāna* respectively). The Arabic loanwords from this phase are consequently heavily focused on these semantic fields with a special focus on the former three. The ones from Persian then show a slightly wider span including the previously mentioned, although the categories mentioned as primary are not in the lead anymore, as well as a number of newly introduced tools, animals and plants related to a change of lifestyle of which there are much fewer in the former case, e.g. *baxt* 'luck', *axır* 'stable', *zincir* 'chain', *afta* 'week', *bağ* 'orchard', *samur* 'sable (animal)', *zerdeli* 'apricot', *xıyar* 'cucumber' (< Persian بخت *baxt*, آخور *âxor*, زنجیر *zanjir*, هفته *hafte*, باغ *bağ*, سمور *samôr*, زردآلو *zard-âlu*, خیار *xiyâr*).

The second phase can then be delimited by the Crimean Khanate's suzerainty of the Ottoman Empire, that is from mid-15th to late 18th century CE. This phase mostly concerns East Karaim, since the speakers of West Karaim left the area earlier in the 14th century. During the Ottoman rule, the Ottoman Turkish language had a strong impact on Crimean Tatar and Karaim when it itself being heavily influenced by Arabic, usually by the proxy of Persian, and Persian loaned a great deal of words of this origin either directly to Karaim or via Crimean Tatar.

The loanwords that entered the language this way are usually but not exclusively from the field of administration and warfare accompanied by a fair number of words from all over the semantic field range that most likely entered via Crimean Tatar when Ottoman Turkish became more common in everyday life, e.g. *askär* 'soldier' (ultimately < Persian لشکر *laškar*), *sikke* 'coin', *xarc* 'expense' (< Arabic سكة *sikka*, خرج *xarj* resp.), *şahar* 'city' (< Persian شهر *šahr*).

For a diachronic analysis of a selection of specific loanwords from this contact situation see Zajączkowski (1961).

## 4.4. Early contact situations in Crimea

The early contact situations in Crimea are already partially covered in the previous section (4.3). This period can be similarly once again divided into two phases, that is one from the arrival of the Karaims to the Crimean Peninsula sometime in the 10th century up to the 14th or 15th century when part of the Karaims left for Galicia and Lithuania and one right after that lasted until the end of Crimean Khanate in the late 18th century.

The former phase covers the initial contact with different Kipchak varieties speakers of which arrived in the area with the Karaims, or the Karaims arrived with them to be more precise, as well as the later emerging languages of Crimean Tatar and to a certain extent also Noghay. Alongside these Turkic languages, there were also diasporic minorities of Italians and Greeks whom the speakers of Karaim also encountered.

This phase has not caused any noteworthy amount of impact on the lexicon of Karaim. The source offers only a handful of loanwords from Italian and Greek that are likely from this period related to a sedentary lifestyle, e.g. *mandra* 'sheepfold' (< Italian *mandria* 'herd' or Greek *μάνδρᾱ mándrā* 'cattle enclosure'), new natural phenomena, sailing, and seaside. However, the majority of Greek loanwords entered Karaim via Turkish in the following phase. Upon consultation, Jankowski & Aqtay (2015) also offer a small number of "imitations of a Noghay form" that generally involve only a depalatalization of alveopalatal voiceless fricative. It is not entirely clear whether these forms should be considered loanwords or simply a sign of Noghay influence on these specific words, therefore they are not counted as such for the purpose of this paper.

The latter phase that only applies to East Karaim is then parallel to the second phase described in the previous section, so for a more detailed context see Section 4.3. As discussed above, the Ottoman Turkish language influenced both of the Turkic languages of Crimea quite heavily. Besides the abovementioned mediated loanwords from Arabic and Persian, Turkish also loaned several of its own words and mediated a variety of loanwords from Greek. A few Turkish loanwords most likely entered Karaim via Crimean Tatar, the local majority's language that itself heavily influenced East Karaim. The source does not indicate the ultimately Turkish or Crimean Tatar borrowings as such.

The Turkish loanwords semantically cover a wide range of fields and usually exist parallelly to a formally very close Karaim word, e.g. the word 'lion' occurs in the original Karaim form *arslan* and the form *aslan* of the Turkish

origin. The mediated Greek loanwords are mostly focused on new materials, e.g. *abonos* 'ebony' (< Turkish *abanoz* < Greek *ἔβενος ébenos*).

While the process of ultimate near-convergence of East Karaim and Crimean Tatar has possibly already begun by the end of this period, it quite surely was not due to lexical borrowings of which there are lower double digits. This subcategory of loanwords proper mostly includes conjunctions and adverbs in the data source. However, they are not included in the preliminary subset of this paper. The process is thus more likely a result of phonological and structural assimilation within the system of the language.

