



LUND
UNIVERSITY

Words without borders:
A study of loanwords in Western Asia

Ravn Kirkegaard

Supervisor: Niklas Erben Johansson

Co-supervisor: Filip Larsson

Centre for Language and Literature, Lund University
MA in Language and Linguistics, General Linguistics
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

June 2023

Abstract

This thesis aims to study the contact situation of the languages Georgian, Northern Kurdish, Turkish, and Western Farsi, through the distribution of loanwords. Many studies on language contact seem not to focus on language contact areas, but rather direct language contact or investigate the contact situation of a single language. The current study, furthermore, wants to highlight the importance of contact areas and the interrelation between languages. Doing so it employs an adapted version of the methodology of the Loanword Typology (LWT) project (Haspelmath & Tadmor, 2009), a project which aimed to study the borrowability of languages on a global scale. Information of borrowed status is collected in all four languages, using a list comprised of 401 concepts. Through the use of dictionaries, each lexical item is categorised as borrowed or very likely not borrowed, and, if borrowed, what donor language the lexical item comes from. All concepts belong to a semantic domain which makes it possible to get a borrowability score for each language, and for each semantic domain. These scores are compared with the results of the LWT project, and analysed separately in order to find which donor languages are significantly represented and how this reflects the social and cultural setting of the contact area. The thesis finds that Northern Kurdish and Western Farsi can be classified as average borrower languages, and Turkish as a high borrower language, when compared to the results of the LWT project. Georgian is excluded from analysis, due to complications in the data. Furthermore, it finds that Arabic is by far the largest donor language, followed by Western Farsi. This also reflects the importance of these two languages in the area historically as high prestige languages. The results are in line with previous research. This thesis contributes with data from languages and a region not part of the LWT project and underlines the importance of studying contact areas and not only global averages and single language contact situations.

Keywords: language contact, loanword typology, borrowings, semantic fields, Western Asia, Georgian, Northern Kurdish, Turkish, Western Farsi.

Acknowledgements

TBA

Contents

List of Figures	5
List of Tables.....	6
Table of languages.....	8
1. Introduction	9
1.1. Aim and research questions.....	9
2. Theoretical background.....	11
2.1. Contact Linguistics: an overview	11
2.2. Types of contact phenomena	13
2.3. Processes of change.....	14
2.3.1. The tree model and the wave model.....	15
2.4. Lexicostatistics and glottochronology	16
2.5. Loanwords and Borrowings	17
2.5.1. What is a borrowing?.....	17
2.5.2. Reasons for borrowing	18
2.5.3. Borrowability and borrowing hierarchies.....	19
3. Methodology	22
3.1. The Loanword Typology project methodology.....	22
3.1.1. Data collection in the LWT project.....	22
3.1.2. Data analysis in the LWT project.....	23
3.2. Language selection	24
3.2.1. Georgian	25
3.2.2. Northern Kurdish.....	26
3.2.3. Western Farsi.....	26
3.2.4. Turkish.....	26
3.3. Word list selection.....	26
3.4. The working process	27
3.4.1. What counts as a loanword.....	28
3.4.2. Synonyms, polysemy, and semantic granularity	29

3.5.	Dealing with dictionaries.....	30
3.5.1.	The dictionaries of this study	30
3.5.2.	Dictionaries in unknown languages.....	30
3.5.3.	Dictionaries in known languages.....	33
3.5.4.	Common problems	34
3.6.	The structure of the dataset.....	35
3.7.	Potential sources of error.....	36
3.7.1.	Issues with the Georgian data.....	36
4.	Results and analysis.....	38
4.1.	Overview of borrowings.....	38
4.1.1.	Borrowings in the semantic domains	38
4.1.2.	Borrowability of word classes	43
4.2.	Donor languages.....	44
4.3.	Overall comparison with the LWT project.....	47
4.4.	Semantic macro groups	50
4.5.	Donor language per semantic domain	52
5.	Discussion	58
5.1.	Recipient languages.....	58
5.2.	Signs of cultural contact	59
5.3.	Methodological implications	61
5.4.	Theoretical implications	63
5.5.	Future research	64
6.	Conclusion.....	65
7.	References	67
	Appendix 1	70
	Appendix 2	75

List of Figures

Figure 1. Illustration of the process of borrowing (arrows) and a borrowing (blue box). The letters indicate proficiency in language A, respectively B.....	14
Figure 2. Illustrations of the tree model and the wave model of language change, repurposed from Carling (2019, p. 18). In the tree model we can see a feature A being kept in one branch and changed to B in another, which in turn develops into feature C, which is kept in one branch and further developed into feature D in the last two branches. In the wave model the same features are present but we can see that they mostly cluster together, indicating that contact might be involved.....	16
Figure 3. Map displaying where the languages in this study are spoken. Turkish in green, Western Farsi in purple, Northern Kurdish in red, and Georgian in yellow. Note that this is an estimate and both Turkish and Western Farsi are spoken also outside of Turkey and Iran.	25
Figure 4. Entry hêstir in Chyet (2003, p. 251).....	28
Figure 5. Entry tane (grain,piece) in Nişanyar Sözlük dictionary.	31
Figure 6. Entries for kalp (heart) in Nişanyar Sözlük dictionary.	32
Figure 7. Entry ხაღაღი (shirt) in the Georgian Explanatory Dictionary.....	32
Figure 8. Etymology of Persian word ârâm آرام ‘quiet, calm’ (Nourai, 2013, p. 390).....	34
Figure 9. Bar plot comparing the highest to lowest loanword ratio between all languages. N.B. that the semantic domains are ordered by percentage of borrowings and not in alphabetical order, the bars do not represent the same semantic domains but only the relation between them in terms of borrowability.	43
Figure 10. Bar plot comparing the borrowability score of the present study and the LWT project.....	49

List of Tables

Table 1. Showing the borrowing scale of Thomason and Kaufman (1988).....	20
Table 2. Percentage of borrowings per semantic domain, as found by the LWT project (Haspelmath & Tadmor, 2009, p. 64).....	21
Table 3. Dictionaries used in the thesis, with references.....	30
Table 4. Columns in the dataset and their explanations.	35
Table 5. The percentage and number of borrowings and total word forms in the dataset per language and the region as a whole.	38
Table 6. Number of borrowings and total count per semantic domain in the region.	39
Table 7. Number of borrowings per semantic domain in Northern Kurdish.....	40
Table 8. Number of borrowings per semantic domain in Western Farsi.....	41
Table 9. Number of borrowings per semantic domain in Turkish.....	42
Table 10. Borrowings per word class.	44
Table 11. Donor language overall contribution of borrowings.	45
Table 12. Donor language contribution to Turkish.	46
Table 13. Donor language contribution to Western Farsi.	46
Table 14. Donor language contribution to Northern Kurdish.	47
Table 15. The languages of this study entered into the overall language results of the LWT project (Haspelmath & Tadmor, 2009, p. 64). Highlights indicate the languages of this thesis.	48
Table 16. Division of semantic domains from LWT into semantic macro groups.....	50
Table 17. Borrowings per semantic macro group and recipient language.	51
Table 18. Each donor language's contribution to the recipient languages per semantic macro domain. Blank cells indicate 0.00%. None represent the percentages of words that are not borrowed.	52
Table 19. Division of donor languages into groups based on geography and genealogy.....	53

Table 20. Contribution of language families to Northern Kurdish in terms of borrowings per semantic domain. Highlights indicate highest row percentage.....	54
Table 21. Contribution of language families to Turkish in terms of borrowings per semantic domain. Highlights indicate highest row percentage.	55
Table 22. Contribution of language families to Western Farsi in terms of borrowings per semantic domain. Highlights indicate highest row percentage.....	56
Table 23. Borrowings for each language with respect to donor language family and semantic macro groups. Highlights indicate highest percentage per semantic macro group (lack of loans excluded). None the percentage or words that are not borrowed.....	57

Table of languages

<i>Language</i>	<i>Glottocode</i>	<i>Language family</i>	<i>Language branch</i>
<i>Arabic</i>	arab1395	Afro-Asiatic	Semitic
<i>Armenian</i>	arme1241	Indo-European	Armenic
<i>Avestan</i>	aves1237	Indo-European	Indo-Iranian
<i>Azerbaijani</i>	nort2697	Turkic	Oghuz
<i>Bulgarian</i>	bulg1262	Indo-European	Balto-Slavic
<i>Central Kurdish</i>	cent1972	Indo-European	Indo-Iranian
<i>Chagatai</i>	chag1247	Turkic	Turkestan Turkic
<i>Chinese (Mandarin)</i>	mand1415	Sino-Tibetan	Sinitic
<i>Cuman</i>	cuma1241	Turkic	Kipchak
<i>English</i>	stan1293	Indo-European	Germanic
<i>French</i>	stan1290	Indo-European	Italic
<i>Georgian</i>	nucl1302	Kartvelian	Georgic
<i>Greek</i>	gree1276	Indo-European	Graeco-Phrygian
<i>Hebrew</i>	hebr1245	Afro-Asiatic	Semitic
<i>Hungarian</i>	hung1274	Uralic	Hungarian
<i>Italian</i>	ital1282	Indo-European	Italic
<i>Latin</i>	lati1261	Indo-European	Italic
<i>Mongolian</i>	mong1331	Mongolic	Central Mongolic
<i>Northern Kurdish</i>	nort2641	Indo-European	Indo-Iranian
<i>Old Persian</i>	oldp1254	Indo-European	Indo-Iranian
<i>Pahlavi</i>	pahl1241	Indo-European	Indo-Iranian
<i>Proto-Iranian</i>	iran1269	Indo-European	Indo-Iranian
<i>Russian</i>	russ1263	Indo-European	Balto-Slavic
<i>Sanskrit</i>	sans1269	Indo-European	Indo-Iranian
<i>Serbian</i>	sout1528	Indo-European	Balto-Slavic
<i>Sogdian</i>	sogd1245	Indo-European	Indo-Iranian
<i>Syriac</i>	clas1252	Afro-Asiatic	Semitic
<i>Turkish</i>	nucl1301	Turkic	Oghuz
<i>Western Farsi</i>	west2369	Indo-European	Indo-Iranian
<i>Zaza</i>	zaza1246	Indo-European	Indo-Iranian

1. Introduction

Studying lexical borrowing has often been used in tracing language genealogy by identifying vocabulary which is resistant to borrowing, i.e. *core vocabulary*, through the use of the lexicostatistic method introduced by Swadesh (1952). This method, although heavily used, has also received a fair share of criticism, since Swadesh relied on his intuition of what could be considered universally important concepts, and not on any actual empirical research (Haspelmath & Tadmor, 2009).

The most recent, large-scale study of loanwords is the Loanword Typology project (LWT) (Haspelmath & Tadmor, 2009), set out with the explicit goal of making a base reference concerning different types of loanwords and providing genuine generalisations across languages, and to create a list of borrowing resistant lexical concepts based on their findings. They studied the composition of vocabulary pertaining to several semantic domains using a crosslinguistic sample of 41 languages and based on this drew conclusions on the borrowability of the different semantic domains.

Although the LWT project is a highly influential crosslinguistic study of loanwords and their impact on language vocabulary, it is much like previous loanword research based on individual language contact situations and on direct language contact. The focus of the present study, conversely, is to investigate a multilingual language contact situation of Western Asia, in order to better understand the relation between deep historical, cultural, and religious ties, and the composition and convergence of vocabulary. The focus is thus not necessarily to study any one language in particular, but rather to paint a picture of a region as a whole.

With this in mind, Western Asia becomes an interesting linguistic contact area to study because there has both been long-time language contact in the region with several large languages being dominant in different areas, geographically, culturally, religiously, and linguistically, while constantly co-existing, and there has not, to the extent of my knowledge, been any studies made on this specific contact area.

The thesis attempts, therefore, to fill this gap by studying the loanword composition of Georgian, Northern Kurdish, Western Farsi, and Turkish. These languages are chosen because of availability of existing etymological data, their being typologically very distinct (with the exception of Western Farsi and Northern Kurdish, both being Indo-Iranian languages), their historic importance in the area, and the fact that they are not part of the Loanword Typology project.

1.1. Aim and research questions

The main interest of this thesis is to study whether a linguistic contact area with an explicitly shared history behaves similarly to the generalisations drawn from the larger crosslinguistic studies on loanword typology, notably the LWT project. The aims are thus, (1) to provide a database of vocabulary of the languages in the area, and (2) to investigate whether the area follows crosslinguistic tendencies

of borrowing, as well as adding data from several languages not studied in the LWT project. In order to achieve these aims, the following research questions are posed:

1. How do the languages Northern Kurdish, Western Farsi, Turkish, and Georgian compare to the languages analysed by the Loanword Typology project?
2. Which are the main donor languages in the region?
3. What signs of linguistic and/or cultural contact can we see based on attested loanwords? That is, what can we say about any potential correlation between donor language and semantic domain or number of borrowings vis-à-vis recipient language?

The first research question relates my findings to that of one of the major works in the field, providing support or possibly extending the analysis of loanword typology to more local, but yet multilingual, contact situations. It could potentially also give a measure to the contact intensity of the area. The second research question intends to give indications of the main prestige languages in the area as well as to see whether there are some donor languages more prominent in some semantic domains than others. The third research question can aid our understanding of the cultural relation between the languages, acting as a linguistic addition to our understanding of the region through historical records and more contemporary cultural and political relations.

In order to answer these research questions, I will employ an adapted version of the methodology proposed by the LWT project. This is for two primary reasons, the first being the already existing and considerably large database it provides, making it possible to both draw from existing research, and add meaningful data that has not yet been studied. The second being that it provides a solid framework for collecting and analysing relevant data, that is, to build a proper database of language vocabulary, and it also provides a list of lexical concepts ready to use. The adaptation is necessary to account for the range of languages, limited access to material, and the time scope of the thesis. Some further analyses using variations of the semantic domains, as well as grouping donor languages by geographical and genealogical affiliations, is used to aid in answering the third research question.

2. Theoretical background

The aim of this chapter is to give some background of the different theories concerning borrowing and language contact, as well as explaining core concepts and notions of the field and related topics. Firstly we discuss contact linguistics, as borrowing ultimately is an expression of language contact. Secondly, we enter the domain of historical linguistics, where we touch upon the processes behind change, different models, and methodologies within the field. In the next subchapter loanwords and borrowings are explained according to both foundational and more contemporary approaches, what defines borrowings, and the current understanding of borrowability. Lastly, short descriptions of the languages relevant for this thesis are presented.

2.1. Contact Linguistics: an overview

Firstly, we should note that languages interacting with each other is a natural process, and unavoidable in states of bilingualism. Bilingualism is the capacity of an individual to use two or more languages (Myers-Scotton, 2002). I will in this thesis use both the terms bilingualism and multilingualism to mean the phenomenon of being able to use two or more languages. Terms such as *mixing*, as well as *interference* have been used to denote contact-induced change, but these terms are not used in this thesis, unless in reference to established theories and terminology. It is practically impossible to find a language that has not, to some extent, been influenced by other languages, and as such the idea of a “pure” language, a language free of any “foreign” structures or forms, have limited to no bearing for scientific research (Haugen, 1950; Matras & Bakker, 2008). It has, however, been extensively used for a very specific type of contact languages often called *mixed languages*, in which there is no single genealogical ancestor as the language draws equally from two or more different languages (Bakker & Matras, 2013). As part of their seminal work on language contact Thomason and Kaufman (1988) use the term *interference* to cover two types of contact-induced changes, namely *borrowing* and *shift*, where the latter deals with the complete conversion to another language, and the former refers to the slow introduction of new elements into a language.

It has, furthermore, been argued by Matras (2020) that language contact ought to be considered a language internal factor of language change. The reasoning behind this view is that language contact is predicated on the multilingual interaction between individuals, in which the speakers use their complete multilingual language repertoire in order to achieve some communicative goal given certain social expectations, dictated by the socio-cultural situation of the speakers. Contact-induced change is therefore a product resulting from the creative use of a speaker’s full linguistic repertoire in a multilingual setting (Matras, 2020, p. 6). Central to this idea is that languages are not truly separate entities within a speaker but intermingling and only actually separated given the social expectations of other speakers. The effect of using the full range of the language repertoire will result in a loosening of social control of language use and an eventual acceptance of variation caused by this interaction between

languages within speakers' multilingual language repertoires, leading eventually to large-scale language change. This is somewhat similar to the approach taken by Myers-Scotton (2002), who argues that language contact is a very real phenomenon, in the sense that languages are in contact, rather than just speakers being in contact, since different languages co-exist and influence each other within the mind of a multilingual speaker. Myers-Scotton (2002, p. 5) writes that "some linguists like to say that to speak of 'language contact' is erroneous, because it is the speakers who are in contact, not the languages." Myers-Scotton herself, however, disagrees with this point of view, holding that languages are in contact with each other in the mental representation of the linguistic knowledge of the speakers. To make it a bit clearer, language contact entails that individual speakers have knowledge of more than one language and actively use that knowledge with others. Whether this leads to language change depends on further factors, such as societal bilingualism, the language communities' acceptance of the use of part or the full range of the population's language repertoires (Matras, 2020). As a result of this definition of language contact and contact-induced change, such changes in the lexical and grammatical structure of a language will be considered, not an *external* but rather an *internal* process.

Here it also becomes important to distinguish the difference between the synchronic processes and the diachronic processes of language contact. Intensity of language contact is tied to the extent of bilingualism in a linguistic community, and as such bilingualism is a very important part of language contact. While bilingualism is not necessary for contact-induced phenomena, such as lexical borrowings, to appear, extensive bilingualism will facilitate borrowing into a language. Contact-induced changes are thus often, although not always, part of a bilingual process in which the speaker is in constant negotiation "[...] between the need to maintain demarcation boundaries among subsets of their repertoire in order to satisfy expectations on communicative behaviour, and the urge to make use of the full [bilingual] repertoire" (Matras, 2020, p. 41). This behaviour essentially results both in the ongoing contact-phenomena studied in bilingual research, and the structural changes of an entire language or language variety, studied by contact linguistics. Bilingual research may, of course, include the study of language contact and contact linguistics, but the main difference tends to manifest itself in the object of study. Bilingualism focusses mostly on the speakers, while contact linguistics takes the structural aspects of lexicon and grammar as its main target (Myers-Scotton, 2002).

To sum up, we can say that language contact is facilitated by language acquisition and societal bilingualism, and lead to the introduction of borrowings into the language. On a synchronic level this can often be seen in the behaviour of bilingual speakers and the spread of loanwords through a language community, and on a diachronic level it manifests as sedimented structures and forms. Furthermore, this entails that change resulting from language contact is determined by the behaviour of the linguistic community and their speakers. This is also a cornerstone in the theory developed by Thomason and Kaufman, who state that "[...] it is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcome of language contact" (Thomason & Kaufman, 1988, p. 35). This is an important aspect of the impact contact linguistics and historical

linguistics have on our understanding of history, since, if language structure were the primary source of language change, we would be hard pressed to motivate any linguistic argument to be used at all for research within the field of history.

