

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/365573879>

# Lexical copies in Khalaj: A contribution to the World Loanword Database (WOLD)

Article in *Turkic Languages* · November 2022

DOI: 10.13173/TL/2022/2/31

CITATIONS

2

READS

71

1 author:



[Mehmet Akkuş](#)

Artvin Coruh University

29 PUBLICATIONS 26 CITATIONS

[SEE PROFILE](#)

# Lexical copies in Khalaj: A contribution to the *World Loanword Database (WOLD)*

**Mehmet Akkuş**

Akkuş, Mehmet 2022. Lexical copies in Khalaj: A contribution to the *World Loanword Database (WOLD)*. *Turkic Languages* 26, 31–52.

This paper provides a quantitative analysis of lexical copies in Khalaj, an endangered Turkic language spoken in a compact language ecology in Central Iran. Surveying lexical copies in a language allows one to trace historical linguistic contact between two or more speech communities. The more intense the language contact is, the more likely it is that lexical items are hierarchically code-copied from a Model Code, in Johanson's terminology. This study thus focuses on the lexical copying (loanword) patterns of two Khalaj varieties, both of which are enclaves at the edges of the language ecology. The database for this study was constructed based on an adapted list of 1,353 items, titled *Loanword Typology list* (Haspelmath & Tadmor 2009). The Khalaj data was obtained through direct elicitations from four Khalaj-speaking language consultants in Iran. These elicitations were undertaken in July and August, 2021, in the villages of Shānegh, and Telkhāb. Since Khalaj lacks standardized orthography, the data was transcribed using IPA (International Phonetic Alphabet) conventions. As for data analysis, *Wörterbuch des Chaladsch* (Doerfer & Tezcan 1980) was used to verify the meanings of Khalaj lexical items and their meanings. A quantitative analysis was conducted to identify the extent of possible sources of lexical copying. The findings revealed that the number of copied lexical items varied between two speech communities, probably because the language ecology has had an impact on the source of items copied.

Keywords: Khalaj, code copying, lexical item, dialects, World Loanword Database

Mehmet Akkuş, Artvin Coruh University, Faculty of Science and Letters, Office No: 221, Şehir Yerleşkesi, Çayağzı Mah., 08000, Merkez/Artvin, Turkey.  
E-mail: mehmetakkus@artvin.edu.tr

## Introduction

In language contact situations, “predicting the outcome” is a quite demanding challenge for linguists (Siemund 2008: 3). However, depending on the duration of contact, its intensity, the presence or absence of imperfect learning, the status of languages between the two or more speech communities, the number of speakers, and purist attitudes of speakers, language contact may result in

- (i) mixed languages (Bakker & Mous 2013),
- (ii) code-switching/code-mixing (Gardner-Chloros 2009),

- (iii) linguistic attrition (Schmid 2013),
- (iv) creoles and pidgins (Winford 2003),
- (v) shift (Thomason and Kaufman 1988),
- (vi) language maintenance (Thomason and Kaufman 1988).

In an intense contact situation, language contact might result in an increase in both lexical copying and lexical or grammatical replication.

In order to contribute to the discussion on the roles of language contact and lexical copying, this study aims to examine lexical copying patterns of two Khalaj varieties in relation to an adapted list of 1,353 lexical items entitled the *Loanword Typology* meaning list, as a contribution to the *World Loanword Database* (WOLD) (cf. 1,460 items in original list, Haspelmath & Tadmor 2009).

### **Lexical copying and the *World Loanword Database* (WOLD): Concepts and issues**

In the literature of contact linguistics, it appears that there is no “one-size-fits-all” model defining “universals” for all contact situations (Siemund 2008), although some rigorous attempts have been made. Johanson’s (2002) *code-copying* framework is one of the most frequently used models. To better understand the fundamental processes of contact-induced language change, Johanson (2002) suggests a descriptive framework of specific language contact circumstances. The process of describing contact-induced language change phenomena in Johanson’s framework is based on theoretical notions that are similar to the ideas used in previous frameworks (Thomason & Kaufman 1988, Weinreich [1953] 1964).

Johanson (2002) uses the term *copying* intentionally, as opposed to Haugen’s (1950) *borrowing*, Weinreich’s ([1953] 1964) *interference*, Matras & Sakel’s (2007) *matter borrowing*, and Heine and Kuteva’s (2006) *transfer* and *replication*. Furthermore, most of these constructs, such as *borrowing* and *interference*, have been described as misleading (Johanson 2002). Therefore, Johanson chooses to use the term “copying” to describe his central concept. Because borrowing suggests the act of “taking” something that must be returned after a certain amount of time, “copying”, which indicates the permanence of the linguistic content copied and accepted into a language, has been utilized throughout this study (Thomason and Kaufman 1988, Weinreich [1953] 1964).

A lexical copy is a word that entered a language’s lexicon as a consequence of copying at some time in its history. Lexical copies differ from native words, i.e. words “which we can take back to the earliest known stages of a language” (Lehmann 1962: 212, as cited in Haspelmath & Tadmor 2009). It is still not fully understood why “some structural elements are ‘attractive’ in the absolute sense that they notably lend themselves to copying” (Johanson 2002: 2) but others are not in a contact situation.

In this regard, Keller (1994) states that there seems to be an invisible hand leading to language change. Johanson (2002) refers to this “invisible hand” as “attractiveness”. In his framework, attractiveness characterizes linguistic elements that are more

prone to be copied, and hence are more attractive, than others. Some linguistic elements, however, may be more resistant to copying.

The possibility of lexical copying is mainly determined by the type of language ecology and contact situation (Haspelmath & Tadmor 2009). For instance, a speech community that is under the sociopolitical control of another group may adopt administrative terms from the socially dominant group. It is thus highly likely that the contact situation will provide typological explanations. In attempting to establish generalizations across a worldwide sample of languages, a lexical database titled the *World Loanword Database* (WOLD) encompassing 41 individual language sub-databases was established based on a fixed list (Haspelmath & Tadmor 2009). The database provides “the (orthographic) form, information about the analyzability of this form, a morpheme-by-morpheme gloss (for analyzable words), information about loanword status, information about the age of the word, and optional further information of various kinds” (Haspelmath & Tadmor 2009: 2). The database provides the Model Code with its meaning, the Basic Code and some information on the copying circumstances for each lexical copying. WOLD is based on a wide array of lexical items, entitled the *Loanword Typology* meaning list. The list comprises 1,460 lexical meanings, most of which can be found in any language, such as ‘eye’, ‘to go’, ‘heavy’, and ‘sun’. It is worth noting that the list is a “meaning list”, and not a “word list”. In the database, each sub-database includes 1,000–2,000 words or counterparts. The number of lexical items in each sub-database varies, since a language may have no counterpart for a given meaning in some circumstances. The 1,460 semantic meanings are divided into 24 fields: the physical world, kinship, animals, the body, food and drink, clothing and grooming, the house, agriculture and vegetation, basic actions and technology, motion, possession, spatial relations, quantity, time, sense perception, emotions and values, cognition, speech and language, social and political relations, warfare and hunting, law, religion and belief, the modern world, and miscellaneous functional words. See Haspelmath & Tadmor (2009) for the full list of semantic meanings of the lexical items. In order to indicate whether a word is a loanword, a 5-point scale of certainty of copying is provided:

- 0. No evidence of copying
- 1. Very little evidence of copying
- 2. Perhaps copied
- 3. Probably copied
- 4. Clearly copied

In order to verify copying, a number of etymological dictionaries on the East Old Turkic and some Turkic varieties (Oghuz, Kipchak, Arghu, etc.) were used as primary sources to detect the lexical items that Khalaj could have taken from its Turkic predecessor. For lexical copying of Persian origin, the *Dehkhoda Dictionary* (Dehkhoda 1998) was consulted.

As a result of language contact, lexical copies may undergo phonological modifications. These modifications are categorized as “highly integrated”, “intermediate”,

or “unintegrated” in WOLD. Haspelmath & Tadmor (2009: 16) highlight the differences as such:

As a rough guideline, unintegrated loanwords were those that kept significant phonological and/or morphological peculiarities of the donor language and were therefore recognizable as loanwords also to speakers with no training in linguistics. On the contrary, highly integrated loanwords were those that had no structural properties that betrayed their foreign origin. Loanwords that had some synchronic properties of the foreign language were marked “intermediate”.

The language sample in WOLD covers 41 languages representing “the world’s genealogical, geographical, typological, and sociolinguistic diversity” (Haspelmath & Tadmor 2009: 3). Some of the project languages, e.g. English and Mandarin, are spoken by tens of millions of people, while others are spoken by only a few thousand or even a few hundred. In the report based on the data analysis, Selice Romani (62.7%) and Tarifit Berber (51.7%) are regarded as *very high borrowers*, with copying rates of over 50% while Mandarin Chinese (1.2%) and Old High German (5.8%) are among the *low borrowers*, with copying rates of less than 10%.

### **The Khalaj language ecology**

Khalaj (ISO-3: klj, Glottolog: turk1303), a highly endangered Turkic language, is spoken as an enclave language in Central Iran, in the valleys on the northeastern side of the central Zagros Mountain range between Hamadān and Qom, mainly in the districts of Khalajastān and Salafchegān in Qom, and the district of Āshtīān, as well as in Farāhān county in Markazī province.