### 4.5. Early contact with Polish and Lithuanian

The period of contact between West Karaim and Polish and Lithuanian began with the relocation of part of the Karaims from the Crimean peninsula to Galicia and Trakai or Troki in Polish, both within the Polish-Lithuanian Commonwealth of which the official language was the Polish language, in the 14th and 15th century and continues mainly up to the 18th century when the Commonwealth crumbles, although the usage of Polish and Lithuanian in the general populace in the area was for quite some time likely not too linguistically affected. Incidentally, sometime during this and the following contact situation, several words of French origin, a language popular among the nobility at the time, entered the Karaim lexicon.

Within the preliminary subset, the loanwords from Polish that by far occur the most frequently in the Trakai dialect primarily belong to the semantic classes of conjunctions and particles, e.g. *aj* 'until; even' (< Polish *aż*), *ale* 'but' (< Polish *ale*) or *ani* 'neither' (< Polish *ani*). The source, however, contains a fair number of nouns from a wide scope of semantic fields that are also of Polish origin, e.g. animal, plant, and material names like *ğuba* or *guba* 'tree fungus' (< Polish *huba*). However, some of the lexemes marked as such might be of Belarusian or Ukrainian origin instead. There are also over a hundred entries that are marked as "(Common) Slavic" in the source for which it is difficult, if not impossible, to determine a single Slavic language of origin. The minority of them that can be further specified does not seem to be of Polish but Belarusian or Ukrainian origin. The Lithuanian language only accounts for two loanwords within the whole source, i.e. *ğrıbus* 'mushroom' (< Lithuanian *grybas*) and *şaka* 'branch' (< Lithuanian *šaka*).

## 4.6. Later contact with the East Slavic languages, Polish, Lithuanian, and others

The final contact situation in the development of the Karaim language can then once again be divided into two phases, that is 1) the pre-Soviet period and 2) the Soviet (and post-Soviet) period.

The first phase began in the late 18th century when the Russian Empire conquered the Crimean Khanate and the Polish-Lithuanian Commonwealth disintegrated with parts of it falling under the Russian Empire, the Habsburg Monarchy, and Prussia, i.e. the later established German Empire. This geopolitical situation indicates the period of the most intense isolation of the Karaim communities from each other with the Crimeans, Volhynians, and part of the Galicians being under Russian rule, the other part of the Galicians being under the Habsburg rule, and the Lithuanian Karaim being under the Prussian, that is German rule.

Whereas the Crimean dialect has understandably continued to be influenced by the Russian and the Ukrainian languages as well as the Crimean Tatar that now similarly occupies a place of, even if still a diametrically bigger one, a minority language, there is a remarkable lack of any influence of German on the remaining two dialects. While the source includes German in the list of abbreviations among other languages of origin, the author of this paper has not managed to find any single entry marked that way. Conclusively, if there is any influence of the German language on West Karaim, it is such a small number of lexemes that it would not be statistically significant.

The second phase starts in 1922 with the establishment of the USSR and can then be further divided into the Soviet and post-Soviet periods. The first period is marked by the severe Sovietization- and Russification-oriented policies. Such policies only strengthened the position of the foreign component in the lexicon of Karaim which was once again for the most part united under the rule of one government. Consequently, the Sovietization of the nations also led to a decline in the usage of the Karaim language which was often only maintained up until then for religiocultural reasons. During this period, one would expect an influx of new words related to the political system, industrialization, and new scientific discoveries. Unfortunately, neither Baskakov et al. (1974) nor Jankowski & Aqtay (2015) offer this type of loanwords. Nevertheless, based on the situation in other Turkic languages of the former Soviet Union and especially Krymchak, a fellow moribund Kipchak Turkic variety sometimes considered a religiolect or ethnolect of Crimean Tatar that is spoken by

the Turkic Crimean followers of rabbinic Judaism. Rebi (2004), although this type of loanwords is also not included in the body part of this dictionary, offers a brief insight into this matter in several of the provided conversational phrases, e.g. *Nomerdä telefon işliyım?* 'Does the phone in the [hotel] room work?', *Sız başqa avtobusqa otırdıñız.* 'You boarded the wrong bus.', *Poyezd qaç daqadan kıtecek?* 'In how many minutes does the train leave?', *Samolët nä vaxıt uçacax?* 'When does the plane depart?' or related things such as *Yox, biletım daa yox*. 'No, I do not have a ticket yet.' among others. Based on that, it can be rather safely assumed that Karaim is in a similar position unless these Russian loanwords have since been replaced by equivalent loanwords from Polish or Lithuanian.