2.2. Types of contact phenomena

Contact-induced changes can come in many different forms, ranging from lexical borrowing to the complete restructuring of the grammar and lexicon of a language, and depend on a multitude of factors, such as prestige, intensity of contact, and typological distance. Interestingly, although the effects of language contact are readily observed and well-documented, there are many different theories regarding the reasons for why the changes occur at all and the process they undergo. It is not, for example, clear to what degree typological distance (differences in language patterns, be they grammatical, semantic, phonological, lexical, or otherwise) plays a role in determining the types of changes experienced, or if prestige can satisfactorily explain borrowings, nor is it entirely understood what the limitations of borrowing events are, or how extensively a language might be affected by language contact. This might have been grounded in the way research in this field has been conducted since, as Muysken (2000) points out, much effort has been put into understanding the visible outcomes of language contact, rather than on the process through which these results come to be. Hence, in this section we will shortly discuss the different types of contact phenomena and the processes behind these.

The first observation we need to make is the difficulty of not conflating the result and the process of borrowing. *Borrowing*, is the process, while *borrowings*, in plural, is the end product of the process. The process is the manner in which a borrowing enters and spreads through a language. It is somewhat unclear as to how this process actually ought to be understood, but Matras has proposed it, essentially, as communication with the intention of achieving some social goal (Matras, 2020), and it seems that it is a process deeply rooted in social behaviour (cf. Thomason & Kaufman, 1988). This process, then, occurs in the interaction between people and might lead to the establishing of a structure or form as a borrowing in a linguistic community. Borrowings or loans are the structures and forms that become part of the linguistic community and, over time, adopted also by speakers not familiar with the original language. Borrowings are thus not the same as *code-switches*, the phenomenon of alternating between languages or inserting elements of one language into another during the same discourse (Myers-Scotton, 2002). Code-switches require knowledge of both languages while borrowings have been adopted by the language community in a broader sense. This is illustrated in Figure 1, where the two boxes represent different languages and the circles different communities within each language. The arrows represent the process of borrowing, i.e. how the words spread through and between populations, while the blue box represents a linguistic unit that is being borrowed.

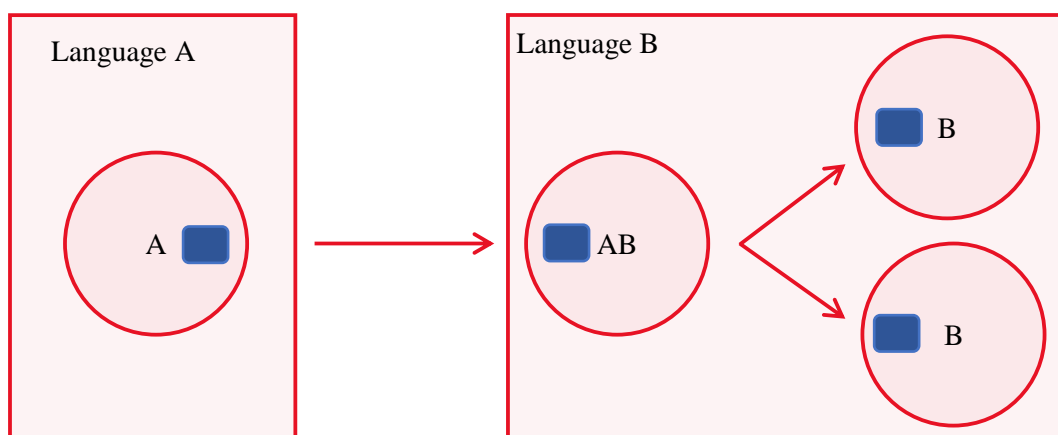


Figure 1. Illustration of the process of borrowing (arrows) and a borrowing (blue box). The letters indicate proficiency in language A, respectively B.

The idea behind the illustration is that the process *borrowing* is predicated on behaviour of individuals when interacting with others, while the *borrowings* are actual linguistic units that one can observe in the language. Furthermore, it becomes clear that code-switching between language A and B can only occur in group AB.

Some important contact processes except borrowing, and code-switching are *shift*, *language acquisition*, *convergence*, and *relexification*. Shift and language acquisition are related since no shift could occur without learning a language in the first place. Shift, however, also indicates a population's switch to a *target language* as their main communicative medium, abandoning the previously dominating language. The term target language is used to denote the language a person seeks to acquire. Convergence and relexification are specific kinds of borrowing with certain conditions attached to them. Convergence represents a bidirectional influence while borrowing tends to be unidirectional (cf. Matras, 2020; Muysken, 2000). Relexification is an extreme form of lexical borrowing that affects large parts of the vocabulary and replaces vocabulary rather than adding to an already existing lexicon (Muysken, 2000).

In the theory of Thomason and Kaufman (1988) language acquisition is a central part of contact situations they call *language shifts* and borrowing, conversely, the main process in *language maintenance* situations. Muysken (2000) agrees with the aforementioned, stating that shifting from one language to another and borrowing are the two most important processes in a language contact situation.

2.3. Processes of change

There are a few processes of language change, not necessarily contact-induced, that are important to historical linguistics, those being *divergence*, *convergence*, *advergence* or *parallel drift*. We have talked about convergence as a contact-induced process of change since this is not likely to happen spontaneously without interaction with and acquisition of other languages, although there may be other causes for convergence to take place (Carling, 2019). Advergence, sometimes also called parallel drift,

is similar to convergence in that it denotes the joint development of two language varieties, that continues to be exposed to each other, and yield similar contact situations as convergence. Usually, however, advergence applies to language varieties with a common ancestor (Renfrew, 2000). Divergence is, as it sounds, the development of a language into varieties with increasing internal differences, eventually leading to a language split into new daughter languages. This is a process present whenever speakers of a language become geographically isolated from each other (Renfrew, 2000).

2.3.1. The tree model and the wave model

Since the purpose of historical linguistics is to chart the change of languages, historical linguistics shares in some sense a connection with *phylogenetics*. Usually phylogenetics refers to the study of evolution and genetic relation between organisms, but interpreted a bit more broadly one could see it as the study of the development and evolution of an object over time, be this object an organism, technological innovation, or language (Dunn, 2014). The method of linguistic phylogenetics provides an important way to “test hypotheses about human dispersals, processes of cultural change, and the evolution of other linguistic subsystems” (Dunn, 2014, p. 190), and he further argues that the aims of the approach are more ambitious than that of regular historical linguistics, which tend to focus on the history of languages rather than on the social and cultural environments related to language change. Although this might be the case, the linguistic phylogenetic approach capitalises on the existence of language phylogenies, which in turn is related to the process of divergence and the *tree model*, rather than on the process of convergence, often related to the *wave model* (Carling, 2019; Renfrew, 2000). The two models are illustrated in Figure 2, and as we can see the tree model shows splits at different stages between which some change occurs, while the wave model shows how the changes cluster together. In cases of convergence, languages tend to become more similar over time as a result of language contact (Renfrew, 2000), which can present issues when identifying language phylogenies. Since the aim of the study in this thesis is to investigate a language contact area, it becomes less interesting to look at language phylogenies and more interesting to look at processes of convergence and advergence and how linguistic features move according to the wave model.

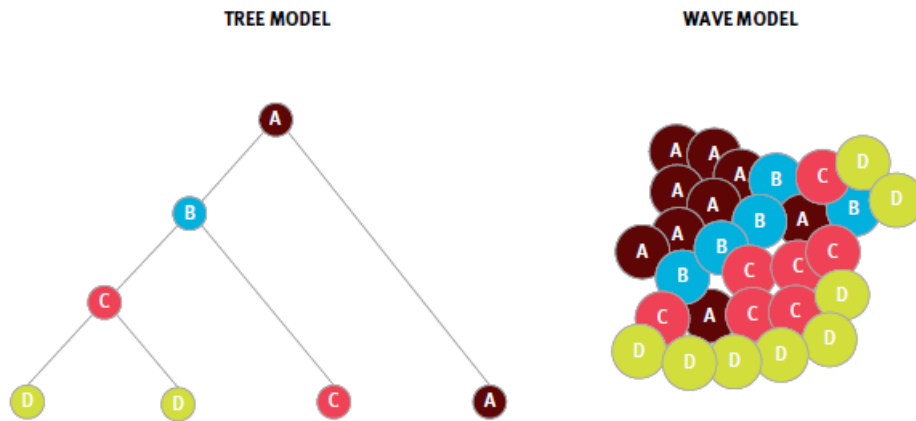


Figure 2. Illustrations of the tree model and the wave model of language change, repurposed from Carling (2019, p. 18). In the tree model we can see a feature A being kept in one branch and changed to B in another, which in turn develops into feature C, which is kept in one branch and further developed into feature D in the last two branches. In the wave model the same features are present but we can see that they mostly cluster together, indicating that contact might be involved.

2.4. Lexicostatistics and glottochronology

The methods *lexicostatistics* and *glottochronology* are based on the fact that languages change over time and that if we can measure the rate of change, we can also figure out the age of a given language and thus provide both family trees and dates for language splits (Dunn, 2014; Swadesh, 1952). In order to do so, Swadesh created a list of a hundred words, consisting of what he considered to be culturally universal and therefore particularly resistant to borrowing. This would mean that these words, also called *basic* or *core vocabulary*, are more likely to be inherited and therefore also change at a slower pace than other parts of the lexicon. The methods have been criticised on many occasions and for various reasons, for the lack of empirical backing of the words in the list (Haspelmath & Tadmor, 2009; Pereltsvaig & Lewis, 2015), for its inconsistency in classifying language genealogy (Pereltsvaig & Lewis, 2015), for the difficulty in determining what word form should be entered for each lexical concept (Pereltsvaig & Lewis, 2015), and for assuming a constant rate of change in lexicon (Dunn, 2014). This is not to mean that lexicostatistical data with basic vocabulary, however not necessarily those created by Swadesh, has not been successful. It has been argued that lexicostatistical methods consistently produce reliable language family trees, comparable to those created by the Comparative Method (Carling, 2019).

Some of the critiques of lexicostatistical methods (which also include phylogenetic methods mentioned above), have been addressed by the research community. For example, dealing with the issue of how Swadesh developed his list of core vocabulary has been one of the main purposes of the Loanword Typology project and they could produce from this a list of concepts that tend to crosslinguistically be very rarely borrowed (Tadmor et al., 2010). The most difficult critique to address is the difficulty of the considerable variation in semantic typology around the world. The methods use a glossing language, also called meta-language, and a large part of the researcher's task is to figure out what word is equivalent to the concept in the glossing language. This is seldom as straightforward as one might first think, as the semantic typology of languages vary considerably. The semantic granularity

of words in different languages often diverge, or a concept might exist in one language but not another, and therefore studying anything related to semantics using another language can be problematic.

2.5. Loanwords and Borrowings

As we have discussed the emergence of borrowings is a common outcome of language contact, together with language shifts, and can give many clues as to language genealogy as well as to understanding the socio-cultural situation of people in the past. It is therefore worth to take some time reflecting on what borrowings and specifically loanwords actually are and how they come about. We also discuss the Loanword Typology (LWT) project and its impact on the field of historical linguistics and contact linguistics.

2.5.1. What is a borrowing?

This seemingly innocent question is fraught with complications, regarding both what defines a borrowing, and not in the least the somewhat problematic connotations of “to borrow”. In Haugen’s influential work on borrowings, he states that “the metaphor implied is certainly absurd, since the borrowing takes place without the lender's consent or even awareness, and the borrower is under no obligation to repay the loan. One might as well call it stealing, were it not that the owner is deprived of nothing and feels no urge to recover his goods” (1950, p. 211). The term is, however, rarely misunderstood and is in frequent use in the literature. The language into which the borrowing is reproduced is often called the *recipient language* and the source language *donor language*.

Loanwords are usually defined in terms of reproduction in one language of patterns found in another (cf. Haspelmath & Tadmor, 2009; Matras, 2020; Thomason & Kaufman, 1988). Perhaps more importantly than defining a borrowing, is to realise that the study of borrowings necessarily has to begin with the behaviour of bilingual speakers (Haugen, 1950), and thus to ground theories on borrowing as a process rather than as an outcome (Matras, 2020). Of course, an important aspect of understanding the process is investigating the structure and composition of the outcomes, as has been done by many scholars, especially in lexicostatistics and phylogenetics (cf. Carling, 2019; Carling et al., 2019; Comrie, 2000; Haspelmath & Tadmor, 2009; Shijulal et al., 2011; Van de Velde & De Smet, 2021; van Hout & Muysken, 1994). The difference lies in that studying the results of borrowing events should be complementary to our understanding of how bilingual interactions work. For example Matras (2020) choose to define borrowing as the process in which linguistic matter and pattern replication occurs. Linguistic matter includes lexemes and morphemes, and linguistic patterns denote structures in which matter is organised. Replication is another word for the reproducing or adoption of the two linguistic types into a recipient language. Other terms, such as *loans* and *transfers* are also in frequent use. In this thesis borrowing will refer to the process, while borrowings or a borrowing will refer to the linguistic

feature moving from one language into another. Loanwords and lexical borrowings will refer strictly to the borrowing of lexical units.

2.5.2. Reasons for borrowing

Although it is not always easy to determine or predict the directionality of borrowings or, necessarily, to understand the processes underlying borrowing, there are some often discussed reasons for why borrowing happens. Usually these are mentioned as (1) cultural gaps, (2) prestige, and (3) dominance (Carling et al., 2019; Epps, 2014; Haspelmath & Tadmor, 2009; Matras, 2020; Myers-Scotton, 2002; Thomason & Kaufman, 1988). Even though these are frequently mentioned they have also been criticised and there are some potential issues related to their use. There are some other ideas more directed towards the process of how borrowings enter a language, notably so an approach presented by Matras (2020).

Perhaps the most obvious reason for borrowing is to fill so called cultural gaps. However, as Haspelmath and Tadmor (2009) mention, since all languages have the option of creating new words from the resources available within them, there is no actual need to borrow the term attached to the innovation. In a study by Brown et al. (1994), both of these strategies are found. Their findings show that Native American languages spoken adjacent to Spanish tend to have a much higher rate of Spanish borrowings for newly introduced items, while those spoken in the neighbourhood of English tend to create neologisms instead. There is thus no inherent need to borrow words, but it also seems that there is no inherent motivation to not borrow either.

One of the reasons for not inventing new words, instead of importing them, is prestige. In some situations there is a wish for one group to allude to or assume the identity of a socially more prestigious group, and one way of doing this is to use the same words, grammar, or way of speaking as that group. This can be a powerful reason for borrowing events to take place, but it does not explain why, for example, a less prestigious language sometimes influences a more prestige language (Thomason & Kaufman, 1988).

Similar issues are related to the notion of dominance, since even though a more dominant language community can exert strong influence on less dominant language group, not uncommonly in situations such as military invasion and colonisation (Myers-Scotton, 2002), it cannot explain why smaller groups can have a strong influence on larger more powerful groups. Some interesting studies regarding these notions have been made and even though they cannot explain all situations they account for many (cf. Carling et al. 2019).

To sum up, although there are many situations in which the more socially, politically, or militarily powerful group is the donor of borrowings, this is not always the case and therefore can neither need or prestige be a truly reliable indicator of the donor language's status in relation to the recipient language (Matras, 2020; Thomason & Kaufman, 1988). Furthermore, it is very important indeed to take note of the fact that all languages borrow, regardless of their status in relation to other languages (Carling

et al., 2019). It has also been argued that in the wake of globalisation issues of dominance and prestige have become much more complicated than they have been historically (Matras, 2020).

2.5.3. Borrowability and borrowing hierarchies

Although it is difficult to make predictions about the directionality of borrowings, there is concrete evidence that some semantic domains are more likely to be borrowed than others and there are some suggested borrowing hierarchies in relation to word classes. It has also remained unclear to what extent typological differences affect what is borrowed, but it seems that social factors are more important than structural ones (cf. Brown et al., 1994; Thomason & Kaufman, 1988). In the work of Thomason and Kaufman (1988), it is argued that the interference conferred by structural differences languages between is directly correlated with the intensity of the contact situation. Lower intensity indicates greater typological interference while in high intensity contact situations, typology has no effect at all and there are no limits to what may or may not be borrowed. One could perhaps still argue that there are certain constraints even in these situations if one allows to draw information from *mixed languages*, i.e. languages that have no clear genealogy and cannot be considered a descendent from a singular ancestor language and arise in situations of extremely intensive contact situations. It seems, however, that these languages tend to either borrow their nouns from one language and their verbs from the other language, or to borrow the entire grammatical structure from one parent language and their lexicon from the other (Bakker, 2008). So in these, admittedly rare, occasions borrowing or shifts seems to follow certain restrictions, that is the “mixing” is not random.

There is, however, some good evidence to support the borrowability of syntactic features, word classes and semantic domains. Since the thesis revolves around lexical borrowings, we will not dwell long on the borrowability of syntactic features but focus on the lexical elements. There has been a multitude of different borrowing hierarchies concerning lexical units, and while most look similar there are some differences.

Thomason and Kaufman (1988) take a holistic view and suggest a borrowing scale which stands in direct relation to the level of contact intensity, see Table 1.

Table 1. Showing the borrowing scale of Thomason and Kaufman (1988).

<i>Contact intensity</i>	<i>Features borrowed</i>
<i>Low to casual contact</i>	Content words (e.g. nouns and verbs)
	Conjunctions, adverbial particles
	Minor phonological, syntactic, and semantic features
<i>Moderate contact intensity</i>	Adpositions, pronouns, low numerals
	Derivational and inflectional affixes
	Slight structural interference
<i>High contact intensity</i>	Structural features
	Change in typological nature
<i>Very high contact intensity</i>	Heavy structural borrowing
	Considerable disruption of typology

Haspelmath and Tadmor (2009) in their project provide some interesting findings as well, firstly content words are borrowed far more than function words. Furthermore, nouns are borrowed to a larger extent than adjectives and adverbs, and verbs are borrowed less than adjectives and adverbs. Matras (2007) provide a more detailed hierarchy,

nouns, conjunctions >> verbs >> discourse markers >> adjectives >> interjections >> adverbs >> other particles, adpositions >> numerals >> pronouns >> derivational affixes >> inflectional affixes.

Interestingly, this hierarchy disputes the two previously mentioned ones. Firstly, in that it puts conjunctions together with nouns and before verbs, which is in direct conflict with the one given by Thomason and Kaufman (1988), and he also puts verbs before adjectives and adverbs. There are, however, some common currents in these hierarchies. Nouns are clearly the most borrowed, followed by verbs, adverbs, and adjectives, in uncertain order, then comes pronouns, and last affixes.