In recent reports, the estimated number of speakers varies from 40,000 (Johanson 2021) to 66,000 (Hatami Khajeh 2016). Although there seems to be no consensus on the exact number of speakers, it is a well-known fact that Khalaj is mainly spoken as a family language in a number of isolated villages in rural areas of the Qom and Markazī provinces. The Khalaj people can also be characterized as an immigrant society in the sense that young people move to big cities (e.g. Tehrān, Arāk, Qom) in the pursuit of a better life, resulting in an increase in unbalanced bi-/multilingualism among younger generations. Since Persian speakers dominate in towns and cities, the proficiency and knowledge of Persian is more widespread amongst urbanized Khalaj. Non-Khalaj speaking – mostly Persian or Azeri speaking – city-dwelling outsiders seeking a more sedentary rural lifestyle are also welcomed, further confining Khalaj to family communication. Khalaj is thus a dominated language.

Persian, the official and national language of Iran, serves as a socially-dominant vernacular for all the Khalaj speakers. It is used for education, administration, and written communication in formal and informal settings. Khalaj, however, is restricted to family communication. It is worth noting that code-mixing and code-switching seem to be everyday practices in all Khalaj-speaking communities. They are hardly used in social media or written communication in general. The written material in

Khalaj includes a few poetry collections, translations, and some folkloric stories. The liturgical languages are basically Arabic and Persian. Khalaj is used only as a family language among older generations. The decrease in language use has driven the language to the brink of death.

As for the Khalaj dialectal varieties, Doerfer states that there are seven main Khalaj dialects (1988: 276):

- a. Western dialect: spoken in Borzābād and Khāltābād,
- b. Northern dialect: Vāsheqan, Mihr-e Zemin, and Chāhāk,
- c. Northeastern dialect: Mushakiya, Espit, Sefidālā, Ahmedābād, Bāgh-e Yek, Selefchegān, etc.
- d. Southern dialect: Shānegh,
- e. Central dialect: Khārrāb,
- f. Central eastern dialect: Moujan, Sorkhādeh, Mānsurābād, etc.
- g. Main dialect: mixed dialects

The Khalaj dialects are spoken in such a compact linguistic geography that speakers of each dialect may be in regular contact with speakers of more than three surrounding languages, including Persian, Azeri (an Oghuz language), and to a lesser extent Tati (Northwestern Iranian). The northern, northeastern, and western dialects, surrounded by the Azeri-speaking sphere, for instance, seem to mainly preserve their peculiar Turkic sounds due to typological proximity between Azeri and Khalaj Turkic. On the other hand, the central, mixed and southern dialects are largely affected by the Western Iranian language Persian and were probably also influenced to a lesser degree by the Northwestern Iranian language Tati. Doerfer's dialect classification (1988) needs to be further studied to give information about the grammatical level of the differences, mainly phonological, lexical, or morpho-syntactical, between the sub-varieties, taking the earliest documentation results into account; however, this is beyond the scope of the present study. Instead, this study aims to zoom in on the interplay between two distinct dialect areas and possible differences in the lexicon.

### Contact Situations: Past and present

The Khalaj are believed to have adopted the prestigious tribal name of the “Oghuz Khalaj”, with whom they once lived at the relatively isolated northwestern edge of the Turkic kaghanate, “the region between Talas and Balāsāghūn” (Johanson 2021: 91). Their language is referred to as Arghu in *Dīwān Luyāt at-Turk* by al-Kāshgarī. A subgroup of the Khalaj tribe then migrated to the south. It was probably around this time that the *Khalaj* ethnonym was mentioned in Bactrian legal documents of the 7th and 8th centuries, which were found in the vicinity of Bamiyan in northern Afghanistan and deciphered by the acclaimed Iranologist Sims-Williams (Tezcan 2018: 244). In these texts, *Khalaj* was written as *χαλασο* in Ancient Greek letters (Sims-Williams 2002: 234). In addition to the ethnonym, a number of Turkic titles were also introduced in these documents. These have historical value as they reveal that the Khalaj

had already migrated to the southwest fringe. It is estimated that some of the Khalaj who lived in northern Afghanistan and the Khorasan moved with the Oghuz tribes and settled in Central Iran, where their variety “developed in a predominantly Persian surrounding and in contact with predecessors of South Oghuz” between the 10th and 11th centuries (Johanson 2021: 131). It is recorded that the Khalaj had already settled in Sāveh, Qom, and Kashan in 1415 during the Timurid era. In the Afshar and Qajar eras in Iran, the predominantly nomadic Khalaj tribes lived in the Fars region together with the Kashkay. This explains the relatively slight Luri influence on the Khalaj morphology and to a lesser degree on its lexicon. Lexical copies in Khalaj that have Luri origins point to a (slight) historical contact with the Lur. The theme of Lur aggression is still alive in Khalaj oral history in the village of Dirmenek in Markazī province. Situated in the vicinity of Āmare and Vafs, the Khalaj have had contact with the Tati language. The contact with Persian can be categorized into two periods characterized by differing sociopolitical and sociocultural settings: (i) the period from the first appearance of the Khalaj in Central Iran, and (ii) the Pahlavi and revolutionary period. In the earlier period, it seems that there was a symbiotic relationship between Turkic-speaking and Persian-speaking peoples. Following the enthronement of Reza Shah, however, large-scale Persianization took place, with the schooling and mass media being predominantly in Persian. Furthermore, as a result of industrialization, began migrating to the relatively industrialized cities in the neighborhood such as Tehrān, Arāk, and Qom in the pursuit of a better life. This led to an increase in the use of Persian among the younger generations. Nevertheless, the Khalaj language maintained its vitality in rural villages for a while, with middle-aged women and even children being monolingual. It was mainly during this time that modern cultural lexemes began to be copied into Khalaj with their Persian names. This development can be regarded as an ongoing process caused by contact-induced language change, as is observed in (almost) all Turkic varieties of Iran. The contact situation, however, is so intricate that at times it is almost impossible to detect the source of copied lexemes. For instance, the Khalaj counterpart of ‘army’ is *qošun* in the Telkhāb variety. It appears to be ultimately copied from Mongolian, but its immediate source is unknown. The word was also used in Persian in the Ilkhanid texts, the expected intermediate Model Code. It might have been copied through either Azeri or Persian.

A systematic analysis of loanwords in Khalaj is a relatively neglected subject in the relevant literature. While some remarks on the Khalaj lexical items and etymologies can be found in Doerfer & Tezcan’s *Wörterbuch des Chaladsch*, to my best knowledge, contact influence on the Khalaj lexemes has not yet been systematically analyzed. However, contact-induced mechanisms in Khalaj have been investigated at the morphosyntactic level (Kıral 2000, Ragagnin 2020).

### Data Elicitation Process and Participants

All the Khalaj data used in this study was obtained from my linguistic fieldwork, carried out in July and August, 2021, in the villages of Shānegh and Telkhāb. My reason for selecting these two varieties is that both of them are situated in a *sui generis* linguistic ecology. Shānegh is a dialect enclave at the southern edge of the language

ecology, while Telkhāb is situated in the vicinity of Azeri-speaking communities in the north. It should also be noted that the data was collected in villages where the language consultants reported that they used their heritage language on a regular basis. Of the 1,460 different lexical items in the *Loanword Typology* list, 107 (7.3% of the database) were either unknown to the Khalaj speakers, — e.g. tropical flora ('cassava', 'manioc', 'yam', 'banyan', etc.) and fauna ('opossum', 'anteater', 'chameleon', 'elk', 'moose', etc.) — that are outside the geographical and cultural living environment of speakers in Iran or were verbs that are differentiated in the language, such as 'to take' and 'to get'. Such items were accordingly omitted from the list, and they were highlighted to check with speakers during fieldtrips. Regarding polysemies, five Khalaj lexical items have two *Loanword Typology* meanings, e.g. [æt] 'meat', 'flesh'. At this stage, native Khalaj speakers from the village of Beharestān, namely Omid (Afshin) Arabgol and Āzer Arabgol were consulted. The Khalaj database thus comprises 1,353 lexical items elicited from four speakers in a series of elicitation sessions in two dialect areas.

These elicitation sessions were carried out over the course of five weeks, since the number of lexical items (1,353) was too high to obtain in just a few sessions. Prior to the data elicitation sessions, I first translated the adapted lexical items of the *Loanword Typology* list into Persian. A Persian native speaker, Furrugh Kemalābādī Ferāhānī, checked the translations. After necessary revisions, the list was divided into six groups in order of the occurrence of field titles in the LWT list. In this respect, the lexical items of the fields *the physical world*, *kinship*, *animals*, and *the body* were prepared in the form of a PowerPoint presentation to help speakers easily visualize the expected lexical items, unless these items were abstract concepts and notions. In the second session, the items pertaining to the fields *food and drink*, *clothing and grooming*, *the house*, and *agriculture and vegetation* were elicited, while *basic actions and technology*, *motion*, *possession*, and *spatial relations* were obtained in the third session. The data concerning *quantity*, *time*, *sense perception*, and *emotions and values* was obtained in the fourth session. The items regarding *cognition*, *speech and language*, *social and political relations*, *warfare and hunting*, *law*, *religion and belief*, *modern world*, and *miscellaneous function words* fields were elicited in a combined session due to the coming religious Ashura ceremonies, observed by almost all the Shia Khalajs.