The second period, after 1991 was then a period of linguistic endangerment of the Karaim language with it not being actively used in everyday life. At this time, a noticeable amount of code-copying can be observed even among the several dozens of elderly speakers of the most alive Trakai dialect (Csató 2001) with a shift to the language of the majority among the younger Karaims. This period, as well as the latter half of the first period, is not going to be dealt with here due to the age of the source material, that is 1974. A brief analysis of several Belarusian loanwords can be found in Wexler (1980).

## 5. Etymological division of the lexicon

There is a total number of 1509 lexemes analyzed in this paper. Out of this number, 1195 or 79.19% belong to the original Turkic base of the Karaim language, 242 or 16.04% are lexemes borrowed before or during the genesis of the language that do not originate in Jewish liturgical languages, 40 or 2.65% can be considered loanwords proper as they were certainly borrowed after the formation of the Karaim language – note that only 11 or 0.73% is of other than Turkish origin – and 32 or 2.12% is of Hebrew or Aramaic origin. Consequently, the total for inherited material amounts to 1437 or 95.23% and the total loanwords proper amount to 72 or 4.77%.

While it might come as surprising that the group of loanwords from JLL is in total almost three times as numerous as the group of "recent" loanwords from the local major languages, that seems to be a trend also exhibited by other Jewish diaspora languages, e.g. the proportional situation in Yiddish is 81-6-13 and 68-12-20 in lexemes and roots for Germanic, Slavic and Semitic

respectively (Foltýn 2022). Nevertheless, both numbers are surprisingly low in the Karaim case. For the context on the categorization see Sections 3 and 4.

Table 1. Etymological categorization of entries per dialect (percentages).

| Component | Trakai | L-H | Crimea |
|---|---|---|---|
| Turkic base | 79.86 | 80.76 | 77.31 |
| Inherited loanwords | 15.10 | 15.03 | 17.63 |
| Loanwords | 1.83 | 1.60 | 4.19 |
| Loanwords from JLL | 3.20 | 2.61 | 0.87 |
| Total IM | 94.97 | 95.79 | 94.94 |
| Total LWs proper | 5.03 | 4.21 | 5.06 |

The results of a separate analysis of each dialect are presented in Table 1. The results show that in the borders of the preliminary dataset, both dialects within the West Karaim grouping, of which the Trakai data consists of 437 entries and the Lutsk-Halych data of 499 entries, exhibit a very similar amount of foreign influence on the lexicon, although the figures for the Trakai dialect are slightly higher. In both dialects, the Turkic base adds up to roughly 80% and the inherited loanwords then account for another approximately 15% without any significant difference in the composition with Arabic loanwords being the most represented, then the Persian and, finally, in single digits Mongolian and Greek. The loanwords proper in both cases make up between 1.5% and 2% and they are exclusively of Slavic origin. However, there is a split in the composition. Whereas the Trakai-dialect data includes mostly loanwords of Polish origin, the data for the Lutsk-Halych dialect show per indications in the source publication Russian as the primary Slavic contributor. The Ukrainian language, however, seems as a more likely source of the words shared among the Slavic languages given the geographical distribution and historical context. The category of loanwords from Jewish diaspora languages then presents another slight split with the Trakai dialect being at over 3% and the Lutsk-Halych dialect being closer to 2.5%. The overall inherited material then constitutes roughly 95% and 96% and the borrowed material constitutes approximately 5% and 4% of the lexical subset for the Trakai and the Lutsk-Halych dialect respectively. As for East Karaim, the subset of data for the Crimean dialect

consisting of 573 entries significantly differs from the previous two. While the Turkic base could be said to at over 77% roughly constitute a similar portion of the lexicon as was the case in the two formerly described dialects, the distribution of loanwords within the subcategories is dissimilar. The inherited loanwords constitute a larger part of the lexicon by about 2.5% and composition-wise lean slightly more towards Arabic over Persian. The categories comprising the "total loanwords proper" category are especially noteworthy. Once again, the loanwords proper constitute a larger part of the lexicon by approximately 2.5%. Having said that, it is to be noted that only 0.35% is of other than Turkish origin, i.e. Slavic, which means that the vast majority or 3.84% of these loanwords come from Turkish that West Karaim was out of reach of. Lastly, the impact of the Jewish liturgical languages is less than 1%. This may be quite possibly attributed to the recurrent geopolitical isolation of the Crimean Peninsula from the rest of the Jewish world as well as the close coexistence with the local larger Muslim nation. Consequently, minding the different composition, the proportions of inherited material versus borrowed material are almost identical to that of the Trakai dialect.