There is not only structure to borrowing in relation to the grammatical category of a word, but perhaps even more clearly in terms of semantic domain. One of the huge achievements of the Loanword Typology project (Haspelmath & Tadmor, 2009), is the large crosslinguistic investigation of borrowability in relation to semantic domains. The amount of borrowed terms per semantic domain is presented in Table 2.

Table 2. Percentage of borrowings per semantic domain, as found by the LWT project (Haspelmath & Tadmor, 2009, p. 64).

<i>Semantic field</i>	<i>Loanwords as % of total</i>
<i>Religion and belief</i>	41.2%
<i>Clothing and grooming</i>	38.6%
<i>The house</i>	37.2%
<i>Law</i>	34.3%
<i>Social and political relations</i>	31.0%
<i>Agriculture and vegetation</i>	30.0%
<i>Food and drink</i>	29.3%
<i>Warfare and hunting</i>	27.9%
<i>Possession</i>	27.1%
<i>Animals</i>	25.5%
<i>Cognition</i>	24.2%
<i>Basic actions and technology</i>	23.8%
<i>Time</i>	23.2%
<i>Speech and language</i>	22.3%
<i>Quantity</i>	20.5%
<i>Emotions and values</i>	19.9%
<i>The physical world</i>	19.8%
<i>Motion</i>	17.3%
<i>Kinship</i>	15.0%
<i>The body</i>	14.2%
<i>Spatial relations</i>	14.0%
<i>Sense perception</i>	11.0%
<i>All words</i>	24.2%

What can be observed from the data is that the fields with a higher borrowability are domains usually subjected to influence from other cultures. Many of them are for example related to powerful institutions, such as religion, law, and socio-political relations, or technological innovations, with clothing, the house, agriculture, food, and warfare coming to mind. The fields with a lower borrowability are related to concepts considered more universal among human cultures, and since most languages have lexical units to express these concepts, there is no exigent need to borrow them from other languages, for example kinship terms, body parts, spatial relations, and senses.

3. Methodology

The starting point of the Loanword Typology project (Haspelmath & Tadmor, 2009) was the experienced need of large crosslinguistic studies on the borrowability of loanwords (Haspelmath, 2008). Morris Swadesh has been hugely influential in the study of vocabulary with his list of core vocabulary. The list, and the notion of core vocabulary has been used by a multitude of studies, often in order to study the genealogical history of a language, or as a basis to create language relations or establish families. It is, however, problematic since it is based on Swadesh's own intuition on what cultural concepts could be considered universal and not on any crosslinguistic empirical study of the vocabulary composition of languages. One of the goals of the LWT project is to rectify this and create a *basic vocabulary list* consisting of the least borrowable words in a large sample of different languages. The project uses a modified version of the Intercontinental Dictionary Series, creating a list of 1460 lexical concepts. The lexical concepts come from 24 distinct semantic domains, established by Buck (1949), and the languages in the study are from different genealogical backgrounds and geographic positions, as to create as even a distribution of languages as possible and avoiding geographically close languages since this could bias the borrowings of a certain area or language. In total 41 languages are part of the study.

The LWT project provides a good methodology for investigating loanwords and borrowability of concepts and semantic domains. The present study uses this methodology but, despite this, it departs slightly from LWT project in terms of aim and therefore some adaptations are necessary. One of the aims of the LWT project was to create a framework from which one can depart when studying borrowability, something this thesis uses in order to study the regional language contact situation concerning the languages Western Farsi, Northern Kurdish, Georgian, and Turkish, none of which are part of the LWT project. Furthermore, this thesis aims to study the cultural contacts through their influence on language lexicon, and to see which languages have contributed the most to the lexicons of the languages studied.

3.1. The Loanword Typology project methodology

The LWT project follows a particular methodology for mapping the loanwords in the languages of their study. The broad idea is to collect a large number of datapoints, words in this case, from a selection of languages chosen to be as representative as possible of the diversity of the world's languages and analyse the borrowability of these words and their respective semantic domains. It is one of the largest studies on loanwords to date and provides an interesting point of reference.

3.1.1. Data collection in the LWT project

The LWT project realised some of the restrictions this puts on the gathering of data, and the contributors were therefore encouraged to remove or add word forms in those cases where a language has a broader or narrower semantic representation of the concept given in English. If a word form could not be found

in the language, a note was taken indicating the reason, which could be either lack of information, that the meaning was irrelevant to the speakers, or that there simply was no corresponding meaning.

When collecting the data, the researchers were instructed to find out what the corresponding word form in their language the concept was represented by, together with a gloss, which should show whether or not the word was analysable, its borrowed status, and the earliest attestation or reconstruction of a word form. The gloss is important in that a morphologically analysable word is not considered borrowed in the approach outlined by the LWT project, but rather a language internal construction. The borrowed status was given a number, 0 through 4, in order to account for the uncertainty of whether a word might be borrowed or not. A score of 0 means no evidence of borrowing, 1 very little evidence, 2 could be borrowed, 3 likely borrowed, and 4 certainly borrowed (Haspelmath & Tadmor, 2009). Note that there is no conclusive method to determine that a word has not *ever* been borrowed, and therefore there is no score for certainly not borrowed. Related to this was the information on earliest attestation of a word form. This information is important for a number of reasons but perhaps more particularly because it gives some indication of the possible borrowability of a word. If a word has been present in a language for a very long time and is not borrowed, that would point to the word, or concept, to be more resistant to borrowing. Apart from this, optional information could be given about the original script, information about the meaning of a word form (if it differs significantly from the LWT concept), calquing, if some parts of the morphology were borrowed, frequency of word form, and which register it belonged to.

When a word form was indicated as borrowed, researchers needed to give additional information. For all loanwords the researchers needed to indicate (1) the donor language, (2) the source word in the donor language, (3) impact on the lexicon on the recipient language, i.e. replacement (of existing term in the recipient language), insertion (where no term existed previously in the recipient language), or coexistence (with existing term in the recipient language), (4) contact situation, i.e. in what context a word was borrowed.

3.1.2. Data analysis in the LWT project

When all the information was collected some additional data treatment was applied in order to make it possible to account for inconsistencies between and within languages in relation to the semantic concepts. The LWT project chose to incorporate the use of a *Borrowability score* between 0 and 1, which indicated the how likely a word was to be borrowed. This directly corresponds to the score given by the contributors for borrowed status. Furthermore, if a word was polysemous, and as a result related to two or more semantic domains or word classes, it was counted as one word split over those domains or word classes. That is, if a word is related to two semantic domains it was counted as $\frac{1}{2}$ for each. This was essentially to prevent the counting of polysemous words more than once as a polysemous word is still just one word, but with a broader semantic scope. Unfortunately, this leads to a rather abstract score system, but it was deemed necessary in order to create a weighted score for borrowability. However, as

noted in *Loanwords in the World's Languages* (Haspelmath & Tadmor, 2009), these polysemous cases were very few.

3.2. Language selection

There are three main factors motivating the choice of the four languages studied in this thesis. The most important one is accessibility to material. Simply put, there has to exist dictionaries with etymological information available, which is considerably more of a challenge than might be first imagined. The second motivation is linguistic diversity, and the third time limitations.

There is a multitude of interesting languages and dialects in the region, and at first a larger set of languages was considered, including, for example, Armenian, Azerbaijani, and Iraqi Arabic. The main issue with these languages is that the material needed to gather the etymological data is not easily available, especially if one does not have a working expertise of relevant languages. Since I do not speak Arabic, Armenian, Azerbaijani, or Russian this poses a major barrier in terms of material access. Hence Turkish is chosen rather than Azerbaijani for this study, although Azerbaijani is closer to the epicentre of the region and would be preferable. The languages of Western Farsi, Northern Kurdish, and Turkish have more or less easily accessible etymological dictionaries. Georgian turns out to be somewhat more difficult, something which is further discussed in *3.7.1 Issues with the Georgian data*. A contributing factor to choosing these is linguistic diversity, since they belong to three very distinct language families, Indo-European, Turkic, and Kartvelian. For this reason, Georgian, a Kartvelian language, is prioritised over Ossetic, an Indo-European language that could have been chosen. The third limitation is time, without which a larger set of languages can be included. Figure 3 shows where the chosen languages are spoken.



Figure 3. Map displaying where the languages in this study are spoken. Turkish in green, Western Farsi in purple, Northern Kurdish in red, and Georgian in yellow. Note that this is an estimate and both Turkish and Western Farsi are spoken also outside of Turkey and Iran.

3.2.1. Georgian

Georgian belongs to the Georgic branch of the Kartvelian language family and is mainly spoken in Georgia with around 4 million speakers, where it is also the official language. Georgian has been a literary language since the 6th century AD and uses its own script, the Georgian script. The Kartvelian homeland is suggested to lie in the Caucasus (Klimov, 1998), where Georgian is still spoken today (Campbell & King, 2013). There are also pockets of Georgian spoken in Iran.

3.2.2. Northern Kurdish

Northern Kurdish, also known as Kurmanji, is an Indo-European language within the Indo-Iranian branch, and relatively closely related to Western Farsi. Kurdish is spoken in a wide area covering parts of Turkey, Syria, Iraq, and Iran. Northern Kurdish is the largest dialect of Kurdish and is spoken by an estimate of 20 million people, although the exact number vary depending on source. There are also pockets of Kurdish speakers in Khorasan (Iran), Baluchistan, Armenia, Azerbaijan, and Turkmenistan. Furthermore, the language has no widely acknowledged standard form and several different scripts are used to write it (Campbell & King, 2013).

3.2.3. Western Farsi

Relative of Northern Kurdish, Western Farsi, also known as Persian or Farsi, is an Indo-European language belonging to the Indo-Iranian branch and is spoken by around 75 million people in Iran, where it also is the official language. Very closely related varieties, Dari and Tajik, are spoken in Afghanistan and Tajikistan respectively. Western Farsi has been a major literary language since the 10th century AD and is written using the Arabic script (Campbell & King, 2013).

3.2.4. Turkish

Turkish belongs to the Oghuz branch of the Turkic language family and is the official language of Turkey, and co-official language of Cyprus together with Greek. It is spoken by around 60 million people in Turkey, as well as by a relatively large number of speakers in Cyprus, Bulgaria, Greece and former Yugoslavia. Turkish has written records dating back to the 13th century, where it is already possible to see lexical influence from Arabic and Persian. As time went on and the literary language of Ottoman Turkish developed, where influence increased significantly (Campbell & King, 2013). Major language reforms, initiated by Kamal Atatürk, the founder of modern-day Turkey, abandoned the Arabic script in favour of the Roman and removed much of the Arabic and Persian influence from Turkish.

3.3. Word list selection

The original word list of the Loanword Typology project is 1,460 words long, and thus much too long to apply to all four languages. It is therefore necessary to create a subsection of the list of a more manageable size.

While the word list originates from the Intercontinental Dictionary Series (IDS) (Key & Comrie, 2023), IDS is not the only large word list used to study lexica. Another one is the NorthEuraLex (NEL) database (Dellert et al., 2019). By intersecting the LWT project word list and the word list of NorthEuraLex a new word list of 396 concepts was created. I then matched this new list, containing the concepts from NorthEuraLex, with the concepts from LWT, which produced a final list of 420 concepts. The reason the two lists turned out different was because some entries in English were both nouns and

verbs (e.g. comb n. - comb v., paint n. - paint v.), or verbs and adjectives (clean v. - clean adj.). These were manually marked, and duplicates removed, leaving a list of 401 concepts (see Appendix 1).

Other lists can be considered, such as intersecting with CLICS (Rzymiski et al., 2019) or Concepticon (List et al., 2023), but these lists are in the former case much too small, consisting of only 201 concepts, and the latter case, larger than that of LWT with 1507 concepts (likely as a result of some words being paired twice in the script run to find the subset). There are other conceivable ways to arrive at a word list of reasonable length, for example choosing certain semantic domains only, e.g. some semantic domains found highly, moderately, and less resistant to borrowings by the LWT project. The main issue with choosing this approach is that a given language contact area might not follow the same distribution of loanwords per semantic domain as the global norm. The resulting dataset might therefore not be entirely representative of the region's lexical borrowing status. Furthermore, a major advantage of using NorthEuraLex over these other methods is the fact that NorthEuraLex provides a set of semantic concepts (with English, German, and Russian as metalanguages), the word forms, in original script and transliterated, as well as several phonetic transcriptions. Furthermore, that NorthEuraLex has information of all of these words gives some indication that these words likely can be found in dictionaries, which reduces the risk of running into the problem of not finding words.

3.4. The working process

Similar to the LWT project I set out to find the corresponding word form for each semantic concept in the list created above. Each concept is looked up in an etymological dictionary for the language in question, some dictionaries are physical books, others are online dictionaries, and some are in English while others are in different languages. The information I am interested in is primarily whether the word is borrowed or not. Since I am not proficient in all languages in the study and there are limitations to the material that I have access to, I cannot afford to be as detailed in our borrowed status as the contributors of the LWT project. Instead of using a scale from 0 to 4, I choose to use a simple Y (for certainly borrowed, and probably borrowed) and N (for probably not borrowed, and very unlikely borrowed) system. The borrowability score assigned is 1 for Y and 0 for N. In cases of polysemous words the score is divided by the total number of word forms, for example Turkish *hava* is polysemous and means both 'air' and 'weather' (borrowed from Arabic) would get a borrowability score of 0.5 each to prevent a word being counted twice.

In the cases where no clear source is mentioned, a note is taken about the information given in the dictionary related to a potential source. This is mostly a concern with *Kurdish-English Dictionary Ferhenga Kurmancî-Inglîzî*, where in many cases the etymological information consists of references to other languages having the same or similar word forms. These are looked at individually and judged to be either likely borrowed or likely not borrowed based on which languages are referenced. In Figure 4 an excerpt from the dictionary is displayed, where we can see there is a reference to Old Iranian, from

whence the Persian word comes, and furthermore is similar to the word forms in Central Kurdish and Zaza. In cases like this, the word is deemed probably not borrowed, hence marked as N for borrowing.

hêstir I هێستەر *pl./f. (JB1-S) tear(s) (water falling from the eyes while crying)* . {also: êsir (L); hêsir (Z-1/Dyd/Çnr/Kp/Srk); hêsir II (Wn); histêrik (Klk); hestir (Grc/Btm); hêsir (Erg); stêrik (Haz); stêrk II (Msr)} {syn: ronî II[2] (Hk); sirişk (K-2)} Cf. O Ir *asru(ka)--> Middle P ars / srešk--> P ašk
 فەرمێسک/sirişk مرشک; Sor esrîn ئەسرین = firmêsk فەرمێسک
 Za hêrs *f.* (Todd)/hesri/hesir *f.* (Mal)

Figure 4. Entry hêstir in Chyet (2003, p. 251).

If no information about a word's etymology is given it is regarded as probably not borrowed. This will, however, very likely affect the end results. The reason this approach is adopted is because, since many of the dictionaries are not etymological but explanatory in nature with additional etymological information, there are too many words that have no etymological information. Since I am interested in loans, I prefer to focus on the word forms I know are certainly borrowed, which would mean explicitly mentioned as such in a dictionary. If there are further issues or uncertainties, they are marked with question marks. When a donor language is presented, it is noted down, as is any information on source word. If more than one donor language is given it is written as [donor language 1] / [donor language 2]. Another kind of potential source is analysability, since we by the methodology of the LWT project consider words that are analysable to be constructions of the language, and thus not borrowed, even if the morphemes themselves are. Each entry is also referenced to one of the dictionaries used.

If a word form given by the list cannot be found in the dictionary and there is an option to look for the concept directly, such as is the case in dictionaries where both target languages are used (cf. *Kurdish-English Dictionary Ferhenga Kurmancî-Inglîzî*) and include these word forms in rows underneath the word form presented by the list.

3.4.1. What counts as a loanword

In this study words count as loanwords if they in a dictionary are mentioned as such. There are, however, some conditions that need to be fulfilled apart from this. Firstly, the word has to be morphologically unanalysable, i.e. it should not be a construction of morphemes that are recognised as such in the language. This would for example exclude words that are compounds where one part of the word is borrowed, which is very common in verbal constructions (for example the use of *faire* in French or *kardan* کردن in Western Farsi), or derivations where the derivational morpheme is a part of the recipient language. I also only look at lexical units and I am not interested in grammatical morphemes. It should be mentioned, however, that some words in the semantic domain *Miscellaneous* are semantically weak

(e.g. *not*, LWT ID 24.06), and this is recognised by the LWT project but they deemed it interesting to include this category, nonetheless.

I am also not discriminatory against age in this case. If a word has been borrowed in any stage of a language's history it is considered a loanword, as long as the dictionary specifies that it is borrowed (otherwise any word could be considered potentially borrowed, since it then just becomes a question of how far back you go).

3.4.2. Synonyms, polysemy, and semantic granularity

Michael Chyet (2003) provides very detailed entries in his dictionary of Northern Kurdish, which also entails that most words have listed synonyms or semantically related words. In some cases these can amount to eleven semantically related words. When there exists one word that is corresponding to the one given in our list, I choose that one and ignore the others. If this is not the case, I have to look into each of the dictionary entries and try to determine which ones are closer to the intended semantic meaning. If it is not possible to determine one, I choose the smallest number possible. This sorting can also be done if there is access to an informant, who can determine which words are relevant and which ones are not, unfortunately, this is not possible in the current study. It should be noted here that I cannot try and keep as many of the synonyms as possible, since many of the dictionaries I work with are in languages not known to me and therefore cannot provide the same level of semantic detail. I have to choose to walk at the pace of the slowest dictionary, so to speak, in order to conserve some balance in the level of detail provided. It is, however, possible that with access to a speaker of the language, more synonyms can be specified to belong to different registers, such as *dog* (formal) and *doggie* (informal), or semantically indispensable for the concept.

Some cases might be trickier than others, such as the word for *arm* in Northern Kurdish, for which several different words are provided, ranging in meaning from fingers to shoulder, wrist to shoulder, elbow to shoulder. It would appear that Kurdish has a higher granularity in terms of arm semantics than English. However, there is also a word in Northern Kurdish, *chepil*, meaning simply arm, seemingly in the English sense. Since this can be considered closer to the English concept it is chosen over the others. While this does provide concern from a standpoint of semantics and over-reliance on English, it is an approach that I choose to employ simply because I do not have proficiency enough in the languages of interest to grasp the actual semantic properties of every word. With more time and resources this can be prevented to a certain extent, but there will always be some inherent problems with using a metalanguage.

The method described above for Chyet's dictionary of Northern Kurdish is applied in the same fashion to all languages.