In the elicitation of the word lists, picture prompts were used together with oral prompts of the Persian lexical items. Using picture prompts instead of a translation task from the dominant language, Persian, appears to be an obvious way to prevent any interference from Persian. However, with relatively more abstract items regarding *cognition*, *miscellaneous function words*, *religion and belief*, etc. it was necessary for the researcher to make use of both types of prompts. In addition to the LWT, a basic 100-item core vocabulary list, referred to as the "Leipzig-Jakarta list" was produced to provide a quantitative analysis, taking the variables of *resistance to borrowing*, *being universal*, *simplicity*, and *stability* into account (Tadmor 2009a).

Table 1 presents demographic information about the consultants including their gender, hometown, age, dialect, and language repertoires. Abbreviations of the names of the consultants are provided.

|                     | Rām                        | Mus                        | Ami               | Hos               |
|---------------------|----------------------------|----------------------------|-------------------|-------------------|
| Age                 | 60                         | 60                         | 29                | 45                |
| Gender              | M                          | M                          | M                 | M                 |
| Home town           | Telkhāb                    | Telkhāb                    | Shānegh           | Shānegh           |
| Language repertoire | Khalaj<br>Persian<br>Azeri | Khalaj<br>Persian<br>Azeri | Khalaj<br>Persian | Khalaj<br>Persian |
| Dialect             | Northwest                  | Northwest                  | South             | South             |

Table 1. Demographic information about the Khalaj consultants

The Khalaj language consultants were between 29 and 60 years of age when the field work was carried out in Iran. As for gender, the consultants were all males. It is worth noting here that female Khalaj-speaking individuals were asked to participate in the data elicitation sessions, but they could not get permission from their male family members, i.e. fathers, husbands, sons, etc., in the highly conservative society. Thus, the Khalaj varieties described in this study reflect male-oriented varieties in general.

The northwestern Khalaj dialect consultants providing the lexical items from Telkhāb, situated at the northwestern edge of the Khalaj dialect area, were Ramazan Ali Telkhābī, and Musa Qazim Telkhābī, who were both 60 at the time of elicitation. They were both born and raised in Telkhāb, and spoke their heritage language, Khalaj, as well as the vernacular language, Persian, on a regular basis. They also stated that they were simultaneous bilinguals. Because of the influence of language shift, and because the societal setting of Khalaj as minority language is a central point in explaining the findings on the frequency and character of lexical copying, the Khalaj proficiency and acquisition patterns of the participants might help predict the pace of ongoing language shift. The elicitation in Shānegh was undertaken with two male Khalaj speakers, Amir Xātībī and Hossein Shāneghī. Amir was 29, while Hossein was a 45-year-old Khalaj-speaking villager. Besides Khalaj, they could speak Persian as a *lingua franca*.

The language consultants were video recorded while providing elicited data. During the sessions, the researcher read the lexical items out loud while pointing at the slides. Each slide contained an image or a representative picture, accompanied by its Persian translation, if this lexical item could be visualized. If not, its Persian translation in Perso-Arabic orthography was presented.

Since Khalaj lacks standardized orthography, the data was transcribed using International Phonetic Alphabet (IPA) conventions. Whenever the researcher needed to verify his interpretations during the transcription process, the items were checked with the consultants.

### Findings: Numbers and lexical copies

Since the identification of lexical copies is of crucial significance for the study, all copies were evaluated word-by-word by two linguists who were working on the Turkic languages of Iran at the time of evaluation. An iterative team collaboration approach was used to assess the data and reduce challenges to internal credibility such as researcher bias, confirmation bias, and structural corroboration. During these iterative stages, the lexical copy rating team convened to come to a consensus on the detection and classification of lexical copies in Khalaj. Whenever there was a point of disagreement in terms of rating, the raters shared their interpretations and assessments, and therefore the rating process was dialogical by its very nature. The raters held discussions to come to an agreement on the lexical copies that they had scored differently. Following a discussion of various ratings, the raters created tables identifying the Model Code(s) from which the likely loanwords were copied.

A crucial source of etymological knowledge about East Old Turkic and some Turkic varieties such as Oghuz, Kipchak, and Arghu is Maḥmūd al-Kāšyarī's (1072–1074) *Dīwān Luḡāt at-Turk* 'Compendium of the Turkic dialects' (Johanson 2021). *An Etymological Dictionary of Pre-Thirteenth Century Turkish* (Clauson 1972) and *Etymological Dictionary of the Turkic Languages* (ESTJ 1974–2003) were also used as primary sources to detect lexical items that Khalaj may have taken from its Turkic predecessor. A complementary source was Räsänen's (1969) relatively short etymological dictionary. *Dīwān Luḡāt at-Turk* was also consulted for detecting Oghuz Azeri copies in Khalaj. For lexical copying of Persian origin, there is the *Dehkhoda Dictionary*, the most comprehensive Persian dictionary ever published, consisting of circa 200 volumes. As for the Tati loanwords, there is Yarshater's (1969) *A Grammar of Southern Tati Dialects*, and Stilo's (1971) *A Grammar of Vafsi-Tati*. Even though there are few glosses in these works, to my knowledge, they are the sole sources for the Tati varieties spoken in the vicinity. For etymologizing the words in Khalaj, my primary resource was Doerfer & Tezcan's (1980) *Wörterbuch des Chaladsch*, which I consulted when I was uncertain of a word's exact semantic field and possible etymology.

Table 2 summarizes the results of the data as regards the incidence for lexical copying by model code in the Shānegh and Telkhāb varieties of Khalaj.

|                                    | Persian | Arabic | Azeri | French | Other | Total copies | Non-copies |
|------------------------------------|---------|--------|-------|--------|-------|--------------|------------|
| Lexical copies in the Shānegh data | 51.8    | 10.6   | 0.0   | 1.6    | 0.7   | 64.7         | 35.3       |
| Lexical copies in the Telkhāb data | 47.9    | 10.6   | 1.3   | 1.4    | 0.8   | 62.0         | 38.0       |

Table 2. Lexical copies in Khalaj by model code in the Shānegh and Telkhāb varieties of Khalaj (percentages)

The figures in Table 2 demonstrate that the Khalaj varieties are among *very high borrowers* (Haspelmath & Tadmor 2009: 57), with a copying rate of over 50%. The data

reveals that the great bulk of lexical copies in Khalaj were copied from Persian, the socially dominant language in Iran. The figures also show that Arabic lexemes were copied at a high rate into the Khalaj varieties. Lying in an enclave the southern edge of the dialect area, the Shānegh variety has evidently not been influenced by Azeri, an Oghuz language. Situated in the vicinity of Azeri-speaking communities to the north, the Telkhāb variety, however, has lexical copies taken from Azeri, with a percentage of 1.3%. Compared to the Shānegh variety, this variety is also characterized by its having more original Khalaj words, probably owing to a number of sociolinguistic factors:

- (a) Telkhāb is not a dialect enclave.
- (b) It has relatively more speakers.
- (c) It is in regular contact with the other Khalaj villages.

These might explain why it is more “conservative” regarding lexical copying than the Shānegh variety.

*Wörterbuch des Chaladsch* (Doerfer & Tezcan 1980) was used to verify the meanings of Khalaj lexical items and their usages. It was also beneficial for making comparisons with the Khalaj lexical items documented some 50 years ago. Doerfer’s *Lexik und Sprachgeographie des Chaladsch* (1987) was also utilized to compare the current dialectal situation. With respect to the relationship between lexical copying and semantic fields, Table 3 presents the percentages of lexical copies by model code and semantic field in the Shānegh enclave.