To summarize, the Trakai and the Lutsk-Halych dialects are almost identical when it comes to lexical composition with only two notable differences, that is 1) the Slavic loanwords in the Lutsk-Halych dialect are mostly traceable to Russian, while the ones in the Trakai dialect generally originate in Polish, and 2) the Trakai dialect exhibits higher level of Hebrew influence. The Crimean dialect then shows a remarkably low degree of Slavic and Hebrew influence but displays a serious impact of Turkish both through the mediated Perso-Arabic loanwords and Turkish loanwords proper.

A brief description regarding the distribution of the loanwords is provided in Section 4.

### 5.1. Distribution by semantic class

The following section focuses on the distribution of loanwords, with a special focus on total loanwords proper, among the semantic classes of nouns, adjectives, verbs, adverbs, and an aggregate category of function words which includes pre- or postpositions, conjunctions, and particles. While the classes of pronouns and numerals are not included, it only seems appropriate to mention that there have been no non-Turkic lexemes found in either of these classes. However, two variations of a borrowing from Turkish present in the Crimean

dialect do fall into the category of indefinite numerals, namely *azbuçıq* and *azbuçuq* "a few". Although the category of interjections exhibits an interesting amount of foreign influence across all etymological sources, they are somewhat outside the scope of this research and therefore also omitted. Once again, the loanwords of ultimately Turkish origin have been discarded since they often exist parallelly to an almost identical Karaim lexeme.

In concordance with the general tendencies of interlanguage borrowing found in Haspelmath & Tadmor (2009), the majority of loanwords are nouns at 185 and 12.26% with total loanwords proper accounting for 31 or 2.05% of them. The amount of borrowed adjectives and verbs then follows in a much smaller number of instances at 51 entries or 3.38% and 41 or 2.27% with total loanwords proper being responsible for only 2 or 0.13% per each. The semantic class that has by far been penetrated the least is the adverbs out of which only 9 or 0.60% is of foreign origin and only 1 or 0.07% is a loanword proper. An interesting case could be made for the class of function words that includes a total of 16 or 1.06% of borrowings with 9 or 0.60% being a loanword proper. The relatively high number of loanwords can be attributed to the fact that it is an aggregate class, however, it is also the only class where the entries affiliated with the category of total loanwords proper constitute a significant part, over a half even. The high borrowability of function words can also be seen in the case studies included in Haspelmath & Tadmor (2009), as well as the preliminary results of a lexical analysis of Yiddish in Foltýn (2022).

The distribution of the components contributing to the lexicon of each Karaim dialect in terms of semantic class is shown in Tables 3 to 5.

Table 3. Distribution of loanwords in Trakai dialect by semantic class (percentages).

| Component | Nouns | Adj. | Verbs | Adv. | FW |
|---|---|---|---|---|---|
| Turkic base | 76.73 | 82.56 | 88.24 | 96.43 | 45.00 |
| Inherited loanwords | 17.33 | 15.12 | 11.76 | 3.57 | 15.00 |
| Loanwords | 0.00 | 0.00 | 0.00 | 0.00 | 25.00 |
| Loanwords from JLL | 5.94 | 2.33 | 0.00 | 0.00 | 15.00 |
| Total IM | 94.06 | 97.68 | 100.00 | 100.00 | 60.00 |
| Total LWs proper | 5.94 | 2.33 | 0.00 | 0.00 | 40.00 |

In terms of total loanwords proper, the data for the Trakai dialect exhibits a surprising amount of borrowing within the class of function words which are generally of Slavic origin. While the absence of borrowed verbs is expectable, the absence of adverbs is not and might be caused by the problems described in Section 5.2.

Table 4. Distribution of loanwords in Lutsk-Halych dialect by semantic class (percentages).

| Component | Nouns | Adj. | Verbs | Adv. | FW |
|---|---|---|---|---|---|
| Turkic base | 76.57 | 79.17 | 89.23 | 93.75 | 90.91 |
| Inherited loanwords | 17.99 | 20.83 | 9.23 | 6.25 | 0.00 |
| Loanwords | 1.26 | 0.00 | 0.77 | 0.00 | 0.00 |
| Loanwords from JLL | 4.18 | 0.00 | 0.77 | 0.00 | 9.09 |
| Total IM | 94.56 | 100.00 | 98.46 | 100.00 | 90.91 |
| Total LWs proper | 5.44 | 0.00 | 1.54 | 0.00 | 9.09 |