3.5. Dealing with dictionaries

Dictionaries are the source of material in this study and there are some considerations that need to be incorporated in how to deal with them. Before we delve into the finer methodological points below it is important to keep in mind that dictionaries are large projects and can be used in nation building or be financed by large government bodies. This means that there might be a certain bias towards, for example, not include borrowed word forms when there exists an inherited one, although the borrowed word form might be far more common. It could also be that some words are so obviously borrowed that the author of the dictionary deems it unnecessary to include them in an etymological account. There are thus many dictionaries that are not purely based on linguistic considerations, but contain, to some degree, a certain political element as well.

In some cases it is not possible for me to deal with the potential ramifications of this for this thesis simply because of lack of alternative sources and time. This does not, however, mean that the data is unusable, it means, rather, that I have to take this into account during analysis, and that it can reveal something about the attitude of the institutions of the language in question.

3.5.1. The dictionaries of this study

The dictionaries used for this thesis are presented in Table 3.

Table 3. Dictionaries used in the thesis, with references.

	<i>Dictionary</i>	<i>Author</i>
<i>Georgian</i>	<i>Etymological Dictionary of the Kartvelian Languages</i>	Klimov (1998)
	<i>Kartwelisches Etymologisches Wörterbuch (the Kartvelian etymological dictionary)</i>	Fähnrich (2007)
	<i>Kartuli enis ganmart'ebiti leksik'oni (the Georgian Explanatory Dictionary)</i>	Čikobava (n.d.)
<i>Northern Kurdish</i>	<i>Kurdish-English Dictionary Ferhenga Kurmancî-Inglîzî</i>	Chyet (2003)
	<i>Etymological Dictionary of Kurdish (vol. 1, and 2)</i>	Tsabolov (2001)
<i>Turkish</i>	<i>Tarihi ve etimolojik türkiye türkçesi lugatı (vol 1, 2, 5, and 7)</i>	Tietze (2002); Tietze et al. (2021a, 2021b); Tietze et al. (2019)
	<i>Nişanyan Sözlük</i>	Nişanyan (2023)
<i>Western Farsi</i>	<i>An Etymological Dictionary of Persian, English and other Indo-European Languages</i>	Nourai (2013)
	<i>Dehkhoda Dictionary</i>	Dehkhoda (n.d.)

3.5.2. Dictionaries in unknown languages

An interesting part of this thesis is to work with dictionaries written in languages I have no working knowledge of. This includes Nişanyan Sözlük, the Georgian Explanatory Dictionary, *Tarihi ve etimolojik türkiye türkçesi lugatı (vol 1, 2, 5, and 7)*, *Etymological Dictionary of Kurdish (vol. 1, and*

2), and the *Kartvelian etymological dictionary*, see Table 3 for references. Fortunately, there are ways to work around this issue. When using Nişanyar Sözlük the interface gives a very clear indication of the word's source. In Figure 5 we can see the source being marked in bold.

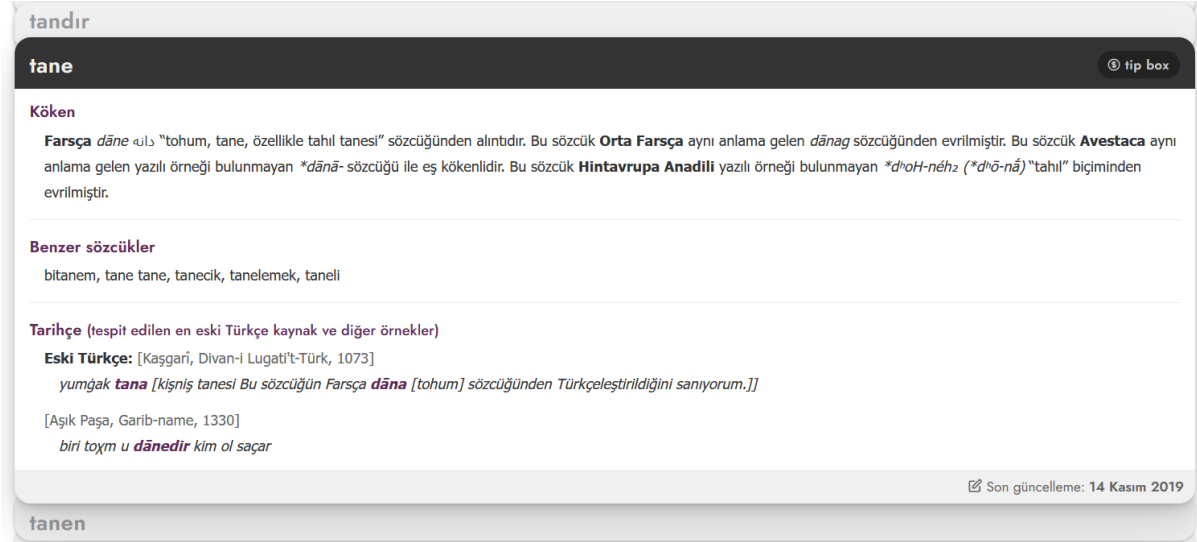


Figure 5. Entry *tane* (grain, piece) in *Nişanyar Sözlük dictionary*.

While this is still in Turkish it is relatively easy to decipher, and since I am not primarily interested in the semantic information, I do not need to actually understand what the entry says. If there are several entries for the same word form, I first see if all entries have the same etymology. If this is the case, I need not look at the entry that specifically corresponds to our concept, but I can enter the etymological information immediately, see Figure 6. If this is not the case, however, and the etymologies differ, I have to consult another dictionary in order to understand some of the entries' information, such as Google Translate (Google, 2023) and Lexin (Berg et al., 2023). The same strategy is used for the Georgian Explanatory Dictionary. It should be kept in mind that the Georgian Explanatory Dictionary is not primarily an etymological dictionary but an explanatory dictionary. The main difference there is that the etymology is shown in brackets at the top, see Figure 7. When using these two dictionaries I search using the script of the languages found in my list, courtesy of NorthEuraLex.

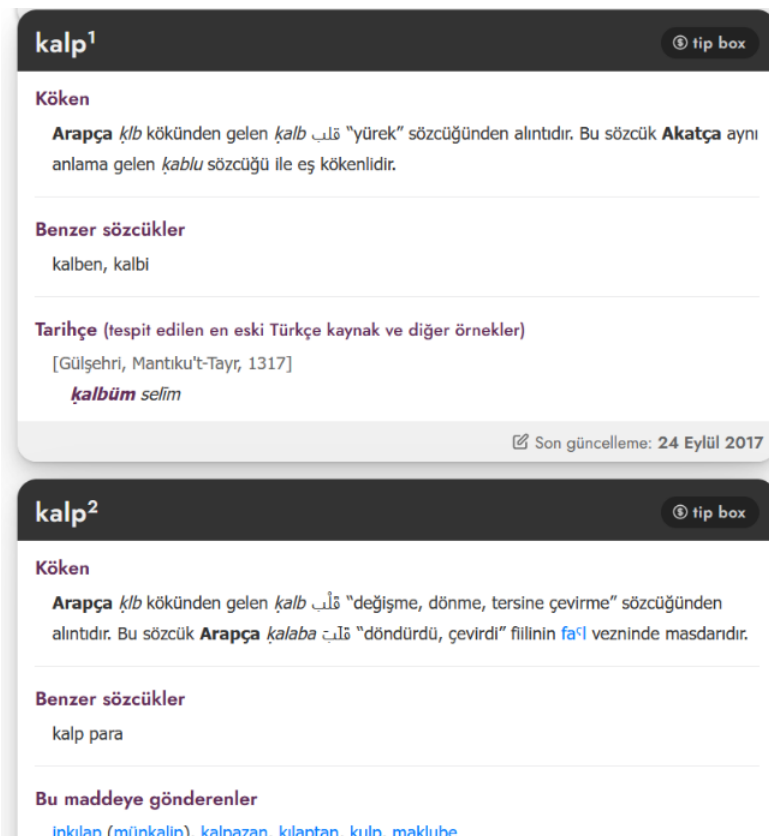


Figure 6. Entries for *kalp* (heart) in *Nişanyar Sözlük* dictionary.

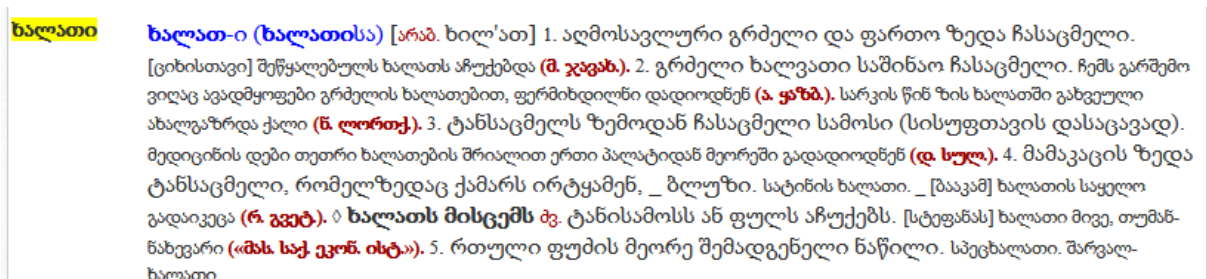


Figure 7. Entry *ხალათი* (shirt) in the *Georgian Explanatory Dictionary*.

Since *Tarihi ve etimolojik türkiye türkçesi lugatı* are physical books the method changes, but broadly stays the same. What changes is the way I find the etymological information, and since there is still a lot of structure to these dictionaries the etymological information is still accessible. In *Tarihi ve etimolojik türkiye türkçesi lugatı*, loanwords are marked with < followed by the recipient language. The *Etymological Dictionary of Kurdish* is difficult to work with without a working knowledge of Russian and as a result when using this dictionary it is necessary to resort to the use of secondary dictionary sources, such as Google Translate and Lexin. There are other issues with this dictionary relating to the choices of transcription for the Kurdish words, and since they differ from the ones used by NorthEuraLex it is a time-consuming enterprise to use this dictionary extensively. Somewhat similar, yet different to the two dictionaries mentioned above, is the *Kartvelian etymological dictionary*. The

same strategies are employed but with the added benefit of the my cursory knowledge of German, making it possible to rely slightly less on secondary dictionaries.

These kinds of sources, for example Google Translate and Lexin, are not lent to translating prose, or any kind of more advanced, lengthy text. This is not the kind of text that I aim to translate, but rather single words, or simple phrases. All in all, neither translations of complicated text nor lexical meaning for purposes of semantic analyses are necessary, and hence these dictionaries are judged to be a relatively reliable source for the purposes of this thesis.

3.5.3. Dictionaries in known languages

These dictionaries include the ones written in English, i.e. *An Etymological Dictionary of Persian, English and other Indo-European Languages*, *Kurdish-English Dictionary Ferhenga Kurmancî-Inglîzî*, and *Etymological Dictionary of the Kartvelian Languages*, as well as the Dehkhoda Dictionary, written in Persian. It should be noted that the Dehkhoda Dictionary is primarily explanatory, but often incorporate the etymology of words.

Some consideration has to be given to the dictionary of Nourai (2013) and Chyet (2003). Nourai very often mentions that words in Western Farsi come from Avestan. Avestan is a language closely related to Old Persian (the ancestor language of Western Farsi) and is much better documented than Old Persian. This would superficially indicate that a large number of words in Persian are old loans from Avestan. However, it is far more likely that these words are inherited from Old Persian, but Avestan is given as the source in its place. In Figure 8 we can see an example of a word probably being inherited rather than being borrowed from Avestan. Similarly Chyet often mentions no clear source of a word as inherited or borrowed but provides a list of similar or related words in other languages for comparative purposes. From this one can arrive at some understanding of the probability of a word being borrowed or not. If for example it is the same or similar words in related languages it is likely an inherited word and not a borrowed one, especially if some of the related languages are spoken in a different geographical area altogether or if an older stage of the language has the same word.

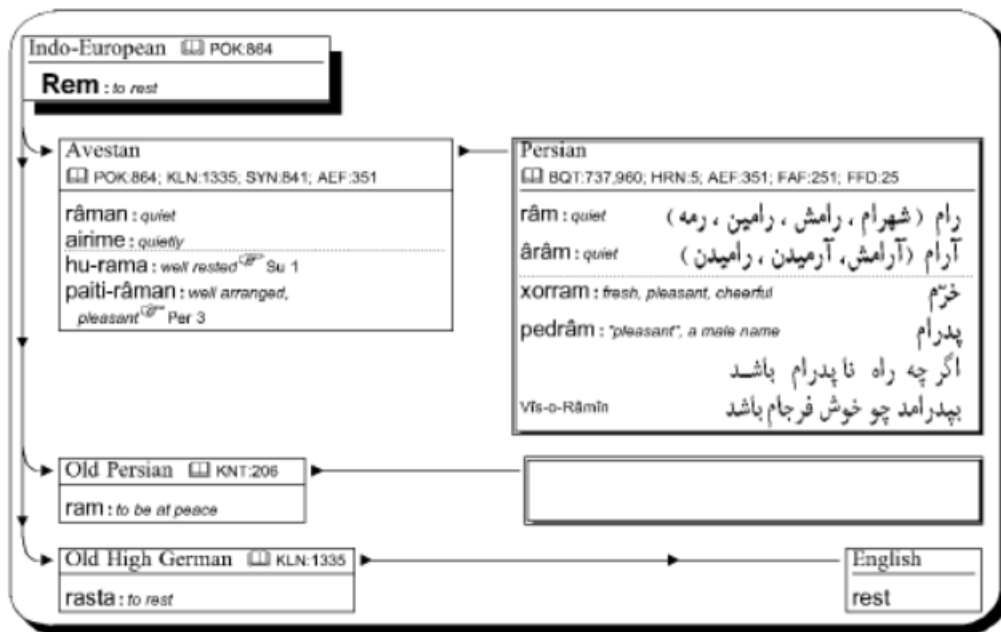


Figure 8. Etymology of Persian word *ârâm* آرام 'quiet, calm' (Nourai, 2013, p. 390).

3.5.4. Common problems

Some problems are reoccurring when working with dictionaries and here I discuss some of them and strategies for dealing with the issues they present. First, it is not uncommon that different dictionaries and databases use different transcription systems for scripts not based on the roman alphabet. Second, many dictionaries have very limited etymological data on loanwords, their focus lying with inherited words instead. Third, it can happen that dictionaries give contradicting information. Fourth, it is in some cases unclear whether a dictionary provides the source word of the donor language or the source word in the original language of the word.

The first issue can be a major problem but, in many cases, not an insurmountable one. The list used in this study includes both roman-based transcriptions and different kinds of phonetic transcriptions. One way to solve the problem is to look at the phonetic transcriptions and see how they relate to the transcription scheme used in the dictionary. Other times the transcription systems are similar enough to not pose a problem at all.

The second issue is problematic, mostly in combination with the sparse accessibility of dictionaries for the languages in this study. These dictionaries are not without use, since they narrow down the number of words you need to consider as possibly borrowed, but the only real way to solve this issue is to find another dictionary with etymological information of loanwords.

The third issue is approached with three principles in this thesis. The first principle is that a borrowed status outweighs no information, in other words, if one dictionary states that a word is borrowed and another gives no etymological information at all, it is considered borrowed. The second principle is that if more than one dictionary says that a word is borrowed, but they give different sources they are marked as [donor language 1] / [donor language 2]. The third principle concerns cases where

two dictionaries give diametrically contradicting information (one dictionary states a word to be borrowed and another to be inherited), in these cases I mark borrowed status as a question mark.

As for the fourth, this is again very difficult to know and cannot be further studied in this thesis. I have to accept that there might be some inconsistencies in the final analysis and be aware of this when analysing.

One last note should be made on Northern Kurdish. All the other languages are national languages, standardised and used within as a national language in the framework of a state, this is not quite so for Northern Kurdish. It is thus worth to keep in mind that this entails that Northern Kurdish might be written in a variety of different ways, using different scripts or spellings.

3.6. The structure of the dataset

The structure of the dataset is displayed in Table 4. Some columns are excluded, as they act only to show metainformation, or are simply superfluous. To view the dataset, the reader is directed to Appendix 2.

Table 4. Columns in the dataset and their explanations.

<i>Column name</i>	<i>Explanation</i>
LANGUAGE	Name of recipient language
GLOTTOCODE	Glottocode from Glottolog
WORD_FORM	Word form either from dictionary or from NorthEuraLex
RAWIPA	IPA transcription, available for all word forms from NorthEuraLex
LEXICAL_CONCEPT	Concept in English from the LWT project
LWT_ID	Concept ID from the LWT project
LWT_SEMANTIC_DOMAIN	Semantic domain as specified by the LWT project
BORROWED	Borrowed status of word form, either Y, N, or ?
BORROWABILITY_SCORE	Numerical score given to borrowed status, Y = 1, N = 0.
SOURCE_WORD	Source word in the donor language
SOURCE_WORD ORTHOGRAPHY	Source word in the script of the donor language (if available)
SOURCE MEANING	Meaning of source word (if available)
DONOR_LANGUAGE	Donor language (abbreviations)
DICTIONARY (BORROWING)	Dictionary from which borrowing status was retrieved
SECONDARY SOURCE	If another dictionary was used to confirm the <i>Dictionary (borrowing)</i>
NOTES	For other observations or important information

3.7. Potential sources of error

One problem is that the composition of the individual language's word lists might not be similar at all, the lists are based on LWT list cross-sectioned with NorthEuraLex, but I do not exactly know how the NEL lists look in the first place. This might mean that there are more of some words in some of the lists on account of the NEL lists not being uniform.

3.7.1. Issues with the Georgian data

During the data collection I noticed the surprising lack of loans in Georgian as compared to the other languages. Georgian has considerably more words in its dataset than the others, reaching 510 word forms (the next largest set is Western Farsi at 459 word forms), but at the same time also a very low amount of loanwords. There are 20 loans in Georgian, amounting to around 4% of the total, which is much lower than any of the other languages. The data has been collected and is available in the complete dataset, but it is excluded from any further analysis as its borrowing ratio is suspiciously low and might introduce unnecessary flaws in the results.

The main reason for not entirely trusting the Georgian data is that I have a very limited insight into the sources for Georgian. The dictionary written in English, *Etymological Dictionary of the Kartvelian Languages* (Klimov, 1998), is specifically for inherited vocabulary and reconstructions of roots, which means that it has essentially no information on loanwords. The *Kartwelisches Etymologisches Wörterbuch* (Fährnich, 2007) is also a book with a focus on the inherited lexicon and furthermore is in German, a language I have limited knowledge of. The Georgian Explanatory Dictionary (Čikobava, n.d.) is, thus, the only dictionary that has any information concerning loanwords and it is entirely in Georgian, which I have no competence in at all. This, together with my decision to count words that have no explicit etymological history as probably not borrowed, means that I have no real idea of the quality of the data. In other words, there might very well be many words from the Georgian Explanatory Dictionary that are in fact borrowed but have not been considered as such because there has not been explicitly shown in the beginning of the entry. When it comes to the other languages, the issue does not quite exist in the same way. With access to more dictionaries, or with better language skills, one could certainly overcome this barrier, but such possibilities have not been presented during the course of this thesis.