|    |                                | Persian | Arabic | Azeri | French | Other | Total copies | Non-copies |
|----|--------------------------------|---------|--------|-------|--------|-------|--------------|------------|
| 1  | The physical world             | 50.0    | 17.2   | 0.0   | 0.0    | 0.0   | 67.2         | 32.8       |
| 2  | Kinship                        | 43.9    | 13.6   | 0.0   | 0.0    | 0.0   | 57.5         | 42.5       |
| 3  | Animals                        | 59.3    | 5.5    | 0.0   | 0.0    | 3.3   | 68.1         | 31.9       |
| 4  | The body                       | 53.6    | 3.9    | 0.0   | 0.0    | 0.7   | 58.2         | 41.8       |
| 5  | Food and drink                 | 52.5    | 12.5   | 0.0   | 2.5    | 0.0   | 67.5         | 32.5       |
| 6  | Clothing and grooming          | 50.9    | 11.3   | 0.0   | 0.0    | 3.8   | 66.0         | 34.0       |
| 7  | The house                      | 70.7    | 4.9    | 0.0   | 0.0    | 0.0   | 75.6         | 24.4       |
| 8  | Agriculture and vegetation     | 76.9    | 7.7    | 0.0   | 0.0    | 0.0   | 84.6         | 15.4       |
| 9  | Basic actions and technology   | 45.4    | 1.3    | 0.0   | 0.0    | 1.3   | 48.0         | 52.0       |
| 10 | Motion                         | 29.2    | 0.0    | 0.0   | 0.0    | 0.0   | 29.2         | 70.8       |
| 11 | Possession                     | 60.5    | 9.3    | 0.0   | 0.0    | 0.0   | 69.8         | 30.2       |
| 12 | Spatial relations              | 33.8    | 5.4    | 0.0   | 0.0    | 0.0   | 39.2         | 60.8       |
| 13 | Quantity                       | 30.8    | 5.1    | 0.0   | 0.0    | 0.0   | 35.9         | 64.1       |
| 14 | Time                           | 50.9    | 12.3   | 0.0   | 0.0    | 0.0   | 63.2         | 36.8       |
| 15 | Sense perception               | 60.4    | 4.2    | 0.0   | 0.0    | 0.0   | 64.6         | 35.4       |
| 16 | Emotions and values            | 46.8    | 29.8   | 0.0   | 0.0    | 2.1   | 79.7         | 21.3       |
| 17 | Cognition                      | 44.0    | 20.0   | 0.0   | 0.0    | 0.0   | 64.0         | 36.0       |
| 18 | Speech and language            | 53.8    | 12.9   | 0.0   | 0.0    | 0.0   | 66.7         | 33.3       |
| 19 | Social and political relations | 91.7    | 2.8    | 0.0   | 0.0    | 0.0   | 94.4         | 5.6        |
| 20 | Warfare and hunting            | 76.9    | 15.4   | 0.0   | 0.0    | 0.0   | 92.3         | 7.7        |

|    |                              |      |      |     |      |     |       |      |
|----|------------------------------|------|------|-----|------|-----|-------|------|
| 21 | Law                          | 42.3 | 42.3 | 0.0 | 0.0  | 0.0 | 84.6  | 15.4 |
| 22 | Religion and belief          | 56.0 | 44.0 | 0.0 | 0.0  | 0.0 | 100.0 | 0.0  |
| 23 | Modern world                 | 44.6 | 19.6 | 0.0 | 30.4 | 5.4 | 100.0 | 0.0  |
| 24 | Miscellaneous function words | 21.4 | 0.0  | 0.0 | 0.0  | 0.0 | 21.4  | 78.6 |
|    | all words                    | 51.8 | 10.6 | 0.0 | 1.6  | 0.7 | 64.7  | 35.3 |

Table 3. Lexical copies in Khalaj by model code and semantic field in the Shānegh variety (percentages)

The data reveals that Persian penetrates almost every semantic field, as it is the socially dominant language in Iran. The findings indicate that Persian enters into all semantic fields. Persian influence is however predominantly concentrated on the semantic fields *social and political relations* (91.7%), *warfare and hunting* (76.9%), *agriculture and vegetation* (76.9%), *the house* (70.7%), *possession* (60.5%), *sense perception* (60.4%), *animals* (59.3%), *religion and belief* (56.0%), *speech and language* (53.8%), *the body* (53.6%), *food and drink* (52.5%), and *the physical world* (50.0%). Persian contributes less than half of the copies in the remaining semantic fields.

Arabic is the second most common model code, accounting for 10.6% of all lexical copies in the database. Arabic copies cover 22 of the 24 semantic categories, according to the findings. In the semantic fields *motion* and *miscellaneous function words*, however, there are no Arabic copies. These data demonstrate that Arabic has a pervasive cultural and religious influence on the Khalaj lexicon, as is common in neighboring languages shaped by Islamic culture. Arabic lexical copies unsurprisingly affect the semantic fields *religion and belief* (44.0%) and *law* (42.3%), certainly profiting from the fact that it functions as a dominant liturgical language, along with Persian, in Iran. It is also worth noting that there are no Azeri lexical copies in this variety, probably because there are no Azeri-speaking villages in the surrounding area. Table 4 provides further lexical copies by semantic word class and model code (percentages) for this variety.

|                | Persian | Arabic | Azeri | French | Other | Total loanwords | Non-loanwords |
|----------------|---------|--------|-------|--------|-------|-----------------|---------------|
| Nouns          | 59.6    | 15.4   | 0.0   | 2.0    | 2.3   | 78.4            | 25.6          |
| Verbs          | 53.5    | 0.0    | 0.0   | 0.3    | 0.0   | 53.8            | 46.2          |
| Adjectives     | 25.8    | 9.2    | 0.0   | 0.0    | 0.0   | 35.0            | 65.0          |
| Adverbs        | 25.0    | 0.0    | 0.0   | 0.0    | 0.0   | 25.0            | 75.0          |
| Function words | 39.6    | 0.0    | 0.0   | 0.0    | 0.0   | 39.6            | 60.4          |
| All words      | 51.8    | 10.6   | 0.0   | 1.6    | 0.7   | 64.7            | 35.3          |

Table 4. Lexical copies by semantic word class and model code (percentages) in the Shānegh variety (percentages)

When sorted by semantic word class, Khalaj loanwords follow the general pattern whereby nouns are the most easily copied word class, as demonstrated in Table 4.

Lexical copies account for over three-fourths of the nouns in the database, whereas verbs account for about half of copies (53.8%) in the relevant word class. No inflecting verbs have been copied from Persian, French or Arabic into Khalaj. Instead, Khalaj makes use of a productive adaptation process, using light verbs. The data analysis also reveals that the Telkhāb variety has adapted and copied verbs either directly from Persian or via Persian as a “bridge language”. In this respect, as Family (2014: 25) has pointed out, Persian light verb constructions (LVCs) “form the foundation of a complex network of verbal collocations that have replaced simple verbs in the language”. For instance, the Persian light verb construction /ændʒām dādæn/ ‘to do’, ‘to make’ constitutes a Persian noun /ændʒām/ ‘performance’, ‘implementation’ and a Persian light verb /dādæn/ ‘to give’. In constructing LVCs, Persian could make use of adapted Arabic nominals as well, e.g. /ædāmæ dādæn/. /ædāmæ/ is an adapted nominalized Arabic word, which is used innovatively with a Persian light verb /dādæn/ ‘to give’, meaning ‘to continue’. As a result of prolonged contact with Persian, Khalaj integrated the same process by using light verbs of Turkic origin, such as /vermæk/ ‘to give’, thus leading to a Khalaj light verb construction: /ædāmæ vermæk/, which is evidently a loan translation from Persian. It is a well-known fact in the relevant literature that light verbs constructions are formed with *kil-*, *et-*, etc., but not *ver-*, in historical and modern Turkic languages. These loan translations of LVCs formed with Arabic nominals were also treated as Persian copies into Khalaj, since Arabic verbs are built around a root consisting of two to five (though generally three) consonants (triliteral or quadriliteral according to the number of consonants) such as ك-ت-ب “k-t-b” ‘to write’ (Ryding 2005). Thus, the source/donor language of a loan is identified by comparing the “original” usage pattern and “bridged” pattern. Khalaj does not copy the original Arabic verbs which are built around a root. Rather, it copies the bridged LVCs through Persian. The lexemes in (1) exemplify a number of loan-translated light verb constructions in the Shānegh variety.

|                              |   |                |
|------------------------------|---|----------------|
| (1) Persian                  | ⇒ | Shā. Khalaj    |
| /jæx zædæn/ ‘to freeze’      |   | /jæx vor-/     |
| /tʃæʃmæk zædæn/ ‘to blink’   |   | /tʃæʃmæk vor-/ |
| /næfæs kæʃidæn/ ‘to breathe’ |   | /næfæs tʃæk-/  |
| /gæz zædæn/ ‘to bite’        |   | /gæzzæ tot-/   |

The data also reveals that the proportions of copies of adjectives, adverbs, and function words are under 50.0%. Table 5 presents the percentages of lexical copies by model code and semantic field in the Telkhāb variety.