The data for the Lutsk-Halych dialect does not include any borrowed adjectives or adverbs. However, it is interesting in that there is a borrowed verb from each of the components contributing to total loanwords proper, i.e. the Slavic *arıştovat'*- 'to arrest' (< Polish *aresztować*) ultimately of Latin origin and *avanla*- 'to do evil' derived from the Hebrew noun *עָוֹן avón* 'an intentional sin' which is not included in the source. Curiously, the source includes another verb of ultimately different origin that entered Karaim via Slavic for the Lutsk-Halych dialect, that is *voyajutt'*- [sic] 'to travel' (< Polish *wojażować* < French *voyage* 'travel') which is not included in the subset. Another figure of significance is the total absence of inherited loanwords in the category of function words.

The results for the Crimean dialect reflect the abovementioned situation of it being strongly influenced by neither the Slavic nor Jewish liturgical languages. Nonetheless, while the absence of any loanwords proper in most of the classes outside of nouns is expected, the presence of an adverb of Hebrew origin presents a point of curiosity. The fact, that it came into existence as a result of the intralanguage derivational process renders it rather insignificant though.

Table 5. Distribution of loanwords in Crimean dialect by semantic class (percentages).

| Component | Nouns | Adj. | Verbs | Adv. | FW |
|---|---|---|---|---|---|
| Turkic base | 72.79 | 88.79 | 90.65 | 91.84 | 71.43 |
| Inherited loanwords | 25.09 | 11.21 | 9.35 | 6.12 | 28.57 |
| Loanwords | 0.71 | 0.00 | 0.00 | 0.00 | 0.00 |
| Loanwords from JLL | 1.41 | 0.00 | 0.00 | 2.04 | 0.00 |
| Total IM | 97.88 | 100.00 | 100.00 | 97.96 | 100.00 |
| Total LWs proper | 2.12 | 0.00 | 0.00 | 2.04 | 0.00 |

## 5.2. Lexical roots

The analysis of unique lexical roots was conducted to avoid certain limitations of the source and the dataset. While the results arisen from the analysis of lexemes also handily present the various degrees of integration of certain categories of loanwords into the derivational system and the figures for a truly used lexical material in traceable textual sources, they do not genuinely represent the actual amount of foreign material in the lexicon of the language or the dialect. In consequence, both the source and the dataset are not entirely consistent in when the derived forms get included as a separate entry and when they do not, most likely due to the lack of presence of a specific form in the textual sources utilized as a base for Baskakov et al. (1974). This, however, leads to such situations as the word for 'journal, magazine', which is derived from a loanword of Arabic origin meaning 'time', i.e. *vaxt* in the Trakai and the Lutsk-Halych dialects and *vaxıt* in the Crimean dialect, being included in all three forms as *vaxtlıx*, *vaxtlık*, and *vaxıtlıq* respectively, while the adjective 'related to some specific time' derived from the same original loanword, i.e. *vaxtlı*, is only mentioned in relation with the Trakai dialect, despite the derivational suffix *-lı* being perfectly productive in all three dialects which thus would in all likelihood have no objections towards such derivation. The loanwords of ultimately Turkish or Crimean Tatar origin are not considered here since they usually mirror an identical Turkic lexical root already present within the lexical base.

There is a total of 440 unique lexical roots, divided into 121, 127, and 192 entries for the three dialects respectively. Overall, the Turkic base accounts for

roughly two-thirds of the data, the category of inherited loanwords accounts for approximately a quarter, the loanwords proper form about one-fiftieth, and the loanwords from the Jewish liturgical languages about one-twenty-fifth. The total for the inherited material is then 411 entries or 93.4% with the total loanwords proper accounting for the remaining 39 entries or 6.6%.

Table 6. Etymological categorization of unique lexical roots per dialect (percentages).

| Component | Trakai | L-H | Crimea |
|---|---|---|---|
| Turkic base | 67.77 | 66.14 | 68.23 |
| Inherited loanwords | 22.31 | 23.62 | 29.69 |
| Loanwords | 3.31 | 3.15 | 1.04 |
| Loanwords from JLL | 6.61 | 7.09 | 1.04 |
| Total IM | 90.08 | 89.76 | 97.92 |
| Total LWs proper | 9.92 | 10.24 | 2.08 |