The dictionaries for Western Farsi are in languages I have extensive knowledge of and one dictionary is explicitly etymological, containing information on borrowings (Nourai, 2013), and the other frequently mentions origin of loans (Dehkhoda, n.d.). The dictionary mainly used for Kurdish has extensive etymological information or provides data as to possible origins, and is furthermore in English (Chyet, 2003). Both of the Turkish dictionaries are in Turkish, but one is about the etymology and history of words, not exclusively inherited lexicon (Tietze, 2002; Tietze et al., 2021a, 2021b; Tietze et al., 2019),

and the other always mentions source, for inherited and borrowed terminology alike (Nişanyan, 2023). Therefore, I can feel more confident in these sources and thus in the data for these languages.

All in all, I believe that the remarkably low borrowing count of Georgian, coupled with (1) the methodological decision of not considering words without etymology as potentially borrowed, (2) the information from the dictionaries, and (3) my lack of language skills, provide ground enough to be suspicious of the quality of the Georgian data. This leads to the final decision of excluding it from the results and analysis, although it is still collected and available in the complete dataset Appendix 2.

4. Results and analysis

In this section I present the results of the study. I will start by presenting the overall results in broad strokes and compare them to the results achieved by the Loanword Typology project, with the goal in mind to answer the question of how the languages, Northern Kurdish, Turkish, and Western Farsi compared to the languages in the project. Then I will go on to present some analysis in an attempt to categorise the data into bite-sized chunks, rather than the pebbles brought about by the limited scope of the thesis, and answer the other research questions; What are the main donor languages in the region? And what signs of linguistic and/or cultural contact can we see through the composition of loanwords?

4.1. Overview of borrowings

First off, we can observe that Turkish is by far the largest recipient of loanwords among the three languages, with around 26.5% of their vocabulary from the list being borrowed. Following Turkish comes Persian, with around 21% borrowed vocabulary. Then Northern Kurdish follows with 16%, see Table 5.

Table 5. The percentage and number of borrowings and total word forms in the dataset per language and the region as a whole.

	Percentage of borrowed terms	Number of borrowed terms	Total number of terms
<i>Turkish</i>	26,48%	121	457
<i>Western Farsi</i>	21,09%	97	460
<i>Northern Kurdish</i>	15,88%	57	359
<i>Region</i>	21,49%	274	1275

Furthermore, as we can see in Table 5 that the region as a whole has an approximate borrowing score of 21.5%, and there is a total of 1275 word forms collected, after all sorting and filtering has been done.

4.1.1. Borrowings in the semantic domains

All words are categorised as part of a semantic domain in order to make it possible to better understand the nature of borrowings in relation to semantics. We need to take into consideration that all kinds of semantic domains are theoretical in their essence, that is, the way that words are categorised affect the semantic analysis. It should furthermore be noted that the semantic domain *Law*, originally part of the LWT project ended up with no word forms in our list and is thus not part at all in the results. The complete word list showcasing which concepts relate to which semantic domains is available in Appendix 1.

As we can see from Table 6, describing the distribution of loanwords per semantic domain, counting all three languages together, there are some immediate issues when it comes to the word count for some of the semantic domains. Due to the limited number of words per language, as a result from

investigating not only one but three languages, there are some semantic domains that are sorely lacking in numbers, especially this should be considered for the domain of *Religion and belief* which only has 6 word forms represented in the data. Religious terms tend to be borrowed frequently and so the percentage of 50% borrowed is not entirely unreasonable, but the low word count is not substantial enough to truly rely on. Apart from *Religion and belief* the top three domains in terms of borrowability are *Possession*, *Cognition*, *Speech and language*, with 48%, 44%, and 42% of the lexicon being borrowed respectively. At the lower end we can see that *Miscellaneous function words*, *Animals*, and *Quantity* are the semantic domains competing for the lowest borrowability in this region, with 11%, 8% and 8% of the lexicon being borrowed respectively.

Table 6. Number of borrowings and total count per semantic domain in the region.

<i>The region</i>	<i>Percentage</i>	<i>Number of borrowed words</i>	<i>Total word count</i>
<i>Religion and belief</i>	50%	3	6
<i>Possession</i>	48%	12	25
<i>Cognition</i>	44%	8	18
<i>Speech and language</i>	42%	10	24
<i>Social and political relations</i>	41%	14	34
<i>Modern world</i>	38%	6	16
<i>The house</i>	34%	13	38
<i>Emotions and values</i>	33%	14	42
<i>Motion</i>	30%	6	20
<i>Spatial relations</i>	27%	32	119
<i>Clothing and grooming</i>	24%	11	45
<i>Time</i>	24%	24	99
<i>Agriculture and vegetation</i>	24%	11,5	48,5
<i>The physical world</i>	23%	26	114
<i>The body</i>	16%	27	169
<i>Basic actions and technology</i>	17%	7	42
<i>Food and drink</i>	15%	9	59
<i>Sense perception</i>	15%	12	79
<i>Kinship</i>	14%	11	76
<i>Warfare and hunting</i>	13%	2	15
<i>Miscellaneous function words</i>	11%	2	18
<i>Animals</i>	8%	7	86
<i>Quantity</i>	8%	6,5	83,5

Moving along to Northern Kurdish, see Table 7. We can see that the top semantic domains in terms of borrowability are *Social and political relations* (44%), *The house* (43%), *Modern world* (40%), excluding *Religion and belief*. Interestingly, when considering the total word count, we can observe that the semantic domains at the bottom tend to have higher counts than those closer to the higher

borrowability end. At the bottom end for Northern Kurdish we can see that, words relating to *Animals* *Food and drink*, as well as *Quantity* are rarely borrowed.

Table 7. Number of borrowings per semantic domain in Northern Kurdish.

<i>Northern Kurdish</i>	<i>Percentage</i>	<i>Number of borrowed words</i>	<i>Total word count</i>
<i>Religion and belief</i>	50%	1	2
<i>Social and political relations</i>	44%	4	9
<i>The house</i>	43%	3	7
<i>Modern world</i>	40%	2	5
<i>Miscellaneous function words</i>	33%	1	3
<i>Agriculture and vegetation</i>	30%	3	10
<i>Basic actions and technology</i>	30%	3	10
<i>Possession</i>	29%	2	7
<i>The physical world</i>	26%	10	39
<i>Cognition</i>	25%	1	4
<i>Speech and language</i>	22%	2	9
<i>Motion</i>	20%	1	5
<i>Warfare and hunting</i>	20%	1	5
<i>Clothing and grooming</i>	18%	2	11
<i>Emotions and values</i>	18%	2	11
<i>Time</i>	17%	5	29
<i>Kinship</i>	15%	3	20
<i>Spatial relations</i>	13%	5	39
<i>Sense perception</i>	8%	2	25
<i>The body</i>	7%	3	45
<i>Animals</i>	4%	1	24
<i>Food and drink</i>	0%	0	14
<i>Quantity</i>	0%	0	26

Table 8 shows the distribution of loans in Western Farsi. The semantic domains most frequently borrowed into are *Speech and language* (67%), *Spatial relations* (41%), *Possession* (40%) and the ones with the lowest score of borrowings are *Social and political relations* (8%), *Animals* (7%), *Miscellaneous function words* (0%).

Table 8. Number of borrowings per semantic domain in Western Farsi.

Western Farsi	Percentage	Number of borrowed words	Total word count
<i>Speech and language</i>	67%	4	6
<i>Religion and belief</i>	50%	1	2
<i>Spatial relations</i>	41%	17	41
<i>Possession</i>	40%	4	10
<i>Emotions and values</i>	38%	6	16
<i>Cognition</i>	33%	2	6
<i>Motion</i>	29%	2	7
<i>Food and drink</i>	27%	6	22
<i>The physical world</i>	24%	10	41
<i>Basic actions and technology</i>	24%	4	17
<i>Warfare and hunting</i>	20%	1	5
<i>Time</i>	19%	7	36
<i>Kinship</i>	19%	5	26
<i>Modern world</i>	17%	1	6
<i>Agriculture and vegetation</i>	15%	3	20
<i>Quantity</i>	15%	4	27
<i>The house</i>	14%	2	14
<i>The body</i>	14%	9	66
<i>Clothing and grooming</i>	13%	2	15
<i>Sense perception</i>	11%	3	27
<i>Social and political relations</i>	8%	1	12
<i>Animals</i>	7%	2	30
<i>Miscellaneous function words</i>	0%	0	8

As for the last language, Turkish, we can observe the following three domains at the top, *Possession*, *Social and political relations*, and *Cognition* with 75%, 69%, and 63% borrowed terminology respectively, see Table 9. At the other end of the spectrum we have *Quantity* with 8% of the words borrowed, *Basic actions and technology* tied with *Warfare and hunting* at 0%. I should also point out that we can see the effect of dealing with polysemous words here, as some entries have decimals. In this case it is the word form *dane* which means both ‘grain’ (*Agriculture and vegetation*) and ‘piece’ (*Quantity*), borrowed from Western Farsi’s *dāne* دانه with the same meaning. This does make the numbers somewhat abstract, but it also becomes a way to protect against the overreliance on English semantic granularity in its role as the metalanguage, by not counting the same polysemous word twice.

Table 9. Number of borrowings per semantic domain in Turkish.

Turkish	Percentage	Number of borrowed words	Total word count
Possession	75%	6	8
Social and political relations	69%	9	13
Cognition	63%	5	8
Modern world	60%	3	5
Religion and belief	50%	1	2
The house	47%	8	17
Speech and language	44%	4	9
Emotions and values	40%	6	15
Motion	38%	3	8
Clothing and grooming	37%	7	19
Time	35%	12	34
Agriculture and vegetation	30%	5,5	18,5
Sense perception	26%	7	27
The body	26%	15	58
Spatial relations	26%	10	39
The physical world	18%	6	34
Miscellaneous function words	14%	1	7
Food and drink	13%	3	23
Animals	13%	4	32
Kinship	10%	3	30
Quantity	8%	2,5	30,5
Basic actions and technology	0%	0	15
Warfare and hunting	0%	0	5

Interestingly enough, among the languages there is a considerable amount of variation in what is being borrowed to what extent. One fascinating example is that concepts relating to *Social and political relations* are very high in loanwords in Northern Kurdish (44%) and Turkish (69%), while, conversely, Western Farsi only reach 8% of borrowings. Similar to this situation is the semantic domain of *The house* which is high in borrowings in Northern Kurdish (43%) and Turkish (47%) and much lower in Western Farsi (14%). Western Farsi is also much lower in concepts borrowed into *Modern world* than the other languages. Western Farsi's 17% compares Northern Kurdish's 40%, and Turkish' 60%. There are some similarities, however, and we can see that for all languages concepts relating to *Animals*, *Quantity* tend to be very low in loanwords.

One further observation we can make is that borrowability is not limited to certain semantic domains. There seems, on a global scale, to be some tendencies as to what semantic domains are usually more crowded with loanwords, but from what we can see here is that in addition to this, if a language has a low amount of borrowings in general, then this applies to all semantic domains, and, vice versa, if a language is rich in loanwords this will reflect broadly on all semantic domains and not just one a few

in particular. This is illustrated in Figure 9. Regardless of which semantic domains that are more or less borrowed, when the highest to lowest borrowed semantic domains are paired between the languages, we can see that Turkish very high, while Northern Kurdish and Western Farsi turning over each other.

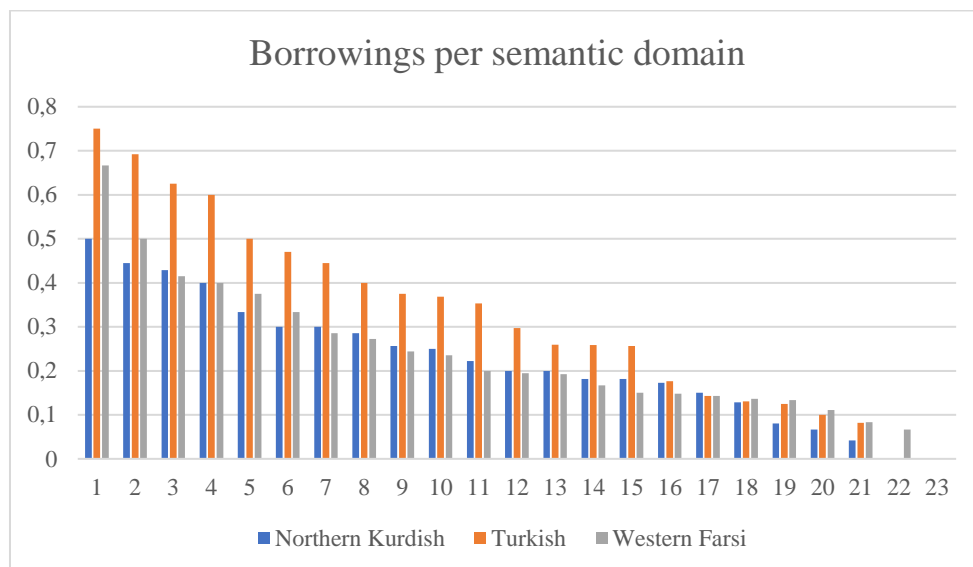


Figure 9. Bar plot comparing the highest to lowest loanword ratio between all languages. N.B. that the semantic domains are ordered by percentage of borrowings and not in alphabetical order, the bars do not represent the same semantic domains but only the relation between them in terms of borrowability.

It is not a perfect match with the overall borrowings per language, but this could indicate that low borrowability is not only tied to specific semantic domains, but that it is something that also covers the entire language, i.e. it is not the case that a language borrows a lot from certain semantic domains and nothing from others to average in the middle, but rather than a language with a low amount of borrowings will have few loanwords across the board.

4.1.2. Borrowability of word classes

As I discussed in the Background chapter there have been many borrowing hierarchies proposed in regard to word classes. Most tend to agree on that nouns are borrowed more frequently than other word classes, and it is followed by verbs, adjectives, and adverbs in different orders depending on specific hierarchies. This seems to be followed broadly by the findings of this study as well, with some caveats. Firstly, we can see from Table 10 that conjunctions are by far the most borrowed terms in the dataset with a borrowability score of 30% , followed by nouns at approximately 19%. This is not all that strange, considering that Matras (2007) put nouns and conjunctions together at the top in his hierarchy. Following this, we have adjectives and adverbs, with adjectives slightly more commonly borrowed at 13.4% than adverbs at 8.6%. Then come prepositions at circa 7% and pronouns at 4%. All of this matches well with the hierarchy of Matras (2007) except that there are no verbs borrowed at all in the dataset. One reason why no verbs are borrowed could be because of the restriction put on analysability. Many verbs in

Northern Kurdish and Western Farsi, for example, are constructed using a light verb (*kardan* in Western Farsi, and *kirin* in Northern Kurdish) in conjunction with a noun and are thus considered internal constructions.

Table 10. Borrowings per word class.

	<i>CNJ</i>	<i>N</i>	<i>A</i>	<i>PRP</i>	<i>ADV</i>	<i>PRN</i>	<i>NUM</i>	<i>V</i>
<i>Ratio</i>	37.5%	25.2%	18.3%	12.5%	11.9%	5.3%	0.0%	0.0%
<i>Total count</i>	8	872	218	8	67	18	47	37

4.2. Donor languages

There are a total of 23 different donor languages, counting language stages as different languages (i.e. Old Persian, Pahlavi, and Western Farsi as separate languages), and 1 unknown origin. As we can see from Table 11, which shows the combined data of all languages, there are four main donor languages in the area, namely Arabic, Western Farsi, Turkish, and Greek. Arabic is by far the largest donor language, representing around 55% of all borrowings in the dataset. Western Farsi comes in at second place and has given rise to approximately 19% of all the loanwords. Turkish has given rise to slightly more than 7% of all loanwords and Greek to right around 5%.

Table 11. Donor language overall contribution of borrowings.

	<i>Source to # of borrowings</i>	<i>Ratio of total borrowings</i>
<i>Arabic</i>	161	55,33%
<i>Western Farsi</i>	56,5	19,42%
<i>Turkish</i>	21	7,22%
<i>Greek</i>	13,5	4,64%
<i>Old Persian</i>	7	2,41%
<i>Sanskrit</i>	5	1,72%
<i>French</i>	4	1,37%
<i>Italian</i>	4	1,37%
<i>Syriac</i>	3	1,03%
<i>Armenian</i>	2	0,69%
<i>Chinese</i>	2	0,69%
<i>Chagatai</i>	2	0,69%
<i>Hebrew</i>	2	0,69%
<i>Bulgarian</i>	1	0,34%
<i>Cuman</i>	1	0,34%
<i>Hungarian</i>	1	0,34%
<i>Latin</i>	1	0,34%
<i>Mongolian</i>	1	0,34%
<i>Pahlavi</i>	1	0,34%
<i>Proto-Iranian</i>	1	0,34%
<i>Russian</i>	1	0,34%
<i>Serbian</i>	1	0,34%
<i>Sogdian</i>	1	0,34%
<i>Unknown</i>	1	0,34%

A lot of the donor languages are donors only to one language. Looking at each language in turn, starting with Turkish, as the language with the largest number of borrowings in the dataset, we can see that the top three donor languages are the exact same as for the region as a whole. As Arabic and Western Farsi both stand for 40.5% each of the loans in Turkish, there is a large drop off in borrowing rate to the third language Greek, representing just shy of 6% of the loanwords, see Table 12. This is not all that surprising, since both Arabic and Western Farsi had a massive influence on the Turkish in use during the Ottoman Empire.

Table 12. Donor language contribution to Turkish.

<i>Turkish</i>	<i>Source to # of borrowings</i>	<i>Ratio of total borrowings</i>
<i>Arabic</i>	50,5	41,00%
<i>Western Farsi</i>	50	41,00%
<i>Greek</i>	7,5	5,86%
<i>Italian</i>	3	2,51%
<i>Chagatai</i>	2	1,67%
<i>French</i>	2	1,67%
<i>Armenian</i>	1	0,84%
<i>Bulgarian</i>	1	0,84%
<i>Cuman</i>	1	0,84%
<i>Hungarian</i>	1	0,84%
<i>Mongolian</i>	1	0,84%
<i>Serbian</i>	1	0,84%

The next language in order of highest number of borrowings is Western Farsi. In Western Farsi, since it cannot borrow from itself, the number one donor language by magnitudes is Arabic at around 81% with the second language Turkish only reaching approximately 6%, see Table 13.

Table 13. Donor language contribution to Western Farsi.