Table 5 gives the percentages of lexical copies in the Telkhāb variety by semantic field and main model codes. The figures demonstrate that Persian penetrates all semantic fields. Persian is predominantly concentrated in the semantic fields of *social and political relations* (88.9%), *agriculture and vegetation* (72.3%), *warfare and hunting* (71.7%), *the house* (70.7%), *possession* (60.5%), *religion and belief* (56%), *speech and language* (53.8%), and *food and drink* (50.0%). In the remaining semantic fields, Persian contributes less than 50% copies.

|    |                                | Persian | Arabic | Azeri | French | Other | Total copies | Non-copies |
|----|--------------------------------|---------|--------|-------|--------|-------|--------------|------------|
| 1  | The physical world             | 45.3    | 17.2   | 4.7   | 0.0    | 0.0   | 67.2         | 32.8       |
| 2  | Kinship                        | 37.9    | 16.7   | 1.5   | 0.0    | 0.0   | 56.1         | 43.9       |
| 3  | Animals                        | 47.3    | 5.5    | 2.2   | 0.0    | 3.3   | 58.3         | 41.7       |
| 4  | The body                       | 47.1    | 3.3    | 1.3   | 0.0    | 0.0   | 51.7         | 48.3       |
| 5  | Food and drink                 | 50.0    | 12.5   | 0.0   | 2.5    | 0.0   | 65.0         | 35.0       |
| 6  | Clothing and grooming          | 47.2    | 11.3   | 1.9   | 3.8    | 0.0   | 64.2         | 35.8       |
| 7  | The house                      | 70.7    | 4.9    | 0.0   | 0.0    | 0.0   | 75.6         | 24.4       |
| 8  | Agriculture and vegetation     | 72.3    | 7.7    | 1.6   | 0.0    | 0.0   | 81.6         | 18.4       |
| 9  | Basic actions and technology   | 45.4    | 1.3    | 0.0   | 0.0    | 1.3   | 48.0         | 52.0       |
| 10 | Motion                         | 29.2    | 0.0    | 0.0   | 0.0    | 0.0   | 29.2         | 70.8       |
| 11 | Possession                     | 60.5    | 9.3    | 0.0   | 0.0    | 0.0   | 69.8         | 30.2       |
| 12 | Spatial relations              | 28.4    | 5.4    | 2.7   | 0.0    | 0.0   | 36.5         | 63.5       |
| 13 | Quantity                       | 20.5    | 5.1    | 0.0   | 0.0    | 0.0   | 25.6         | 74.4       |
| 14 | Time                           | 45.6    | 12.3   | 0.0   | 0.0    | 0.0   | 57.9         | 42.1       |
| 15 | Sense perception               | 54.1    | 4.2    | 4.2   | 0.0    | 0.0   | 62.5         | 37.5       |
| 16 | Emotions and values            | 42.6    | 29.8   | 4.2   | 0.0    | 2.1   | 78.7         | 21.3       |
| 17 | Cognition                      | 44.0    | 20.0   | 0.0   | 0.0    | 0.0   | 64.0         | 36.0       |
| 18 | Speech and language            | 53.8    | 12.9   | 0.0   | 0.0    | 0.0   | 66.7         | 33.3       |
| 19 | Social and political relations | 88.9    | 2.8    | 0.0   | 0.0    | 0.0   | 91.7         | 8.3        |
| 20 | Warfare and hunting            | 71.7    | 15.4   | 2.6   | 0.0    | 2.6   | 92.3         | 7.7        |
| 21 | Law                            | 42.3    | 42.3   | 0.0   | 0.0    | 0.0   | 84.6         | 15.4       |
| 22 | Religion and belief            | 56.0    | 44.0   | 0.0   | 0.0    | 0.0   | 100.0        | 0.0        |
| 23 | Modern world                   | 44.6    | 19.6   | 0.0   | 30.4   | 5.4   | 100.0        | 0.0        |
| 24 | Miscellaneous function words   | 21.4    | 0.0    | 0.0   | 0.0    | 0.0   | 21.4         | 78.6       |
|    | all words                      | 47.9    | 10.6   | 1.3   | 1.4    | 0.8   | 62.0         | 38.0       |

Table 5. Lexical copies in Khalaj by model code and semantic field in the Telkhāb variety (percentages)

Arabic is the second model code, contributing 10.6% of lexical copies to the database. The findings reveal that Arabic copies cover 22 out of 24 semantic fields. There are, however, no Arabic copies in the semantic fields *motion*, and *miscellaneous function words*. These findings highlight that Arabic has exerted a widespread cultural and religious impact on the Khalaj vocabulary, as is frequently observed in the neighboring languages under the influence of Islamic culture.

Contrary to the Shānegh variety, the Telkhāb variety copies Azeri words, clearly due to its shared ecology with Azeri-speaking villages at the northwestern edge of the Khalaj-speaking language ecology. This results in a particular kind of contact. It is evident that Azeri has left its mark on 10 fields, meaning that Azeri copies contribute to the semantic fields *the physical world* (4.7%), *sense perception* (4.2%), *emotion and values* (4.2%), *spatial relations* (2.7%), *warfare and hunting* (2.6%), *animals* (2.2%), *clothing and grooming* (1.9%), *agriculture and vegetation* (1.7%), *kinship* (1.6%), and *the body* (1.3%).

The data reveals that the French influence is observed in the semantic field *modern world* (30.4%), as well as in the fields (modern) *food and drink* and *clothing*. These

copies of ultimately French origin probably made their ways to Khalaj via Persian. Here, it is worth noting that the Iranian intellectuals may have introduced these French copies in the early 19th century when “the French language gained a social prestige in the society and influenced the choice of foreign language teaching later in the country” (Farhady & Hezaveh & Hedayati 2010: 10). Almost all members of Iranian social strata later became acquainted with these “educated” copies (Schulte 2009: 241). French loanwords that entered Khalaj via Persian are not totally different as opposed to Arabic LVCs. Table 6 further provides lexical copies by semantic word class and model code (percentages) for this variety.

|                | Persian | Arabic | Azeri | French | Other | Total loanwords | Non-loanwords |
|----------------|---------|--------|-------|--------|-------|-----------------|---------------|
| Nouns          | 53.8    | 13.2   | 1.7   | 2.1    | 1.3   | 72.1            | 27.9          |
| Verbs          | 52.6    | 0.0    | 0.0   | 0.3    | 0.0   | 52.9            | 47.1          |
| Adjectives     | 23.3    | 9.1    | 1.7   | 0.0    | 0.0   | 34.1            | 65.9          |
| Adverbs        | 25.0    | 0.0    | 0.0   | 0.0    | 0.0   | 25.0            | 75.0          |
| Function words | 37.2    | 2.3    | 0.0   | 0.0    | 0.0   | 34.9            | 65.1          |
| All words      | 47.9    | 10.6   | 1.3   | 1.4    | 0.8   | 62.0            | 38.0          |

Table 6. Lexical copies by semantic word class and model code (percentages) in the Telkhāb variety

As in the Shānegh variety, nouns have the highest proportion of copies of any word class in the Telkhāb variety: circa three-quarters (72.1%). The other content word classes lag behind nouns in terms of copy proportions, with the proportion for verbs being 52.9%. The data indicates that the proportion of copies of adjectives, adverbs, and function words are under 50.0%.

The findings also indicate that the Telkhāb variety has adapted and copied verbs either directly from Persian or via Persian as a “bridge language”. In this variety, even though Azeri exerts a slight influence on nouns (1.7%), it appears that there is no Azeri impact on verbs. The number of Persian verbs copied into the Telkhāb Khalaj is slightly less (52.6%) than in the Shānegh variety (53.5%). A number of light verb constructions in the Telkhāb variety are exemplified in (2).

|                              |   |               |
|------------------------------|---|---------------|
| (2) Persian                  | ⇒ | Tel. Khalaj   |
| /jæx zædæn/ ‘to freeze’      |   | /buzla-/      |
| /tʃæjmæk zædæn/ ‘to blink’   |   | /gœz vɔr-/    |
| /næfæs kæfidæn/ ‘to breathe’ |   | /næfæs tʃæk-/ |
| /gæz zædæn/ ‘to bite’        |   | /gæz vɔr-/    |

As in the Shānegh variety, nouns have the highest proportion of copies of any word class in the Telkhāb variety: circa three-quarters (72.1%). The other content word classes lag behind nouns in terms of loanword proportions, with the proportion for

verbs being 52.9%. The data indicates that the proportion of copies of adjectives, adverbs, and function words is under 50.0%.

Besides the LWT, taking the variables of *resistance to borrowing*, *being universal*, *simplicity*, and *stability* into account, a basic 100-item core vocabulary list, referred to as the “Leipzig-Jakarta list” was produced (Tadmor 2009a). The results of applying the Leipzig-Jakarta vocabulary list to both the Telkhāb and Shānegh varieties are presented in Table 7.