The results of the lexical-root analysis for each dialect are presented in Table 6. In comparison to the results of the lexical analysis in Section 5., the general patterns seem to be identical with the remarkable exception of the Lutsk-Halych data. Both of the West Karaim dialects are still significantly closer to each other than to East Karaim. However, the results here indicate the Lutsk-Halych dialect to be more impacted by foreign influence than the Trakai dialect, bar the loanwords proper category in which the latter still leads, even if only by a small difference. Across the board, the base forms about two-thirds of the unique roots. While the Lutsk-Halych dialect is the most and the Crimean dialect the least overall impacted by foreign influence in general terms, the difference is only roughly 2%. The inherited loanwords form about 23% and 30% of the data in West and East Karaim respectively with identical dominance of loanwords of Arabic origin over Persian as in the general lexical analysis. The total inherited material then accounts for up to 90% and 98% of the lexical stock of West and East Karaim. With the absence of Turkic loanwords, the other two categories of loanwords proper at roughly 3% and 1% and JLL loanwords at approximately 7% and 1% for West and East Karaim respectively confirm the previously found pattern of West Karaim being more

influenced by both the Slavic majority and the original Jewish languages than East Karaim. Consequently, the figures for the total number of loanwords proper are roughly 10% for the Trakai and the Lutsk-Halych dialects and 2% for the Crimean one.

## 5.3. Integration of loanwords

The integration of borrowed lexical material into the phonological system of Karaim varies depending on the age of the borrowing (for Slavic influence on Karaim phonology, see Németh 2021). Whereas the Perso-Arabic loanwords are generally modified to adhere to the rules of intra-stem vocalic harmony, the newer loanwords are usually not. The Crimean dialect is an exception in this case as is shown by the examples of Slavic borrowings adapted to vowel harmony in Section 2.3 in (1) and in the text. They all, however, adhere to the extra-stem vowel harmony. An interesting aspect of the adaptation of Perso-Arabic loanwords is that in a variety of cases, a shift of vowels from back to front vowel harmony can be observed. Some of the cases can be attributed to neighboring palatalized consonants, some are the result of Turkish influence, e.g. *amanat* and *emanet* 'reserve' in West and East Karaim respectively (< Turkish *emanet* < Arabic *ʔamāna* 'deposit'). The suffixes adopt the vowel harmony according to the last syllable of the word.

Morphologically, as is visible from the differences in the analysis of entries and isolated roots, nearly all foreign roots are completely incorporated in both the derivational and also the inflectional system – although the newer Baltic and Slavic roots take part in the former quite rarely, according to the source, e.g. Turkic derivational suffix *-li* with roots of Arabic origin, i.e. *ayipli* 'shameful', *aqıllı* 'wise', Persian origin, i.e. *azatlı* 'free', verbal suffixes *-la(n/t)-* with roots of Arabic, Persian and Hebrew origin respectively, i.e. *arlan* 'to be ashamed', *asayışlan* 'to delight, revel', *avanla-* 'to do evil', or with inflectional suffixes *gufum* 'my body' (< Hebrew גוף *guf*) and *krivdam* 'my harm' (< Polish *krzywda*) with 1p. possessive *-(u)m*, *dorlar* 'generations' (< Hebrew דור *dor*) with plural suffix *-lar*, *kurdu çatırnı* '(s/he) built the tent' (< Persian چادر *čadur*) with accusative suffix, *arfanın* 'of the harp' (< Russian *ápфa árfa*) with genitive suffix, with only very few exceptions of borrowings being used as a morphologically frozen adverbial complement, e.g. Lutsk-Halych *zabavu et-* 'to play; to have fun' (< Ukrainian *заба́ва zabáva* 'fun') where the original form in the accusative case is fixed for this verb with the

literal translation of 'to do fun'. At the same time, the noun *zabava* 'game; fun' of the identical origin borrowed in nominative functions in the morphological system without issues. The extent of morphological integration goes as far as being used in partial calquing, e.g. the Trakai *vağalar* 'scales' (< Polish *waga*) being fixed with the plural suffix as plurale tantum in concordance to the word in other local languages, be it Russian *весы́ vesý*, Ukrainian *ва́ги váhy* or Lithuanian *svarstyklės*.

## 6. Limitations

The results presented in this paper have several limitations related to the source of the data and the preliminary nature of them.

Several of the problems with Baskakov et al. (1974) have been voiced during the course of this paper – no mention of loanwords from other Turkic languages in Section 4.4, the lack of modernization in Section 4.6, and the lack of consistency in terms of inclusion of derived forms in Section 5.2. Additional issue regarding the incorrectness of some of the provided etymologies in the source or intentional omission of Hebrew and Slavic words is discussed in Altbauer (1980). The author is aware of these limitations and describes the reasoning behind the procedural steps taken to minimize their impact on the results, if there are any that require being taken, in the respective sections except for the lack of modernization that the author is not capable of reliably covering at this moment.