<i>Western Farsi</i>	<i>Source to # of borrowings</i>	<i>Ratio of total borrowings</i>
<i>Arabic</i>	78	81,25%
<i>Turkish</i>	6	6,25%
<i>Sanskrit</i>	5	5,21%
<i>French</i>	2	2,08%
<i>Greek</i>	2	2,08%
<i>Chinese</i>	1	1,04%
<i>Hebrew</i>	1	1,04%
<i>Sogdian</i>	1	1,04%

Northern Kurdish, as can be suspected, has a lot of borrowings from Arabic, almost half of the borrowed words in Northern Kurdish comes from Arabic, see Table 14. After Arabic comes Turkish, which stands for around 23% of the loanwords in Northern Kurdish, and then Old Persian with a total of 7 loans, roughly 12%.

Table 14. Donor language contribution to Northern Kurdish.

<i>Northern Kurdish</i>	<i>Source to # of borrowings</i>	<i>Ratio of total borrowings</i>
<i>Arabic</i>	28	49,12%
<i>Turkish</i>	13	22,81%
<i>Old Persian</i>	7	12,28%
<i>Syriac</i>	3	5,26%
<i>Western Farsi</i>	2	3,51%
<i>Armenian</i>	1	1,75%
<i>Proto-Iranian</i>	1	1,75%
<i>Russian</i>	1	1,75%
<i>Unknown</i>	1	1,75%

4.3. Overall comparison with the LWT project

The Loanword Typology project is one of the largest studies on loanwords and borrowability, but none of the languages studied in this thesis are represented in the project, and so part of this thesis is to compare the languages Northern Kurdish, Turkish, and Western Farsi with the Loanword Typology project. Table 15 shows the languages of this study compared to the results per language of the LWT project. From this we can easily see that, as deemed by the LWT project, Turkish is considered a high borrower language, while Western Farsi and Northern Kurdish are average borrower languages.

Comparing the borrowability of the semantic domains we can see in Figure 10 that in the regional average per semantic domain in our study is often considerably lower than that of the global average as presented by the LWT project. In a couple of instances are the averages of languages in this study higher than that of the LWT average, i.e. *The physical world*, *Speech and language*, *Spatial relations*, *Religion and belief*, *Possession*, *Motion*, *Emotions and values*, and *Cognition*. There are a couple of potential explanations why the average is lower than that of the global. One, of course, could be that in this region words, with a few exceptions, are borrowed far less than that of other languages around the world, however, this is rather unlikely. The region is a historically well-integrated one in cultural, economic, religious, and political terms, and therefore we can expect there to be considerable exchange also of lexical units. Another is that I consider words that are not noted explicitly as borrowed as not borrowed. This essentially increases the risk of undetected loanwords, which in turn would lower the overall borrowability score. An additional reason could be the NorthEuraLex database, which is used for lexicostatistical research and thus might be biased towards basic vocabulary, i.e. to words of low borrowability. Hence, it might be difficult to draw any strong conclusions based on the comparison of the dataset of this thesis with that of the LWT project, but it is still valid to make inferences based on the relation between the three languages, since they are all based on the same list.

Table 15. The languages of this study entered into the overall language results of the LWT project (Haspelmath & Tadmor, 2009, p. 64). Highlights indicate the languages of this thesis.

<i>Borrowing type</i>	<i>Languages</i>	<i>Total words</i>	<i>Loanwords</i>	<i>Loanwords as % of total</i>	
<i>Very high borrowers</i>	Selice Romani	1431	898	62.80%	
	Tarifiyt Berber	1526	789	51.70%	
<i>High borrowers</i>	Gurindji	842	384	45.60%	
	Romanian	2137	894	41.80%	
	English	1504	617	41.00%	
	Saramaccan	1089	417	38.30%	
	Ceq Wong	862	319	37.00%	
	Japanese	1975	689	34.90%	
	Indonesian	1942	660	34.00%	
	Bezhta	1344	427	31.80%	
	Kildin Saami	1336	408	30.50%	
	Imbabura Quechua	1158	350	30.20%	
	Archi	1112	328	29.50%	
	Sakha	1411	409	29.00%	
	Vietnamese	1477	415	28.10%	
	Swahili	1610	447	27.80%	
	Turkish	460	121	26.50%	
	Yaqui	1379	366	26.50%	
	Thai	2063	539	26.10%	
	Takia	1123	291	25.90%	
	<i>Average borrowers</i>	Lower Sorbian	1671	374	22.40%
		Hausa	1452	323	22.20%
Mapudungun		1236	274	22.20%	
White Hmong		1290	273	21.20%	
Western Farsi		459	89	21.00%	
Kanuri		1427	283	19.80%	
Dutch		1513	289	19.10%	
Malagasy		1526	267	17.50%	
Zinacantán Tzotzil		1217	195	16.00%	
Northern Kurdish		361	57	15.90%	
Wichí		1187	188	15.80%	
Q'eqchi'		1774	266	15.00%	
Iraqw		1117	162	14.50%	
Kali'na		1110	156	14.10%	
Hawaiian		1245	169	13.60%	
Oroqen		1138	137	12.00%	
Hup		993	114	11.50%	
Gawwada		982	111	11.30%	
Seychelles Creole		1879	201	10.70%	
Otomi		2158	231	10.70%	
<i>Low borrowers</i>	Ket	1030	100	9.70%	
	Manange	1009	84	8.30%	
	Old High German	1203	70	5.80%	
	Mandarin Chinese	2042	25	1.20%	

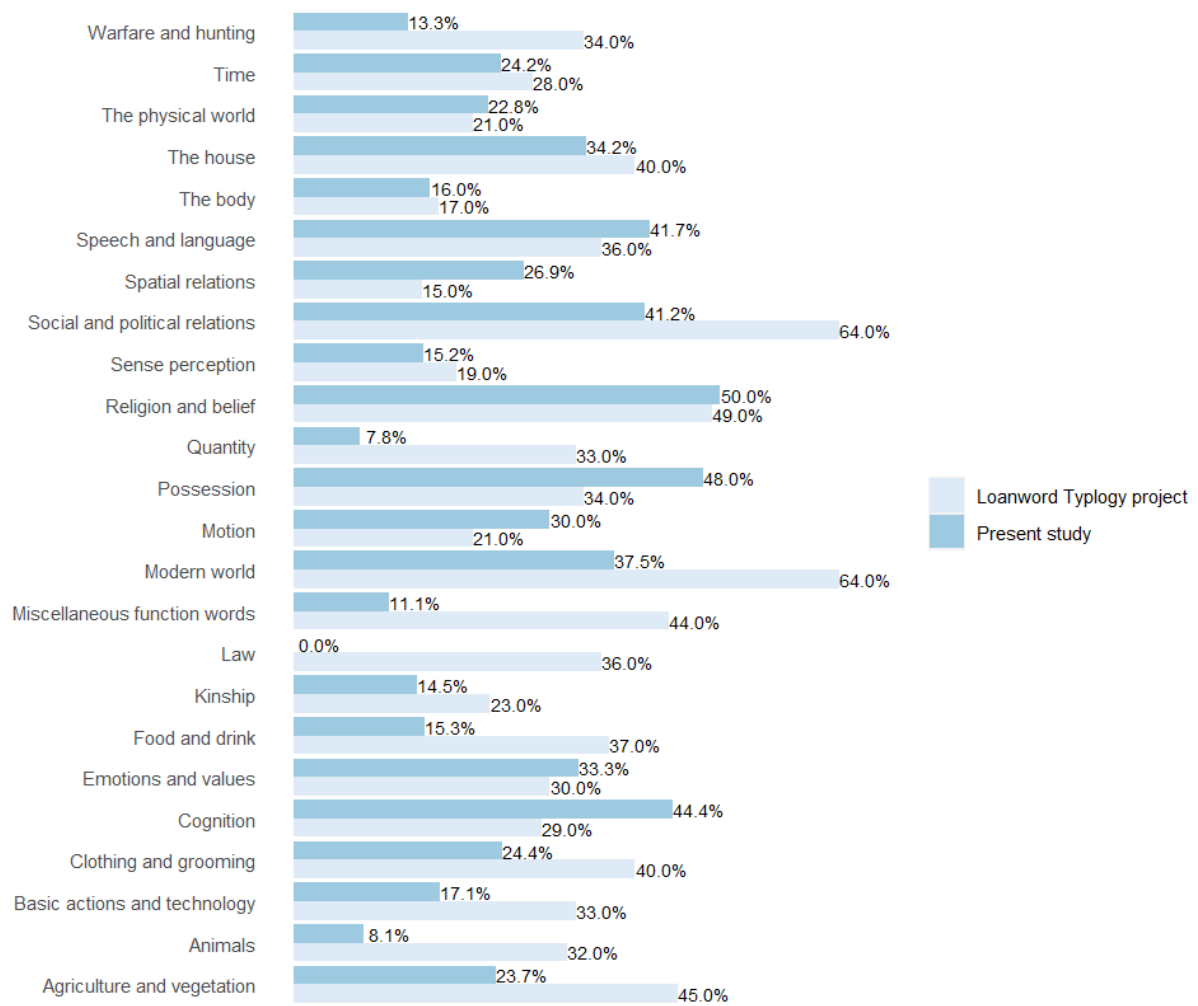


Figure 10. Bar plot comparing the borrowability score of the present study and the LWT project.

4.4. Semantic macro groups

As we can see from the tables shown previously there is a large amount of very small donor languages and semantic domains with very low amounts of borrowings. In order to make some further analysis and in an attempt to amass some larger numbers per category (donor language or semantic domain), I introduce groupings of the semantic domains and the donor languages. I divide the semantic domains used in the LWT project into three major semantic macro groups, namely concepts related to the *Individual*, to *Culture*, and to *Nature*. The semantic domains from Buck (1949) are somewhat opaque and therefore, in most cases, I went through the entire domain to see into which of the three macro groups it would better fit. *Individual* refers to concepts that in some way relate to your individual experience, such as emotions, body parts, senses. *Culture* refers to concepts linked to culture, material or social, such as cultural objects (cf. *The house*), social relations, human activities. *Motion* is a slightly confusing domain as it mostly contains words such as *road*, *path*, *boat*, *bridge*, which belong in a cultural domain. *Nature* refers to natural phenomena, animals, and spatial relations. Consult Table 16 to see which semantic domains belong to which semantic macro groups.

Table 16. Division of semantic domains from LWT into semantic macro groups.

<i>LWT Semantic domain</i>	<i>Semantic macro group</i>
<i>Basic actions and technology</i>	Culture
<i>Clothing and grooming</i>	Culture
<i>Food and drink</i>	Culture
<i>Kinship</i>	Culture
<i>Modern world</i>	Culture
<i>Motion</i>	Culture
<i>Possession</i>	Culture
<i>Quantity</i>	Culture
<i>Religion and belief</i>	Culture
<i>Social and political relations</i>	Culture
<i>Speech and language</i>	Culture
<i>The house</i>	Culture
<i>Time</i>	Culture
<i>Warfare and hunting</i>	Culture
<i>Cognition</i>	Individual
<i>Emotions and values</i>	Individual
<i>Sense perception</i>	Individual
<i>The body</i>	Individual
<i>Agriculture and vegetation</i>	Nature
<i>Animals</i>	Nature
<i>Spatial relations</i>	Nature
<i>The physical world</i>	Nature
<i>Miscellaneous function words</i>	Other

The results of borrowings from this division are presented in Table 17. This table shows some results that cannot really be seen as clearly from the semantic domains of the LWT project. For the purpose of analysing semantics the concepts relating to *Miscellaneous function words* is not considered a semantic domain, as these are very semantically weak words.

One clear result is that the least borrowed concepts are those related to the *Individual*, except in the case of Turkish, where it actually is the one borrowed into the most. In Western Farsi the semantic macro group of *Nature* is the one with the highest borrowing rate, while in Northern Kurdish it is the macro group *Culture* that is the highest. *Culture* is also considerably more borrowed into than *Nature* in Turkish. This could be an indication that Western Farsi is a culturally dominant language and therefore has a lesser need to borrow specifically cultural concepts, as Western Farsi is also a major donor of cultural borrowings.

Table 17. Borrowings per semantic macro group and recipient language.

Language	Semantic macro group	Number of borrowed word forms	Total number of word forms	Percentage
Western Farsi	Culture	43	204	21,1%
	Individual	20	115	17,4%
	Nature	32	132	24,2%
	Other	0	8	0,0%
Northern Kurdish	Culture	29	159	18,2%
	Individual	8	85	9,4%
	Nature	19	112	17,0%
	Other	1	3	33,3%
Turkish	Culture	61,5	218,5	28,1%
	Individual	33	108	30,6%
	Nature	25,5	123,5	20,6%
	Other	1	7	14,3%
Region	Culture	133,5	581,5	23,0%
	Individual	61	308	19,8%
	Nature	76,5	367,5	13,2%
	Other	2	26	7,7%

Investigating each language individually provides two major observations, see Table 18. Firstly, that Turkish has received more loanwords from Western Farsi (9.9%) than Arabic (8.4%) relating to concepts in the *Nature* semantic macro group, and secondly, that although Northern Kurdish has received most of its borrowings in *Individual* from Arabic (3.5%) and Turkish (3.5%), there are 0% from Western Farsi but 2.4% from Pahlavi and Old Persian together. This could maybe indicate that loanwords from Iranian languages have declined over time, in favour for Arabic and Turkish. Another very possible reason could be that it might be difficult to be certain whether words in Northern Kurdish are inherited from a previous language stage, or borrowed from Western Farsi, as they are two relatively closely related languages.

Table 18. Each donor language's contribution to the recipient languages per semantic macro domain. Blank cells indicate 0.00%. None represent the percentages of words that are not borrowed.

<i>Donor language</i>	<i>Northern Kurdish</i>			<i>Turkish</i>			<i>Western Farsi</i>		
	<i>Culture</i>	<i>Individual</i>	<i>Nature</i>	<i>Culture</i>	<i>Individual</i>	<i>Nature</i>	<i>Culture</i>	<i>Individual</i>	<i>Nature</i>
<i>Arabic</i>	9,49%	3,53%	9,73%	10,71%	14,41%	8,40%	13,81%	14,75%	22,96%
<i>Armenian</i>			0,88%		0,90%				
<i>Bulgarian</i>				0,45%					
<i>Chinese</i>							0,48%		
<i>Chagatai</i>				0,45%		0,76%			
<i>Cuman</i>					0,90%				
<i>Farsi</i>	1,27%			10,27%	12,61%	9,92%			
<i>French</i>				0,45%	0,90%		0,95%		
<i>Greek</i>				1,79%		2,29%	0,95%		
<i>Hebrew</i>				0,00%			0,48%		
<i>Hungarian</i>				0,45%					
<i>Italian</i>				1,34%					
<i>Mongolian</i>				0,45%					
<i>Russian</i>	0,63%								
<i>Serbian</i>				0,45%					
<i>Syriac</i>	1,90%								
<i>Turkish</i>	2,53%	3,53%	5,31%				2,38%		0,74%
<i>None</i>	82,28%	90,59%	82,30%	72,32%	69,37%	78,63%	79,05%	83,61%	76,30%
<i>Old Persian</i>	1,90%	1,18%	1,77%						
<i>Proto-Iranian</i>		1,18%							
<i>Sanskrit</i>							1,43%	1,64%	
<i>Sogdian</i>							0,48%		

4.5. Donor language per semantic domain

Apart from looking at the semantic macro groups it is interesting to see if there is any structure to be seen in the division of the donor languages into larger groups. I divide the donor languages into groups based on geographical position, language family, and main family branch, see Table 19.

Table 19. Division of donor languages into groups based on geography and genealogy.

<i>Donor language</i>	<i>Geographical region</i>	<i>Family</i>	<i>Branch</i>
<i>Arabic</i>	East Mediterranean	Afro-Asiatic	Semitic
<i>Armenian</i>	East Mediterranean	Indo-European	Armenic
<i>Bulgarian</i>	Europe	Indo-European	Balto-Slavic
<i>Chinese</i>	Central and East Asia	Sino-Tibetan	Sinitic
<i>Chagatai</i>	Central and East Asia	Turkic	Turkestan Turkic
<i>Cuman</i>	Central and East Asia	Turkic	Kipchak
<i>Western Farsi</i>	West and South Asia	Indo-European	Indo-Iranian
<i>French</i>	Europe	Indo-European	Italic
<i>Greek</i>	Europe	Indo-European	Graeco-Phrygian
<i>Hebrew</i>	East Mediterranean	Afro-Asiatic	Semitic
<i>Hungarian</i>	Europe	Uralic	Hungarian
<i>Italian</i>	Europe	Indo-European	Italic
<i>Latin</i>	Europe	Indo-European	Italic
<i>Mongolian</i>	Central and East Asia	Mongolic-Khitian	Mongolic
<i>Old Persian</i>	West and South Asia	Indo-European	Indo-Iranian
<i>Pahlavi</i>	West and South Asia	Indo-European	Indo-Iranian
<i>Proto-Iranian</i>	West and South Asia	Indo-European	Indo-Iranian
<i>Russian</i>	Europe	Indo-European	Balto-Slavic
<i>Serbian</i>	Europe	Indo-European	Balto-Slavic
<i>Sanskrit</i>	West and South Asia	Indo-European	Indo-Iranian
<i>Sogdian</i>	Central and East Asia	Indo-European	Indo-Iranian
<i>Syriac</i>	East Mediterranean	Afro-Asiatic	Semitic
<i>Turkish</i>	East Mediterranean	Turkic	Oghuz

Based on this we can see in Table 20 that in Northern Kurdish Afro-Asiatic (55.2%) is the largest donor language family, followed by Indo-European and Turkic tied at 22.4%. Highlights indicate the highest row percentage. We can, furthermore, see that almost all of the highlights are in the column for Afro-Asiatic borrowings.

Table 20. Contribution of language families to Northern Kurdish in terms of borrowings per semantic domain. Highlights indicate highest row percentage.

	Afro- Asiatic	Indo- European	Turkic
<i>Northern Kurdish</i>	55,2%	22,4%	22,4%
<i>Agriculture and vegetation</i>	33,3%	33,3%	33,3%
<i>Animals</i>	100,0%	0,0%	0,0%
<i>Basic actions and technology</i>	66,7%	33,3%	0,0%
<i>Clothing and grooming</i>	0,0%	50,0%	50,0%
<i>Cognition</i>	100,0%	0,0%	0,0%
<i>Emotions and values</i>	50,0%	0,0%	50,0%
<i>Kinship</i>	66,7%	33,3%	0,0%
<i>Miscellaneous function words</i>	0,0%	100,0%	0,0%
<i>Modern world</i>	100,0%	0,0%	0,0%
<i>Motion</i>	100,0%	0,0%	0,0%
<i>Possession</i>	100,0%	0,0%	0,0%
<i>Religion and belief</i>	100,0%	0,0%	0,0%
<i>Sense perception</i>	0,0%	50,0%	50,0%
<i>Social and political relations</i>	50,0%	25,0%	25,0%
<i>Spatial relations</i>	40,0%	40,0%	20,0%
<i>Speech and language</i>	100,0%	0,0%	0,0%
<i>The body</i>	33,3%	33,3%	33,3%
<i>The house</i>	66,7%	0,0%	33,3%
<i>The physical world</i>	63,6%	0,0%	36,4%
<i>Time</i>	60,0%	20,0%	20,0%
<i>Warfare and hunting</i>	0,0%	100,0%	0,0%

Turkish has borrowed more from Indo-European (54.1%) than from Afro-Asiatic (41.8%). Turkish also has a more diverse division of donor family per semantic domain, but with a majority of the fields under Indo-European, see Table 21.