|    | Persian (IPA symbols) | English “translation” | Shā. Khalaj (IPA symbols) | Tel. Khalaj (IPA symbols) |
|----|-----------------------|-----------------------|---------------------------|---------------------------|
| 1  | /a:tæf/               | ‘fire’                | /ʎuot/                    | /hət/                     |
| 2  | /bi:ni:/              | ‘nose’                | /burun/                   | /burun/                   |
| 3  | /ræftæn/              | ‘to go’               | /varmaq/                  | /varmaq/                  |
| 4  | /a:b/                 | ‘water’               | /suv/                     | /suv/                     |
| 5  | /dæha:n/              | ‘mouth’               | /a:yuz/                   | /a:yuz/                   |
| 6  | /zæba:n/              | ‘tongue’              | /til/                     | /til/                     |
| 7  | /xu:n/                | ‘blood’               | /qa:n/                    | /qa:n/                    |
| 8  | /ostaxa:n/            | ‘bone’                | /simik/                   | /symyk/                   |
| 9  | /tə/                  | ‘2SG pronoun’         | /sæn/                     | /sæn/                     |
| 10 | /ri:fɛ/               | ‘root’                | /ri:fɛ/                   | /ri:fɛ/                   |
| 11 | /a:mædæn/             | ‘to come’             | /kɛlmæk/                  | /kɛlmæk/                  |
| 12 | /pestan/              | ‘breast’              | /mæmæk/                   | /mæmæk/                   |
| 13 | /ba:ra:n/             | ‘rain’                | /ʎa:ɣuʃ/                  | /ʎa:ɣiʃ/                  |
| 14 | /mæn/                 | ‘1SG pronoun’         | /mæn/                     | /mæn/                     |
| 15 | /na:m/                | ‘name’                | /a:t/                     | /a:t/                     |
| 16 | /ʃæpɛʃ/               | ‘louse’               | /bet/                     | /bit/                     |
| 17 | /pær/                 | ‘wing’                | /pær/                     | /pær/                     |
| 18 | /gəʃt/                | ‘flesh/meat’          | /æt/                      | /æt/                      |
| 19 | /dæst/                | ‘arm/hand’            | /qəʃl/, /æʃl/             | /qəʃl/, /æʃl/             |
| 20 | /mægæs/               | ‘fly’                 | /mægæs/                   | /mægæs/                   |
| 21 | /ʃæb/                 | ‘night’               | /ketʃɛ/                   | /ketʃɛ/                   |
| 22 | /gu:f/                | ‘ear’                 | /qulɛχ/                   | /qulɛq/                   |
| 23 | /pəʃti gærdæn/        | ‘neck’                | /boʃʎun/                  | /boʃʎun/                  |
| 24 | /dər/                 | ‘far’                 | /hirax/                   | /hiraq/                   |
| 25 | /kærdæn/              | ‘to do/make’          | /iʃ kermæk/               | /iʃ kormæk/               |
| 26 | /xu:nɛ/               | ‘house’               | /hæv/                     | /hæv/                     |
| 27 | /sæŋg/                | ‘stone/rock’          | /ta:ʃ/                    | /ta:ʃ/                    |
| 28 | /tælx/                | ‘bitter’              | /tæʎχ/                    | /ha:tʃʎug/                |
| 29 | /gəʃtæn/              | ‘to say’              | /hæʎmaq/                  | /ha:ʎmaq/                 |
| 30 | /dændə:n/             | ‘tooth’               | /ti:ʃ/                    | /ti:ʃ/                    |
| 31 | /mo/                  | ‘hair’                | /qil/                     | /tyk/                     |
| 32 | /bozorg/              | ‘big’                 | /ʎəɣun/                   | /jækæ/                    |
| 33 | /jæk/                 | ‘one’                 | /bi:/                     | /bi:/                     |
| 34 | /ki:/                 | ‘who’                 | /kim/                     | /kim/                     |
| 35 | /ə/                   | ‘3SG pronoun’         | /ə/                       | /ə/                       |
| 36 | /zædæn/               | ‘to hit, ‘beat’       | /vurmaq/                  | /vurmaq/                  |
| 37 | /pa:/                 | ‘leg’, ‘foot’         | /hadaq/                   | /qitʃ/                    |
| 38 | /ʃa:χ/                | ‘horn’                | /ʃa:χ/                    | ---                       |
| 39 | /i:n/                 | ‘this’                | /bo/                      | /bo/                      |
| 40 | /ma:hi/               | ‘fish’                | /ma:hi/                   | /ba:luk/                  |
| 41 | /di:ruz/              | ‘yesterday’           | /æŋgir/                   | /æŋgir/                   |
| 42 | /χordæn/              | ‘to drink’            | /itʃmæk/                  | /itʃmæk/                  |
| 43 | /seʎah/               | ‘black’               | /qaru/                    | /qara/                    |

|    |                  |                     |                |              |
|----|------------------|---------------------|----------------|--------------|
| 44 | /na:f/           | ‘navel’             | /na:f/         | /kyndyk/     |
| 45 | /i:sta:dæn/      | ‘to stand’          | /turmaq/       | /turmaq/     |
| 46 | /yæz zædæn/      | ‘to bite’           | /yæzzæ tutmaq/ | /yaz vurmaq/ |
| 47 | /pɔʃt/           | ‘back’              | /ha:rt/        | /ha:ya/      |
| 48 | /ba:d/           | ‘wind’              | /jel/          | /jel/        |
| 49 | /dud/            | ‘smoke’             | /dud/          | /tytyn/      |
| 50 | /tʃɛ/            | ‘what?’             | /ne:sɛ/        | /ne:sɛ/      |
| 51 | /bætʃtʃɛ/        | ‘child’             | /kuokɛ/        | /kɛ:l/       |
| 52 | /tɔχmi mory/     | ‘egg’               | /numuxte/      | /numurtya/   |
| 53 | /da:dæn/         | ‘to give’           | /vermæk/       | /vermæk/     |
| 54 | /nɔ/             | ‘new’               | /hingi/        | /jengi/      |
| 55 | /sɔχtæn/         | ‘to burn (intr.)’   | /kenmæk/       | /kɔnmæk/     |
| 56 | /næ/             | ‘not’               | /dɛg/          | /dɛg/        |
| 57 | /χɔb/            | ‘good’              | /hɔvul/        | /ha:vul/     |
| 58 | /da:nɛstæn/      | ‘to know’           | /bilmæk/       | /bilmæk/     |
| 59 | /za:nu/          | ‘knee’              | /ti:z/         | /ti:z/       |
| 60 | /ʃi:nɛ/          | ‘sand’              | /ʃin/          | /ʃin/        |
| 61 | /χændidæn/       | ‘to laugh’          | /kilmæk/       | /gylmæk/     |
| 62 | /ʃæni:dæn/       | ‘to hear’           | /ɛʃitmæk/      | /ɛʃitmæk/    |
| 63 | /χa:q/           | ‘soil’              | /tɔrɔpɔ:x/     | /turpɔ:q/    |
| 64 | /bærg/           | ‘leaf’              | /bælg/         | /karpɑ:q/    |
| 65 | /qærmæz/         | ‘red’               | /quzɪl/        | /qirmiz/     |
| 66 | /kæbæd/          | ‘liver’             | /ta:lɛx/       | /ta:laq/     |
| 67 | /pɛnha:n kærdæn/ | ‘to hide’           | /bæktɔmæk/     | /kɛʃitmæk/   |
| 68 | /pɔst/           | ‘skin’, ‘hide’      | /tæri/         | /tæri/       |
| 69 | /mækidæn/        | ‘to suck’           | /itʃmæk/       | /æmmæk/      |
| 70 | /hæml kærdæn/    | ‘to carry’          | /kærtɔmæk/     | /qavzamaq/   |
| 71 | /murtʃɛ/         | ‘ant’               | /murtʃæni/     | /qaruntʃa/   |
| 72 | /sængi:n/        | ‘heavy’             | /æyir/         | /qurs/       |
| 73 | /gæræftæn/       | ‘to take’           | /almaq/        | /almaq/      |
| 74 | /pi:r/           | ‘old’               | /qarru/        | /qarru/      |
| 75 | /χordæn/         | ‘to eat’            | /jemæk/        | /jemæk/      |
| 76 | /ra:n/           | ‘thigh’             | /bɔt/          | /bɔt/        |
| 77 | /qɔloft/         | ‘thick’             | /kɔyɔn/        | /kɔyɔn/      |
| 78 | /tɔla:ni/        | ‘long’              | /u:zu:n/       | /yira:q/     |
| 79 | /dæmidæn/        | ‘to blow’           | /futtʃæmæk/    | /pyflemæk/   |
| 80 | /tʃub/           | ‘wood’              | /hæyɑtʃ/       | /hayɑtʃ/     |
| 81 | /dævidæn/        | ‘to run’            | /hirmæk/       | /tɛzmæk/     |
| 82 | /ɔftɑ:dæn/       | ‘to fall’           | /tʃɪmæk/       | /tyʃmæk/     |
| 83 | /tʃæʃm/          | ‘eye’               | /kez/          | /kɔz/        |
| 84 | /χa:kɛstær/      | ‘ash’               | /kil/          | /kyl/        |
| 85 | /dɔm/            | ‘tail’              | /dɔm/          | /qudruq/     |
| 86 | /sæg/            | ‘dog’               | /i:t/          | /i:t/        |
| 87 | /gerjɛ kærdæn/   | ‘to cry’, ‘to weep’ | /hiylɛmæk/     | /haylamaq/   |
| 88 | /bæstæn/         | ‘to tie’            | /va:ɫmaq/      | /va:maq/     |
| 89 | /didæn/          | ‘to see’            | /kermæk/       | /kɔrmæk/     |
| 90 | /ʃiri:n/         | ‘sweet’             | /ʃi:rɪn/       | /ʃi:rɪn/     |
| 91 | /tænɛf/          | ‘rope’              | /tænɛf/        | /tænɛf/      |
| 92 | /sa:jæ/          | ‘shade’, ‘shadow’   | /sa:jæ/        | /kɔlgæ/      |
| 93 | /pærændæ/        | ‘bird’              | /pærændæ/      | /pærændæ/    |
| 94 | /næmæk/          | ‘salt’              | /tɔz/          | /tɔz/        |
| 95 | /qɔtʃæk/         | ‘small’             | /qutʃulu/      | /bala/       |
| 96 | /væsi:ʔ/         | ‘wide’              | /kɔyɔn/        | /jekke/      |
| 97 | /sæta:ræ/        | ‘star’              | /sæta:ræ/      | /kulduduz/   |
| 98 | /ʌn)dær/         | ‘in’                | /itʃ/          | /itʃ/        |

|     |               |                        |              |            |
|-----|---------------|------------------------|--------------|------------|
| 99  | /sæχt/        | ‘hard’                 | /sæχt/       | /tʃætɪn/   |
| 100 | /χord kærdæn/ | ‘to crush’, ‘to grind’ | /χurt fæmæk/ | /toɣramaq/ |