To ensure the most objectively valid results, the methodology still requires some adjustments and improvements, and the subset database calls for expansion. An additional detailed analysis of the Perso-Arabic and the Mongolic element needs to be conducted in order to determine what portion belongs to the pre-Karaim and the Karaim phases, as well as a further consultation of additional sources such as Jankowski & Aqtay (2015), Smętek (2016), Németh (2020), Sulimowicz-Keruth (2022), among others.

## 7. Conclusion

In conclusion, the author is aware that the data presented here is still quite limited and finds it thus appropriate to remain careful in any generalizing statements. However, the results for the subset show a clear lexical division

between West and East Karaim. The lexicon of the West Karaim dialects includes less Perso-Arabic and more Slavic and Hebrew loanwords than that of the Crimean dialect. Despite the gap in the adoption of borrowings of JLL origin, all three dialects follow the pattern displayed also in Yiddish and quite possibly other Jewish diaspora languages in which the borrowings of JLL origin form a noticeably larger portion of the lexicon that the ones originating in the language(s) of the majority.

The semantic distribution both in terms of semantic fields and classes adheres to the contextual expectations and the general patterns of lexical linguistic contact exhibited in other published case studies, cf. Haspelmath & Tadmor (2009).

For any definitive statements, further research including a focused examination of specific elements is required.

## Acknowledgment

This work was supported by the Masaryk University doctoral scholarship (projects MUNI/A/1093/2022 and MUNI/A/1399/2023).

## References

Akhiezer, Golda. 2010. Karaites. In: *YIVO Encyclopedia of Jews in Eastern Europe*. https://yivoencyclopedia.org/article.aspx/Karaites (accessed 2021, January 23)

Altbauer, Moshé. 1980. O tendencjach dehebraizacji leksyki karaimskiej i ich wynikach w „Słowniku karaimsko-rosyjsko-hebrajskim" [On the de-Hebraization tendencies in the Karaim lexicon and their results in the "Karaim-Russian-Hebrew dictionary"]. *Harvard Ukrainian Studies* 3/4(1). 51–60.

Baskakov, Nikolaj A., Seraja M. Shapshal & Ananiasz Zajączkowski. 1974. *Karaimsko-russko-pol'skij slovar'* [A Karaim-Russian-Polish dictionary]. Moscow: Russkiy jazyk.

Brook, Kevin A. 2014. The genetics of Crimean Karaites. *Karadeniz Araştırmaları* 42. 69–84.

Bunis, David M. 2011. Judezmo: The Jewish language of the Ottoman Sephardim. *European Judaism: A Journal for the New Europe* 44(1). 22–35.

Csató, Éva Á. 2001. Syntactic code-copying in Karaim. In: Östen Dahl & Maria Koptjevskaja-Tamm (eds), *Circum-Baltic languages*. Volume I: *Past and Present*, 271–284. Amsterdam: John Benjamins.

Csató, Éva Á. 2011. Litvanya Karaycası. *JofEL* 1(1). 33–45.

Devlet, Nadir. 1992. *Milletlerarası Çağdaş Türk Alfabeleri Sempozyumu: 18–20 Kasım 1991* [International Contemporary Turkish Alphabets Symposium: 18–20 November 1991]. Istanbul: Marmara Üniversitesi. Türkiyat Araştırmaları Enstitüsü.

Foltýn, Matyáš. 2022. Foreign lexical items in the lexicon of modern Yiddish. Paper presented at the Small Languages, Big Ideas [SLBI] conference, Oldenburg (DE), 30 Jun 2022.

Harviainen, Tapani. 2003a. The Karaites in Eastern Europe and the Crimea: An Overview. In Meira Polliack (ed.), *Karaite Judaism. A guide to its history and literary sources*, 633–656. Leiden: Brill.

Harviainen, Tapani. 2003b. The Karaites in contemporary Lithuania. In: Meira Polliack (ed.), *Karaite Judaism. A guide to its history and literary sources*, 827–854. Leiden: Brill.

Haspelmath, Martin & Uri Tadmor (eds). 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin: Mouton de Gruyter.

Jacobs, Neil G. et al. 1994. Yiddish. In: Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 388–419. London: Routledge.

Jankowski, Henryk. 2013. Karaim, Hebrew component in. In: Geoffrey Khan et al. (eds), *Encyclopedia of Hebrew language and linguistics*, vol. 2, 444–447. Leiden: Brill.

Jankowski, Henryk. 2024. Karaim language studies – A historical overview. *Vilnius University Proceedings* 48. 19–49.