Table 21. Contribution of language families to Turkish in terms of borrowings per semantic domain. Highlights indicate highest row percentage.

	<i>Afro-Asiatic</i>	<i>Indo-European</i>	<i>Mongolic-Khitani</i>	<i>Turkic</i>	<i>Uralic</i>
<i>Turkish</i>	41,8%	54,1%	0,8%	2,5%	0,8%
<i>Agriculture and vegetation</i>	28,6%	71,4%	0,0%	0,0%	0,0%
<i>Animals</i>	25,0%	75,0%	0,0%	0,0%	0,0%
<i>Clothing and grooming</i>	14,3%	85,7%	0,0%	0,0%	0,0%
<i>Cognition</i>	40,0%	60,0%	0,0%	0,0%	0,0%
<i>Emotions and values</i>	66,7%	33,3%	0,0%	0,0%	0,0%
<i>Food and drink</i>	100,0%	0,0%	0,0%	0,0%	0,0%
<i>Kinship</i>	66,7%	33,3%	0,0%	0,0%	0,0%
<i>Miscellaneous function words</i>	0,0%	100,0%	0,0%	0,0%	0,0%
<i>Modern world</i>	33,3%	33,3%	0,0%	33,3%	0,0%
<i>Motion</i>	50,0%	50,0%	0,0%	0,0%	0,0%
<i>Possession</i>	50,0%	50,0%	0,0%	0,0%	0,0%
<i>Quantity</i>	33,3%	66,7%	0,0%	0,0%	0,0%
<i>Religion and belief</i>	0,0%	100,0%	0,0%	0,0%	0,0%
<i>Sense perception</i>	33,3%	66,7%	0,0%	0,0%	0,0%
<i>Social and political relations</i>	22,2%	66,7%	11,1%	0,0%	0,0%
<i>Spatial relations</i>	30,0%	60,0%	0,0%	10,0%	0,0%
<i>Speech and language</i>	100,0%	0,0%	0,0%	0,0%	0,0%
<i>The body</i>	50,0%	43,8%	0,0%	6,3%	0,0%
<i>The house</i>	12,5%	75,0%	0,0%	0,0%	12,5%
<i>The physical world</i>	71,4%	28,6%	0,0%	0,0%	0,0%
<i>Time</i>	45,5%	54,5%	0,0%	0,0%	0,0%

Western Farsi is completely filled with Afro-Asiatic loans, as we can see in Table 22, which is hardly surprising since Arabic is the largest donor language overall for Western Farsi. Most semantic domains are dominated by Afro-Asiatic loans, but a few are from other families, such as Indo-European, and Turkic. The Sino-Tibetan loan is the word *chai* چای ‘tea’ and is only a single loan.

Table 22. Contribution of language families to Western Farsi in terms of borrowings per semantic domain. Highlights indicate highest row percentage.

	Afro- Asiatic	Indo- European	Sino- Tibetan	Turkic
<i>Western Farsi</i>	82,3%	10,4%	1,0%	6,3%
<i>Agriculture and vegetation</i>	100,0%	0,0%	0,0%	0,0%
<i>Animals</i>	100,0%	0,0%	0,0%	0,0%
<i>Basic actions and technology</i>	25,0%	75,0%	0,0%	0,0%
<i>Clothing and grooming</i>	0,0%	0,0%	0,0%	100,0%
<i>Cognition</i>	100,0%	0,0%	0,0%	0,0%
<i>Emotions and values</i>	100,0%	0,0%	0,0%	0,0%
<i>Food and drink</i>	50,0%	16,7%	0,0%	33,3%
<i>Kinship</i>	80,0%	20,0%	0,0%	0,0%
<i>Modern world</i>	0,0%	0,0%	100,0%	0,0%
<i>Motion</i>	50,0%	0,0%	0,0%	50,0%
<i>Possession</i>	75,0%	25,0%	0,0%	0,0%
<i>Quantity</i>	100,0%	0,0%	0,0%	0,0%
<i>Religion and belief</i>	0,0%	100,0%	0,0%	0,0%
<i>Sense perception</i>	66,7%	33,3%	0,0%	0,0%
<i>Social and political relations</i>	100,0%	0,0%	0,0%	0,0%
<i>Spatial relations</i>	100,0%	0,0%	0,0%	0,0%
<i>Speech and language</i>	100,0%	0,0%	0,0%	0,0%
<i>The body</i>	88,9%	11,1%	0,0%	0,0%
<i>The house</i>	100,0%	0,0%	0,0%	0,0%
<i>The physical world</i>	90,0%	0,0%	0,0%	10,0%
<i>Time</i>	100,0%	0,0%	0,0%	0,0%
<i>Warfare and hunting</i>	0,0%	100,0%	0,0%	0,0%

To sum up, what we can see from these tables are that some of the languages seem to be influenced by different families. Turkish is languages where not only one language family has been dominant as a donor language, while Western Farsi and Northern Kurdish have been influenced almost exclusively by Afro-Asiatic borrowings.

Combining the two divisions, i.e. that of family and semantic macro groups, gives Table 23. Interestingly, this pushes the languages into firmer spheres of lexical influence. Turkish is firmly influenced by Indo-European loans and Northern Kurdish and Western Farsi by Afro-Asiatic loans. This is an interesting observation in that even if Arabic, an Afro-Asiatic language, is the absolute giant of borrowings overall and in top position as a donor language in all three languages, Indo-European languages outweigh the Afro-Asiatic ones in Turkish. Another very interesting observation to be made is that the two Indo-European languages, Northern Kurdish, and Western Farsi, have received far less loans from Indo-European languages than the non-Indo-European language Turkish, despite the fact that all three languages are in direct contact with important Indo-European languages. Western Farsi is bordering Urdu speaking populations to the east and Armenian to the north, and Kurdish is spoken in

close proximity to Western Farsi and to Armenian. Thus, there is a potentially a case to be made that the dataset might indicate that the language genealogy has less impact on lexical borrowings than other factors.

Table 23. Borrowings for each language with respect to donor language family and semantic macro groups. Highlights indicate highest percentage per semantic macro group (lack of loans excluded). None the percentage or words that are not borrowed.

	Afro-Asiatic	Indo-European	Mongolic-Khitian	None	Sino-Tibetan	Turkic	Uralic
<i>Northern Kurdish</i>	8,9%	3,6%	0,0%	83,9%	0,0%	3,6%	0,0%
<i>Culture</i>	11,3%	3,8%	0,0%	82,4%	0,0%	2,5%	0,0%
<i>Individual</i>	3,5%	2,4%	0,0%	90,6%	0,0%	3,5%	0,0%
<i>Nature</i>	9,7%	2,7%	0,0%	82,3%	0,0%	5,3%	0,0%
<i>Other</i>	0,0%	50,0%	0,0%	50,0%	0,0%	0,0%	0,0%
<i>Turkish</i>	10,8%	14,0%	0,2%	74,2%	0,0%	0,6%	0,2%
<i>Culture</i>	10,8%	14,9%	0,5%	73,0%	0,0%	0,5%	0,5%
<i>Individual</i>	14,6%	14,6%	0,0%	70,0%	0,0%	0,9%	0,0%
<i>Nature</i>	8,4%	12,2%	0,0%	78,6%	0,0%	0,8%	0,0%
<i>Other</i>	0,0%	10,0%	0,0%	90,0%	0,0%	0,0%	0,0%
<i>Western Farsi</i>	16,6%	2,1%	0,0%	79,8%	0,2%	1,3%	0,0%
<i>Culture</i>	14,3%	3,8%	0,0%	79,1%	0,5%	2,4%	0,0%
<i>Individual</i>	14,8%	1,6%	0,0%	83,6%	0,0%	0,0%	0,0%
<i>Nature</i>	23,0%	0,0%	0,0%	76,3%	0,0%	0,7%	0,0%
<i>Other</i>	0,0%	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%

5. Discussion

5.1. Recipient languages

The recipient languages in this thesis, Northern Kurdish, Turkish, and Western Farsi, are languages not covered by the major study conducted by Haspelmath and Tadmor (2009), and any information gleaned from the data collected from these languages might prove valuable. The languages are of two different categories of borrowers, i.e. Turkish is categorised as a high borrower, and Western Farsi and Northern Kurdish as average borrowers. That Turkish is a high borrower language is not surprising, since it is a language that has been politically, religiously, and socially influenced by both Arabic and Western Farsi speaking communities. Northern Kurdish is a language that is spoken across borders in Western Asia but is not recognised as an official language in the countries where it is most widely spoken. As such it would not be surprising to see Northern Kurdish as a potential high borrower of lexical units, as it would draw from the other more socially and politically established national languages in the region, such as Arabic, Western Farsi, and Turkish. Western Farsi, on the other hand, is an exceedingly important language in the region historically and culturally. Western Farsi was also used during the Ottoman Empire as a courtly language and was used extensively as a literary language connecting regions across Asia (Dabashi, 2012). What is interesting then is that Western Farsi, a prestige language and dominant in many social, cultural, and political domains, has a higher borrowing rate than Northern Kurdish. From the data we can see that almost all borrowings in Western Farsi are from Arabic, around 80%, followed by Turkish at the considerably more modest 6%. The borrowings seem to thus come from other prestige languages, as Arabic and Turkish represent the two major languages in the area. As for the difference between Northern Kurdish and Western Farsi, it might be due to methodological limitations.

Turkish is in the dataset the highest borrower, with most loans coming from Arabic and Western Farsi. This is hardly surprising considering the linguistic history of Turkish during the Ottoman Empire. Ottoman Turkish was influenced by Arabic and Western Farsi to such an extent that some researchers have argued for it being a mixed language (cf. Bakker, 2008; Németh, 1953). While this is not the current view of the matter, it gives us some indication to why Arabic and Western Farsi are the main donor languages. What is fascinating, however, is that despite the massive language reforms that, in an attempt to, among other things, rid the Turkish language of its Arab and Persian influence, Atatürk (considered to be the founder of modern-day Turkey) introduced in early 20th century, Turkish remains a language containing many borrowings from the two neighbouring languages. Another interesting aspect of this relationship is that Arabic and Western Farsi have given rise to the same number of loanwords in Turkish. It is worth to note that Arabic was spoken extensively within the borders of the Ottoman Empire, while Western Farsi was not, and that Western Farsi was used as a courtly language. It seems that Turkish has borrowed concepts relating to the *Individual* mostly from Arabic, *Nature* from Western Farsi, and almost tied Arabic and Western Farsi for *Culture*. Examples of concepts belonging to the

Individual group are concepts for body parts or functions, mental states and emotions, and perception. Possibly, since many of these relate to medicine, the explanation for this could be that Arabic has historically been the main scientific language in the region (Dabashi, 2012).

As we can see from the data there seems to be a rather large difference in the amount of loans absorbed by the different semantic domains (Figure 10). The only semantic domains that come close in borrowing rate when comparing the region to the LWT global average are *Modern world, Social and political relations, The house, and Kinship* as well as *Religion and belief*. Although I will not hazard a guess as to the potential reasons behind the considerable differences, I will argue that it strengthens the importance of studying linguistic areas specifically in terms of loanwords, rather than global crosslinguistic samples or individual languages. These are, of course, also very important aspects to be studied, but region-specific research will provide different results that cannot be discerned or inferred from studying single languages. That is, in order to better understand the context in which a language is situated, we need to consider not only the language in isolation but also its relation to neighbouring languages.

One interesting aspect of this is that there is not necessarily any need for two linguistic communities to be in contact for their languages to be so. Two languages can, in some sense, be in contact through another language, a *relay language*. Now, for this to have any meaning, some assumptions have to be accepted. Firstly, this is a purely theoretical construction. What I mean by this is that the speakers of the languages in question do not need to be aware of the original source of the words, but that we, as researchers, are and that is what gives meaning to the origin language being in contact with the recipient language. Secondly, that lexical loans carry with them their etymological history and origin. In other words, lexical borrowings might be changed by the relay language, but should be considered a word from the original source. One could perhaps see this in terms of lexical genealogy. Each word has its own evolutionary tree, unique to itself, and it cannot be changed retroactively, its origin stays fixed, and as such it is ultimately a word from one language. Thirdly, this should not be considered in a contact situation where you are interested in the direct contact between languages and their speakers, searching for the ultimate origin of a word makes no sense, as this is no indication of the speakers being in contact.

5.2. Signs of cultural contact

It is clear that Arabic is the major donor language in the area, followed by Western Farsi. This in itself is indicative of their importance in the region, and although we have much data to indicate that high status languages tend to become major donor languages, the results from this thesis further support the argument.

One interesting change in donor language importance can perhaps be seen in Northern Kurdish. In Northern Kurdish, while Arabic is the clear major donor language, there is a lack of loans from Western Farsi but we can see that there are some loans from Old Persian and Pahlavi. This could possibly

allude to a shift in language importance. One can imagine that Northern Kurdish borrowed more from Old Persian and Pahlavi as they were very important languages before Arabic, but as Arabic spread through the region speakers of Northern Kurdish began to favour Arabic loans over loans from Western Farsi, the modern descendent of Old Persian and Pahlavi. All in all it seems as if Arabic and Turkish are the main prestige languages in relation to Northern Kurdish.

Although Western Farsi has seemingly had very little effect on the lexicon of Northern Kurdish, it is an important donor language in the region. It does not seem as if being a major donor language affects the languages borrowing rate negatively, as can be seen from the data, with Western Farsi being the second largest recipient of loans, and by previous research (Carling et al., 2019). However, I cannot help but note that the semantic domain of *Social and political relations* has high borrowing scores in Northern Kurdish and Turkish, and Western Farsi furthermore being the donor of 25% of the Northern Kurdish and 55.57% of the Turkish lexicon for that very domain. So, overall, it does not seem to be the case that a language being a major donor language, and thus likely a high prestige language, negatively impacts the borrowing rate of the given language, but, interestingly, there might be a relation between giving and receiving loans from particular semantic domains. If Western Farsi is a prestige language in social and political domains, as was the case in the Ottoman Empire, it makes sense for other languages to borrow Western Farsi word forms for important concepts in those domains, while Western Farsi already in possession of these high prestige words would not necessarily want to exchange them for others.

Another interesting find, which is indicated by Thomason and Kaufman (1988) supported by previous research (Brown et al., 1994), is that genealogical factors seem to have a lesser impact on which languages a recipient language borrows from. This would, to me, indicate that borrowings, as based on the results of this thesis, are more tightly connected to cultural elements than any other. Most donor languages, especially the large ones, in the region are also geographically in close proximity to the recipient languages, Western Farsi, Northern Kurdish, and Turkish, and thus probably is less significant in the impact on the languages' lexicon.

Finally then, it can be said that there are indeed some visible signs of cultural contact to be seen in the lexical data. The first is that of the major importance of Arabic and Western Farsi, from which speakers of other languages have borrowed many lexical items. Even Turkish has inspired quite a few loans, so we can, in some sense, see three potential centres for borrowing in Arabic, Western Farsi, and Turkish. It is thus possible to, based on the multipolar borrowing centres, establish some relative status, although it becomes somewhat difficult with Arabic, since we have no information as to their borrowing situation. Arabic is a donor language to all, indicating some deep cultural ties to the rest of the region. Western Farsi is the second largest donor language and also displays a major role in its cultural influence in the region, surpassing Arabic in some cases. It is, however, itself almost exclusively influenced by Arabic. This would, to my mind, indicate a very high status of Western Farsi, only matched, or bested by that of Arabic. It would seem as well, at least superficially, that Arabic's high presence in the semantic

domains used in this study is connected to words relating to the sciences, while Western Farsi might conceivably, be closer associated with more cultural concepts. It can also be worth to mention that Arabic has a very large number of speakers in Western Asia, and has had over a long period of time, which might also contribute to its influence in the neighbouring languages' lexicon.

5.3. Methodological implications

There are many types of decision that are made when deciding on a methodology, and they provide initial conditions that greatly affect the final result of any study. I would like to, in particular, discuss a couple of those choices here, namely what a loanword is and what counts as a loanword, the discrepancy between language and dictionary, and the ramifications of using a meta-language such as English.

To define what a loanword is no simple task. For example, one could include within the term loanword or borrowing, any element that has at some point in a language's history been adopted by its speakers. One could perhaps argue that a loan has to be recognised as such by its speakers, and once naturalised morphologically and phonologically to the recipient language and no longer seen as a "foreign" element, it has changed substantially enough for it to be a construction of the recipient language and not a loan. Alternatively, one could perhaps view only borrowings that are from the contemporary language stage to be actual borrowings, as any borrowing from previous language stages have, in fact, been inherited. All of these approaches will dramatically affect what counts as a loanword. In many respects, this is a theoretical discussion, based on whether we consider loans to be carriers of a history in and of themselves, which we, as researchers, observing from afar, are able identify the origin of, or whether loans ought to be what users actually perceive them to be. Borrowings should therefore be recognised as theoretical constructions, and the exact definition of which to vary depending on the needs of the study conducted, and critically discussed accordingly.

In addition to the questions around letting time and adaption to the recipient language be a defining feature of borrowings, there are also problems with considering absence of analysability to be a defining feature of borrowings. The LWT project considers morphologically analysable word forms to not be borrowed, an approach adopted in this study. This has very likely affected the outcomes and can be a contributing factor as to why there are fewer loans per semantic domain than in the LWT project. It was adopted because part of this thesis is to compare the languages in the region to the languages in the LWT project, and I felt there could not be too much distance between the data collected for this thesis and that of the LWT project. However, the aim of the LWT project is not as much to study language contact as borrowability, and in that context it makes sense to not consider morphologically analysable word forms to be borrowings. From the perspective of language contact, however, it makes much less sense, as a morpheme, or lexeme used in compounds, borrowed from another language is still an indication of language contact. This is a choice I probably would change in any future study on language contact. A potential solution to this issue could have been to consider all unknown loans as "half borrowed" or "maybe borrowed" and given a borrowability score of 0.5. This would make the data

even more abstract and it would rely on the assumption that roughly half of the unknowns are borrowed and half unlikely to be so, thus I feel it risks inflating the number of borrowed terms, and this approach, furthermore, cannot be used when analysing for donor languages as this is simply not known. All in all, it can have a positive impact on the overall borrowability scores, but at the cost of some uncertainty and would not be useful when looking at cultural contact at all, which is why this is not an approach I adopted.