Table 7. Core vocabulary in the Shānegh and Telkhāb varieties of Khalaj based on the Leipzig-Jakarta list

Table 7 reveals that the northern dialect, Telkhāb Khalaj, retains the lexical items of Turkic origin (including those of Oghuz origin, e.g. /qaruntʃa/, /tʃætɪn/, etc.), in contrast to the Shānegh variety. When it comes to the percentage of the lexical copies in the varieties, Table 8 presents the results of the data as regards the percentage of lexical copying by model code based on the Leipzig-Jakarta list in the Shānegh and Telkhāb varieties of Khalaj.

|                                    | Persian | Arabic | Azeri | French | Other | Total copies | Non-copies |
|------------------------------------|---------|--------|-------|--------|-------|--------------|------------|
| Lexical copies in the Shānegh data | 22.0    | 0.0    | 0.0   | 0.0    | 2.0   | 24.0         | 76.0       |
| Lexical copies in the Telkhāb data | 9.09    | 0.0    | 8.08  | 0.0    | 0.0   | 17.2         | 82.8       |

Table 8. Core vocabulary by semantic word class and model code (percentages) (based on the Leipzig-Jakarta list) in the Shānegh and Telkhāb varieties of Khalaj.

The data analysis reveals that the proportion of the lexical copies is astonishingly low in the database, as demonstrated in Table 8. It is evident that the Shānegh and Telkhāb varieties of Khalaj mainly retain the core vocabulary inherited from the Common Turkic. It is also worth noting that the Telkhāb variety appears to be more “conservative” (82.8% non-copies) than the Shānegh variety (76.0% non-copies), a dialect enclave in the southern edge. Moreover, the Telkhāb variety has apparently copied Azeri lexemes (8.08%), leading to a rise in the total percentage of lexical items of Turkic origin (91.9%). The data also suggests that the core vocabulary is more invulnerable to copying.

### Integration of lexical copies

Contrary to the modern Turkic languages, sound harmony in Khalaj is weakened. Thus, lexical copies violating harmony rules retain their original forms with minor changes when copied both into the Telkhāb and Shānegh varieties, such as Shā. /sætɑ:ræ/ ‘star’, Sha. /gɔla:bi:/ ‘pear’, Sha. /xærgu:f/ ‘rabbit’, as displayed in examples (3).

|     |                    |   |             |             |
|-----|--------------------|---|-------------|-------------|
| (3) | Persian            | ⇒ | Shā. Khalaj | Tel. Khalaj |
|     | /xærgu:f/ ‘rabbit’ |   | /xærgu:f/ - |             |
|     | /tænɐf/ ‘rope’     |   | /tænɐf/     | /tænɐf/     |
|     | /murtʃɛ/ ‘ant’     |   | /murtʃæni/  | -           |
|     | /gɔla:bi:/ ‘pear’  |   | /gɔla:bi:/  | /gɔla:bi:/  |

As observed in examples (3), vowels are not changed to follow sound harmony rules in the Khalaj varieties in this study. This “rule” can also be extended to Arabic lexical copies probably because these copies (4) made their way to Khalaj via Persian. The contact-induced change at the phonological level, which is almost certainly due to the prolonged and intense sociolinguistic contact (Doerfer 1971), is highly likely to have been triggered by the pressure of socially dominant languages as well as by a lack of socially-driven motivations to use the Khalaj Turkic language for communication in the region (Bosnalı 2016).

|     |                 |         |             |             |
|-----|-----------------|---------|-------------|-------------|
| (4) | Arabic ⇒        | Persian | Shā. Khalaj | Tel. Khalaj |
|     | /sæxri/ ‘cliff’ | /sɔxrɛ/ | /sɔxrɛ/     | /sɔxrɛ/     |
|     | /sæhɛl/ ‘shore’ | /sɛhɛl/ | /sɔhɛl/     | /sɛhɛl/     |
|     | /mɔwʒɛ/ ‘wave’  | /mɔvdʒ/ | /mɔvdʒ/     | /mɔvdʒ/     |

This explanation also seems to apply to French lexical copies (5) in Khalaj. As stated earlier, these “educated” copies of ultimately French origin probably entered Khalaj via Persian.

|     |                      |           |             |             |
|-----|----------------------|-----------|-------------|-------------|
| (5) | French ⇒             | Persian   | Shā. Khalaj | Tel. Khalaj |
|     | /plasti:k/ ‘plastic’ | /plastik/ | /plastik/   | /plastik/   |
|     | /film/ ‘movie’       | /film/    | /film/      | /film/      |

Consonant clusters in Persian (6), as well as in French (7) are not broken up even though they are normally broken up by prothetic or epenthetic vowels in Turkic languages. For Sakha, see Pakendorf & Novgorodov (2009).

|     |                  |             |             |
|-----|------------------|-------------|-------------|
| (6) | Persian ⇒        | Shā. Khalaj | Tel. Khalaj |
|     | /bærg/ ‘leaf’    | /bælg/      | -           |
|     | /sæχt/ ‘hard’    | /sæχt/      | -           |
|     | /tælyχ/ ‘bitter’ | /tælyχ/     | -           |

  

|     |                      |           |             |             |
|-----|----------------------|-----------|-------------|-------------|
| (7) | French ⇒             | Persian   | Shā. Khalaj | Tel. Khalaj |
|     | /plasti:k/ ‘plastic’ | /plastik/ | /plastik/   | /plastik/   |
|     | /film/ ‘movie’       | /film/    | /film/      | /film/      |

The integration of lexical copies into the Khalaj language, almost certainly as a result of the historical contact situation, is evidently influenced by the loss or weakening of sound harmony (Doerfer 1988), as well as by (morpho-)phonemic weakening.

### Discussion and conclusion

This study has drawn a number of conclusions with regard to the lexical copying (loanword) patterns of two Khalaj varieties, namely Telkhāb and Shānegh.

First of all, the data revealed that Khalaj, an endangered and socially dominated language, is a *high borrower* language in Haspelmath & Tadmor's terms, irrespective of the dialect areas, with a copying rate of over 40%, similar to Selice Romani (Elšík 2009), Tarifiyt Berber (Kossmann 2009), and Gurindji (McConvell, 2009). In other words, the data analysis revealed that the findings correlate with similar studies on other endangered languages in asymmetric contact situation. For instance, Selice Romani is an endangered Romani variety spoken by multilingual Hungarian and Slovak speakers in southern Slovakia and northern Hungary, with a copying rate of 62.7%. It only functions as a family language (Elšík 2007, 2009). Elšík (2009) indicated that nouns constituted the highest ratio (75.6%) of lexical copies of all word classes. The lexically-copied verbs (45.1%), adjectives (51.7%), adverbs (50.0%), and function words (30.9%) lagged behind nouns. These findings correlate with the Khalaj data analysis presented in this study. In the Khalaj data, it is the nouns that constitute the highest proportion (Tel.=72.1%, Shā.=78.4%) of lexical copies, followed by verbs (Tel.= 52.9%, Shā.=53.8%), adjectives (Tel.= 34.1%, Shā.=35.0%), adverbs (Tel.= 25.0%, Shā.=25.0%), and function words (Tel.= 34.9%, Shā.=39.6%).

Similarly, Tarifiyt Berber, a language spoken in northeastern Morocco, is among the very high borrower languages, with a copying rate of 51.7%. The Tarifiyt Berber speakers are bilinguals, speaking Moroccan Arabic and their vernacular language in an asymmetrical language constellation. Congruent with the Selice Romani and Khalaj data, it is the nouns that constitute the highest proportion (56.1%) of lexical copies of all word classes, followed by adjectives (52.7%), verbs (44.1%), adverbs (40.0%), and function words (39.5%).

In addition, Gurindji, a severely endangered language of the Pama-Nyungan family, is only spoken by older people in Australia. McConvell (2009) reported that the percentage of lexical copies was 45.6% in Gurindji, with adverbs being the highest copied category, with a ratio of 50.0%, followed by verbs (49.7%), nouns (48.8%), adjectives (40.7%), and function words (15.4%). Contrary to Khalaj, Selice Romani, and Tarifiyt Berber, it is the adverbs that constituted the highest proportions of loan-words in Gurindji.

The common variables of the *very high borrower* languages are sociolinguistic in nature: (i) widespread bi-/multilingualism, (ii) being a socio-politically marginalized minority language, (iii) lacking a standard variety, and (iv) "permissiveness towards copying" (Tadmor 2009a: 58).