Jankowski, Henryk & Gülay Aqtay. 2015. *A Crimean Karaim-English dictionary: 10 000 entries*. Poznań: Adam Mickiewicz University, Department of Asian Studies.

Johanson, Lars & Éva Á. Csató (eds). 2022. *The Turkic languages* (2nd edn.). London: Routledge.

Kahn, Lily & Aaron D. Rubin (eds). 2016. *Handbook of Jewish languages*. Leiden: Brill.

Kaleta, Petr. 2015. *Tajemné etnikum z Krymu: osudy příslušníků karaimské emigrace do meziválečného Československa* [The mysterious ethnic group from Crimea: the fates of members of the Karaim emigration into the inter-war Czechoslovakia]. Prague: KLP.

Kavitskaya, Darya. 2010. *Crimean Tatar*. Munich: LINCOM Europa.

Kizilov, Mikhail. 2009. *The Karaites of Galicia. An ethnoreligious minority among the Ashkenazim, the Turks, and the Slavs. 1772–1945*. Leiden & Boston: Brill.

Kizilov, Mikhail. 2014. National inventions: The imperial emancipation of the Karaites from Jewishness. In Roland Cvetkovski & Alexis Hofmeister (eds), *An empire of others. Making ethnographic knowledge in Imperial Russia and the USSR*, 369–394. Budapest &New York: Central European University Press.

Kocaoğlu, Timur. 2006. *Karay. The Trakai dialect*. München: LINCOM Europa.

Kohler, Kaufmann & Abraham de Harkavy. 1906. Karaites and Karaism. In *The Jewish Encyclopedia*. New York: Funk & Wagnalls.

Kotler, Igor. 1994. Karaites. In David Levinson (ed.), *Encyclopedia of world cultures: Vol VI. Russia and Eurasia / China*, 162–165. New York: G.K. Hall & Company.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Musayev, Kenesbay. M. 1964. *Grammatika karaimskogo jazyka* [Grammar of the Karaim language]. Moscow: Nauka.
Németh, Michał. 2015. A historical phonology of Western Karaim. The process of its diversification into dialects. *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 132. 167–185.
Németh, Michał. 2020. *The Western Karaim Torah. A critical edition of a manuscript from 1720*. Leiden & Boston: Brill.
Németh, Michał. 2021. Karaim, Slavic Influence on. In Marc L. Greenberg (ed.), *Encyclopedia of Slavic Languages and Linguistics Online*. Leiden: Brill.
Olach, Zsuzsanna. 2017. Emergence of a new written culture. The use of Hebrew script among the Krimchaks and the Karaim. *Acta Orientalia Vilnensia* 13. Vilnius: Vilnius University Press. 61–78.
Pełczyński, Grzegorz. 2004. *Karaimi polscy* [The Polish Karaims]. Poznań: PTPN.
Rebi, David. 2004. *Krymčakskij jazyk: Krymčaksko-russkij slovar'* [The Krymchak language: Krymchak-Russian dictionary]. Simferopol: Dolja.
Sakel, Jeanette. 2007. Types of loan: Matter and pattern. In: Matras Yaron & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 15–29. Berlin & New York: Mouton de Gruyter.
Smętek, Dorota. 2016. Some lexical and phonological features of the language of a Crimean Karaim Bible translation. *Türk Dilleri Araştırmaları* 26(2). 227–240.
Sulimowicz-Keruth, Anna. 2022. Z archiwum Aleksandra Mardkowicza: listy Zaracha Zarachowicza. *Almanach Karaimski* 11. 109–197.
Tjahlyj, Mikhailo. I. 2007. Karaïmy. In Valerij A. Smolij et al. (eds.). *Entsyklopedija istoriï Ukraïny: T. 4: Ka-Kom* [An encyclopedia of the history of Ukraine. Vol. 4: Ka-Kom]. Kyiv: Naukova dumka.
Weinreich, Uriel. 1979. *Languages in contact. Findings and problems*. Hague: Mouton Publishers.
Wexler. Paul. 1980. The Byelorussian impact on Karaite and Yiddish. *The Journal of Belarusian Studies* 4. 99–111.
Zajączkowski, Włodzimierz. 1959. Die mongolischen Elemente in der Karaimischen Sprache. *Folia Orientalia* 1. 296–302.
Zajączkowski, Włodzimierz. 1961. Die arabischen und neupersischen Lehnwörter im Karaimischen. *Folia Orientalia* 3. 177–212.
Zucker, George. K. 2001. Ladino, Spanyolit, El Kasteyano Muestro. *Shofar* 19(4). 4–14.