Further, the data demonstrated that Khalaj has undergone strong lexical influence from its immediate neighbor, copying over half of the database. The lexical influence of Persian has been especially strong, since it functions as the official language, and medium of instruction. The Arabic language also has exerted a strong influence on Khalaj. A similar case of Arabic influence was also found in a study by Tadmor on Malay-Indonesian (2009b). Many lexical copies of Arabic origin were eventually copied through Persian. Persian thus acted as a "bridge language" for Arabic lexical items in Khalaj.

The data analysis based on the Leipzig-Jakarta list suggested that the number of copied lexical items in the core vocabulary were quite low compared to the mainly cultural items in the database.

There also existed a difference in the number of the lexical copies between the two varieties, probably due to the fact that the language ecology had an impact on the source of items copied. The findings revealed that the Shānegh variety, an enclaved variety in the southern edge of the dialect area, has been influenced by Persian, Arabic, and French. Situated in the vicinity of Azeri-speaking communities in the north, Telkhāb variety, however, has lexical copies taken from Azeri, as well as Persian, Arabic, and French.

As for the implications of findings for contact theory, this study suggests that different percentages of loanwords per word class imply the type of language contact in the sui generis multilingual setting of Khalaj. It is evident that while copying content words is often an expected phenomenon in balanced language contact settings, copying of function words takes often place in relatively “heavy” contact settings such as language mixing and shift. There are LVC copying examples where lexical copying interferes with the morphosyntax in the Khalaj data, signaling relatively heavy contact-induced language change. However, using core vocabulary on a regular basis appears to make such items more invulnerable to copying.

Last but not least, the limited elicitations obtained from four language consultants during the field work limits the generalizability of the findings. Moreover, the present study focused only on the contribution of the lexical copies in the two varieties of the Khalaj language. However, there are five other Khalaj dialect areas, in which other Iranian languages, e.g. Tati and Luri, might exert influence on lexical copies. These factors were beyond the scope of this study. That the age of the speakers could have an effect on whether a loanword is provided for a certain meaning or not (i.e. with younger speakers being more prone to copying) is another important variable to control for in further research.

### Acknowledgements

The study has received no financial support. I wish to thank the anonymous reviewer(s) for their valuable contributions. There are no conflicts of interest to declare.

### Abbreviations

|      |                         |
|------|-------------------------|
| Ar.  | Arabic                  |
| LVC  | Light Verb Construction |
| LWT  | Loanword Typology       |
| M    | Male                    |
| Shā. | Shānegh                 |
| Tel  | Telkhāb                 |
| WOLD | World Loanword Database |

## References

- Bakker, Peter & Mous, Maarten 2013. *Mixed Languages*. Oxford: Oxford University Press.
- Bosnali, Sonel 2016. Değininim durumundaki azınlık dillerinin sürdürümü [Maintenance of minority languages in the case of contact]. *Border Crossing*, 6/2, 292–304.
- Clauson, Gerard 1972. *An etymological dictionary of pre-thirteenth century Turkish*. Oxford: Clarendon Press.
- Dehkhoda, Allameh Aliakbar 1998. *Dehkhoda dictionary*. Tehran: Tehran University.
- Doerfer, Gerhard 1971. *Khalaj materials*. Bloomington: Indiana University Publications.
- Doerfer, Gerhard 1987. *Lexik und Sprachgeographie des Chaladsch* 1–2. Wiesbaden: Harrassowitz.
- Doerfer, Gerhard 1988. *Grammatik des Chaladsch* [Khalaj grammar]. (Turcologica 4.) Wiesbaden: Harrassowitz.
- Doerfer, Gerhard & Tezcan, Semih 1980. *Wörterbuch des Chaladsch (Dialekt von Charrab)* [Dictionary of Khalaj. Dialect of Charrab]. Budapest: Akadémiai Kiadó.
- Elšík, Viktor 2007. *Affix extraction: A case study on Hungarian Romani*. Unpublished PhD Dissertation. Prague: Univerzita Karlova.
- Elšík, Viktor 2009. Loanwords in Selice Romani, an Indo-Aryan language of Slovakia. In: Haspelmath, Martin & Tadmor, Uri (eds.) *Loanwords in the world's languages: A comparative handbook*. Berlin: Walter de Gruyter. 260–303.
- ESTJ = Severtjan, Ėrvand V. 1974. *Ėtimologičeskij slovar' tjurkskix jazykov. Obščetjurkskie i mežtjurkskie osnovy na glasnye* [Etymological Dictionary of Turkic Languages. Common Turkic and Inter-Turkic Vowel Stems]. Moskva: Nauka.
- Family, Neiloufar 2014. *Semantic spaces of Persian light verbs: A constructionist account*. Leiden: Brill.
- Farhady, Hossein & Hezaveh, Fattaneh Sajadi & Hedayati, Hora 2010. Reflections on foreign language education in Iran. *The Electronic Journal for English as a Second Language* 13/4, 1–18.
- Gardner-Chloros, Penelope 2009. *Code-switching*. Cambridge University Press.
- Haspelmath, Martin & Tadmor, Uri (eds.) 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin: Walter de Gruyter.
- Hatami Khajeh, Belal 2013. *İran'da yaşayan Halaç Türklerinin sosyo-kültürel yapısı* [The socio-cultural structure of Khalaj Turks living in Iran]. Unpublished MA Thesis. Ankara: Hacettepe University.
- Haugen, Einar 1950. The analysis of linguistic borrowing. *Language* 26, 210–231.
- Heine, Bernd & Kuteva, Tania 2006. *The changing languages of Europe*. Oxford: Oxford University Press.
- Johanson, Lars 2002. *Structural factors in Turkic Language contacts*. London: Curzon.
- Johanson, Lars 2021. *Turkic*. Cambridge: Cambridge University Press.
- Keller, Rudi 1994. *On language change: The invisible hand in language*. London: Routledge.
- Kıral, Filiz 2000. Reflection on -miş in Khalaj. In: Johanson, Lars & Utaş, Bo (eds.) *Evidentials, Turkic, Iranian and neighbouring languages*. Berlin-New York: Mouton de Gruyter. 89–102.
- Kossmann, Maarten 2009. Loanwords in Tarifit, a Berber language of Morocco. In: Haspelmath & Tadmor eds. 2009: 191–214.
- Lehmann, Winfred P. 1962. *Historical linguistics: An introduction*. New York: Holt, Rinehart & Winston.

- Matras, Yaron & Sakel, Jeanette (eds.) 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter.
- McConvell, Patrick 2009. Loanwords in Gurindji. In: Haspelmath & Tadmor (eds.) 2009: 790–822.
- Pakendorf, Brigitte & Novgorodov, Innokentij N. 2009. Loanwords in Sakha (Yakut), a Turkic language of Siberia. In: Haspelmath & Tadmor (eds.) 2009: 496–524.
- Ragagnin, Elisabetta 2020. Major and minor Turkic language islands in Iran with a special focus on Khalaj. *Iranian Studies* 53/3–4, 573–588.
- Räsänen, Martti 1969. *Versuch eines etymologischen Wörterbuch der Türksprachen* [Essay on an etymological lexicon of Turkic]. Helsinki: Suomalais Ugrilainen Seura.
- Ryding, Karin C. 2005. *A reference grammar of modern standard Arabic*. Cambridge: Cambridge University Press.
- Schmid, Monika S. 2013. First language attrition. *Linguistic Approaches to Bilingualism* 3/1, 94–15.
- Schulte, Kim 2009. Loanwords in Romanian. In: Haspelmath & Tadmor (eds.) 2009: 230–259.
- Siemund, Peter 2008. Language contact: Constraints and common paths of contact-induced language change. In: Siemund, Peter & Kintana, Noemi (eds.) *Language contact and contact languages*. Amsterdam: John Benjamins Publishing Company. 3–11.
- Sims-Williams, Nicholas 2002. Ancient Afghanistan and its invaders: Linguistic evidence from the Bactrian documents and inscriptions. In: Sims-Williams, Nicholas (ed.) *Indo-Iranian languages and peoples*. Oxford: Oxford University Press. 225–242.
- Stilo, Donald L. 1971. *A grammar of Vafsi-Tati: An application of a transformational computer model*. Unpublished PhD Dissertation. Michigan: University of Michigan.
- Tadmor, Uri 2009a. Loanwords in the world's languages: Findings and results. In: Haspelmath & Tadmor (eds.) 2009: 55–75.
- Tadmor, Uri 2009b. Loanwords in Indonesian. In: Haspelmath & Tadmor (eds.) 2009: 686–716.
- Tezcan, Semih 2008. 8. yüzyılda Afganistan'da Halaç prensesleri, günümüzde Orta İran'da Halaçlar ve dilleri [Khalaj princesses in Afghanistan in the 8th century, The Khalaj and their language in Central Iran today]. *Hacettepe Üniversitesi Türkiyat Araştırmaları* 28, 231–252.
- Thomason, Sarah Gray & Kaufman, Terrence 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Weinreich, Uriel 1964 [1953]. *Languages in contact: Findings and problems*. The Hague: Mouton.
- Winford, Donald 2003. *An introduction to contact linguistics*. Oxford: Blackwell Publishing.
- Yarshater, Ehsan 1969. *A grammar of Southern Tati dialects*. Berlin: Walter de Gruyter Mouton.