Returning to the issue of the slight discrepancy between contemporary language use and a more historically holistic study, there is a noteworthy discordance also between language in use and dictionaries. While language contact is a process, and inherently social in its workings, the construction of a dictionary is not necessarily so. Dictionaries are often based on standardised language varieties, and, in their essence, not as prone to change as language in use. Used language, however, belongs on a continuum, where the boundaries between different lects are by definition fuzzy, and this discrepancy makes it unpractical to precisely capture any language in a dictionary. By lects, I here allude to different kinds of varieties of a language, such as idiolects, sociolects, and dialects. This does not make them in any way useless as an object of study, but it becomes important to be aware of when studying loanwords. In a standard language it becomes possible to make choices on what to include and what to exclude, and they do not necessarily represent the same complexities in terms of language contact, as a dialect would, rather, they will perhaps exhibit the more permanent and by cultural institutions accepted changes. In this thesis standard languages are the objects of study, but there could be much of value for language contact research to focus, not only on standard languages, but dialects and sociolects, in order to deepen our understanding of borrowing as a process, as was my initial aim, although unfortunately not realisable due to circumstances.

Another point to discuss is the use of a meta-language such as English. Semantic typology is an important field of study to further our understanding of crosslinguistic semantic tendencies. The use of one language to discuss another language's semantics is problematic since the semantic granularity of languages differ significantly. A lucid example of this is the difference in semantic scope of the verb *to think* in English and the equivalents *att tänka* 'to cogitate', *att tycka* 'to have an opinion' and *att tro* 'to believe' in Swedish. All of three Swedish counterparts have very different meanings, but they are all captured with one word in English, and by using English one might be led to think of the certain differences and similarities as more or less prominent than one would by using Swedish. Thus, by using English as a meta-language in this study, to give meaning to the concepts might very well lead to enforcing English semantic granularity onto the languages studied here. This approach is still adopted in this thesis, however, because there are no other practical ways to deal with long wordlists in many different languages. This has likely had impact on the final results, with certain words being counted fewer or more times than they should have, as they are subsumed or split by force. Some measures are implemented to deal with this, however, for example, the decision to split borrowability scores across polysemous words. Using the previous example with *to think* and *att tänka*, *att tycka*, *att tro*, this would

mean that if studying English, using Swedish, we would get three pairs *to think – att tänka, to think – att tycka, to think – att tro*, all of which would get one third each, since English *to think* is only one word and should only be counted once. Using English to study Swedish, you would have to divide the three up into three different concepts and give them one point each. A typical example of this in the data is the terms aunt and uncle, which in most of the languages of this thesis have much finer distinctions than the English binary. One could perhaps imagine using a meta-language based on the idea of *semantic primes*. The idea behind semantic primes is to “decompose [complex meanings and culture-specific meanings] into simple concepts which can be found in all languages” (Goddard & Wierzbicka, 2014, p. 2). To use semantic primes would take considerable effort, perhaps a whole thesis’ worth in itself, and, furthermore, the theory has been criticised and several problems have been identified (Evans, 2010). It is doubtful whether it would perform better.

5.4. Theoretical implications

Some of the results found in this study could have more general theoretical implications. While the data collected for this thesis might not be perfect and cannot account for a complete overview of the region, there are some interesting findings.

One of these findings, which in itself might not be a striking observation, but one that I have not seen explicitly mentioned in the literature, is the even distribution of loans. It would seem that if a language is a high borrower, it will be higher in all semantic domains relative to a language that is an average or low borrower. There are two implications to be observed. The first implication is that languages have even distribution of loanwords, i.e. there is no language that has no borrowings in half of the semantic domains and gone through an almost complete shift in the other half. The second implication would be that one could look at a single semantic domain, which tends to be very high in borrowings, and based on only that make an estimate as to how open to borrowings a language is. In some ways, this is the inverse of the basic vocabulary lists, where the aim is to measure resistance to borrowings.

Another question that can be raised from inspecting the data, is whether language genealogy has a lesser impact on borrowings than other factors, such as social, cultural, political, or geographical ones. It is not a controversial statement, as it has been shown that language typology plays a limited to negligible role in the effects of more intense language contact situations (cf. Thomasson & Kaufman, 1988). It is worthwhile to remember that the foundation to all language contact lies in the behaviour of speakers and is thus a phenomenon deeply rooted in the social. The results of the current study can be interpreted as supporting the idea that genealogy and typology has a weaker influence as compared to other factors, including such notions as prestige and dominance. This is also in line with contemporary approaches language contact, and supports previous studies (cf. Brown et al., 1994).

5.5. Future research

Before delving into more fanciful ideas to extend this research and expand into other fields, I would have liked to include a much larger set of languages, including for example Azerbaijani, Armenian, Arabic, Central Kurdish, redo the data collection Georgian using better sources, as well as a much more exhaustive word list including perhaps all of the words in the original LWT word list for maximal comparative purposes. Especially Arabic would be an interesting addition, as it is also a major donor language.

The first matter I would like to raise, is the study of semantic change. Originally this was a core part of the thesis and was what initially drove me to consider this thesis project to start with. Unfortunately, it simply had to be cut back due to the time provided to write this thesis. There are many interesting examples of curious semantic changes visible through loanwords and etymology. Some of these semantic changes could be studied by looking at how the words have developed in the different languages they have been adopted by. It is not impossible that, due to the difference in semantic typology and make-up, loanwords will change differently, and I suspect some of these changes to have certain structure to them. There is also a particularly curious example of semantic change that I stumbled upon during the data collection, and it goes as follows: *fincan* ‘cup’ in Kurmaji, from Arabic *finjān* ‘cup’, from Western Farsi *pengān* ‘cup’, from Aramaic *pinka* ‘dish, plate’, from Greek *pinax* ‘board, plank, tablet, platter, register’. This type of semantic change could be categorised as different types of conceptual changes. By collecting many more words, and reviewing their semantic changes, categorising them in relation to conceptual and semantic change, time, and donor/recipient language and inheritance, there is a distinct possibility that some structure of semantic change could emerge. Part of this is to also include a chain of borrowings by investigating the etymology of the source words of each borrowing and continuing doing that several steps back into the loanwords history.

Another extension of this thesis would be to study not just standard languages but dialects specifically. This can then be more directly connected to behavioural approaches to language contact. Looking at specific dialects can also make it possible to study different kinds of intensity of language contact, and how different compositions of surrounding languages might affect the outcome of language contact. Some examples could be, studying the Georgian speaking enclave in Iran as compared to different dialects in Georgia, comparing different types of Hebrew-based languages, of which there are a multitude in Western Asia (Kahn & D. Rubin, 2015), or looking at dialects in areas of varying linguistic diversity.

These are some of the ideas of how one could proceed, following from this thesis. Some might be implausible, and most probably need serious revision, but I do believe the underlying questions motivating these inquiries are of value and should be pursued at some capacity.

6. Conclusion

Language contact is a field of study deeply connected to a multitude of different disciplines within linguistics, including but not restricted to socio-linguistics, psycho-linguistics, and historical linguistics. Studying language contact can give us fascinating insights into the historical and cultural ties between languages and the specific field of language contact concerned with borrowings is a very tangible way to see the effects left on a language from another.

The aim of this thesis is to investigate the language contact situation in Western Asia, a region with many languages from different families and typological dispositions, but also deeply intertwined cultures, that have been in constant exchange and contact. The languages studied in particular here are Northern Kurdish, Western Farsi, and Turkish. Using a methodology based on the Loanword Typology project I have studied the lexicons of these languages with the intention of describing and analysing their borrowing composition. More specifically, this thesis compares the languages in terms of borrowing rate with the Loanword Typology project and investigates the linguistic contact area of the region. The following research questions are asked:

1. How do the languages Northern Kurdish, Western Farsi, Turkish, and Georgian compare to the languages analysed by the Loanword Typology project?
2. Which are the main donor languages in the region?
3. What signs of linguistic and/or cultural contact can we see through the composition of loanwords? That is, what can we say about any potential correlation between donor language and semantic domain or amount of borrowings vis-à-vis recipient language?

The results of the thesis can provide some answers to these questions. It should be noted that since Georgian is not part of the analysis, it has to be excluded also from the first research question. However, in regard to the first research question we can ascertain that, if compared to the results of the Loanword Typology project, Turkish can be considered a high borrowing language, Northern Kurdish and Western Farsi as average borrowing languages. The region, with a borrowing rate of around 21.5%, when considered as one unit would come at an average borrower place. The second research question has the very simple answer, Arabic is unquestionably the dominant donor language for all languages studied, followed by Western Farsi. But there are also a number of other donor languages present amounting to around 20% of the total number of borrowings, which means that the area is also some diversity with regard to borrowings. The third research question can be addressed with the following observations, judged by the massive Arabic influence in the three languages, it seems likely to conclude that Arabic has been a very prestigious or dominant language in the region. Furthermore, there is some

indications to say that Arabic loans are especially geared towards concepts relating to sciences, while Western Farsi, as the second largest donor language, is more present in cultural loans. Remember, however, that Arabic has a large presence also in the cultural loans. It would also seem that Northern Kurdish has been more influenced by earlier stages of Persian than the modern stage, possibly because of the importance of Arabic. Furthermore, the data indicates that if a language is a major donor language in a semantic domain it will also borrow less into the very same domain. In addition to all this, there is no obvious structure to the differences in borrowability of the semantic domains as the borrowability differs considerably from the results of the LWT project.

It should be kept in mind that the data is by default tricky to analyse by virtue of difficulties in accessing material, methodological choices concerning what should be counted as a loanword that arose from the material, differences in semantic granularity between the languages, the somewhat coarse grouping of semantic macro domains, the potential bias in the dictionaries in regards to which word forms to include, and the problems in assessing the differences between lexis within a language which affect the inclusion or exclusion of certain apparent synonyms.

The findings of this thesis are more or less in line with what could be expected from previous research. The main contribution, however, is the choice of languages and the regional focus. None of the languages studied in this thesis are part of the Loanword Typology project, which, gives some important indications regarding the borrowability of Northern Kurdish, Western Farsi, and Turkish. Finally, this thesis also puts emphasis on the importance of studying languages in contact in a linguistic contact area, rather than as single entities.

7. References

- Bakker, P. (2008). Mixed languages as autonomous systems. In Y. Matras & P. Bakker (Eds.), *The Mixed Language Debate*. De Gruyter Mouton.
- Bakker, P., & Matras, Y. (Eds.). (2013). *Contact Languages: A Comprehensive Guide* De Gruyter Mouton. <https://doi.org/doi:10.1515/9781614513711>.
- Berg, D., Berg, S., Hult, A.-K., Holmer, L., Sjögreen, C., Sköldberg, E., Gronostaj, M. T., & Malmgren, S.-G. (2023). *Lexin*. Kungliga Tekniska Högskolan. <https://lexin.nada.kth.se/lexin/>
- Brown, C. H., Callaghan, C. A., Casson, R. W., Etxebarria, J. M., Hill, J. H., Lehman, F. K., Oswalt, R. L., & Proulx, P. (1994). Lexical Acculturation in Native American Languages [and Comments and Reply]. *Current Anthropology*, 35(2), 95-117. <http://www.jstor.org/stable/2744174>
- Buck, C. D. (1949). *dictionary of selected synonyms in the principal Indo-European languages: a contribution to the history of ideas*. Chicago University Press.
- Campbell, G. L., & King, G. (2013). *Compendium of the world's languages*.
- Carling, G. (2019). *The Mouton atlas of languages and cultures*. <https://doi.org/10.1515/9783110367416-202>
- Carling, G., Cronhamn, S., Farren, R., Aliyev, E., & Frid, J. (2019). The causality of borrowing: Lexical loans in Eurasian languages. *PLoS ONE*, 14(10), e0223588. <https://doi.org/10.1371/journal.pone.0223588>
- Chyet, M. L. (2003). *Kurdish-English Dictionary Ferhengê Kurmancî-Inglîzî*. Yale University Press.
- Čikobava, A. (n.d., 2023 April). *Kartuli enis ganmart'ebiti leksik'oni [Georgian Explanatory Dictionary]*. ენის მოდელირების ასოციაცია [Language Modeling Association]. <http://ena.ge/explanatory-online>
- Comrie, B. (2000). Language contact, lexical borrowing, and semantic fields. *Studies in Slavic and General Linguistics*, 28.
- Dabashi, H. (2012). *The world of Persian literary humanism*. Harvard University Press.
- Dehkhoda, A. A. (n.d.). *Dehkhoda Dictionary* [Online dictionary]. Dehkhoda Lexicon Institute and International Center for Persian Studies. <https://dehkhoda.ut.ac.ir/fa/dictionary>
- Dellert, J., Daneyko, T., Münch, A., et al., & Lang Resources & Evolution. (2019). Version 0.9). <https://doi.org/https://doi.org/10.1007/s10579-019-09480-6>
- Dunn, M. (2014). Language phylogenies In Claire Bower & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 190-211). Taylor & Francis Group.
- Evans, N. (2010). Semantic typology. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 504-533). Oxford University Press.
- Fähnrich, H. (2007). *Kartwelisches Etymologisches Wörterbuch*. Brill. <https://doi.org/https://doi.org/10.1163/ej.9789004161092.i-874>

- Goddard, C., & Wierzbicka, A. (2014). *Words and meanings. lexical semantics across domains, languages, and cultures* (First edition. ed.). Oxford University Press.
- Google. (2023). *Google Translate*. Google. <https://translate.google.com/?hl=sv&ie=UTF-8&sl=auto&tl=sv>
- Haspelmath, M. (2008, 01/01). Loanword typology: Steps toward a systematic crosslinguistic study of lexical borrowability. *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*, 43-62 (2008).
- Haspelmath, M., & Tadmor, U. (2009). *Loanwords in the world's languages a comparative handbook*. De Gruyter Mouton.
- Haugen, E. (1950). The Analysis of Linguistic Borrowing. *Language*, 26(2), 210-231. <https://doi.org/10.2307/410058>
- Kahn, L., & D. Rubin, A. (Eds.). (2015). *Handbook of Jewish Languages*. Brill. <https://doi.org/https://doi.org/10.1163/9789004297357>.
- Key, M. R., & Comrie, B. (Eds.). (2023). *The Intercontinental Dictionary Series* [Online database, accessed Feb 2023]. Max Planck Institute for Evolutionary Anthropology. <https://ids.clld.org/>.
- Klimov, G. A. (1998). *Etymological dictionary of the Kartvelian languages*. Mouton de Gruyter.
- List, J. M., Tjuka, A., van Zantwijk, M., Blum, F., Ugarte, C. B., Rzymiski, C., Greenhill, S., & Forkel, R. (Eds.). (2023). *CLLD Concepticon 3.1.0* [Online data set, accessed Feb 2023]. Max Planck Institute for Evolutionary Anthropology. <https://concepticon.clld.org/>.
- Matras, Y. (2020). *Language contact* (2nd ed.). Cambridge University Press.
- Matras, Y., & Bakker, P. (Eds.). (2008). *The Mixed Language Debate: Theoretical and Empirical Advances*. De Gruyter Mouton. <https://doi.org/doi:10.1515/9783110197242>.
- Muysken, P. (2000). *Bilingual speech a typology of code-mixing*. Cambridge University Press.
- Myers-Scotton, C. (2002). *Contact linguistics. Bilingual encounters and grammatical outcomes*. Oxford University Press.
- Németh, J. (1953). Zur Kenntnis der Mischsprachen (Das doppelte Sprachsystem des Osmanischen.). *Acta Linguistica Academiae Scientiarum Hungaricae*, 3(1/2), 159-199. <http://www.jstor.org/stable/44309058>
- Nişanyan, S. (2023). *Nişanyan Sözlük* [Online dictionary].
- Nourai, A. (2013). *An etymological dictionary of Persian, English and other Indo-European languages*. Xlibris Corp.
- Pereltsvaig, A., & Lewis, M. W. (2015). *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge University Press. <https://doi.org/DOI:10.1017/CBO9781107294332>
- Renfrew, C. (2000). At the Edge of Knowability: Towards a Prehistory of Languages. *Cambridge Archaeological Journal*, 10(1), 7-34. <https://doi.org/10.1017/S0959774300000019>

- Rzymiski, C., Tresoldi, T., & et al. (2019). *The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies* [Online database, accessed Feb 2023] [https://doi.org/DOI: 10.1038/s41597-019-0341-x](https://doi.org/DOI:10.1038/s41597-019-0341-x)
- Shijulal, N.-S., Johann-Mattis, L., Hans, G., Heiner, F., Russell, D. G., William, M., & Tal, D. (2011, 06/22/). Networks uncover hidden lexical borrowing in Indo-European language evolution [research-article]. *Proceedings: Biological Sciences*, 278(1713), 1794-1803. <http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.41314858&site=eds-live&scope=site>
- Swadesh, M. (1952, 08/22/). Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos [research-article]. *Proceedings of the American Philosophical Society*, 96(4), 452-463. <http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.3143802&site=eds-live&scope=site>
- Tadmor, U., Haspelmath, M., & Taylor, B. (2010, 10/19). Borrowability and the Notion of Basic Vocabulary. *Diachronica*, 27, 226-246. <https://doi.org/10.1075/dia.27.2.04tad>
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics* [Book]. Univ. of California Press. <http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat07147a&AN=lub.977998&site=eds-live&scope=site>
- Tietze, A. (2002). *Tarihi ve etimolojik Türkiye Türkçesi lügatı / 1, A - E* (1. baskı ed.). Simurg.
- Tietze, A., Demir, N., Yılmaz, E., & Türkiye Bilimler, A. (2021a). *Tarihî ve etimolojik Türkiye Türkçesi lügatı 2. cilt C-E* (İkinci baskı ed.). Türkiye Bilimler Akademisi.
- Tietze, A., Demir, N., Yılmaz, E., & Türkiye Bilimler, A. (2021b). *Tarihî ve etimolojik Türkiye Türkçesi lügatı 5. cilt M-P* (İkinci baskı ed.). Türkiye Bilimler Akademisi.
- Tietze, A., Tezcan, S., & Türkiye Bilimler, A. (2019). *Tarihî ve etimolojik Türkiye Türkçesi lügatı 7. cilt, S-Ş* (Birinci baskı ed.). Türkiye Bilimler Akademisi.
- Tsabolov, R. L. (2001). *Etymological Dictionary of Kurdish*. Academy of Sciences.
- Van de Velde, F., & De Smet, I. (2021). Markov Models for Multi-state Language Change. *Journal of Quantitative Linguistics*, 1-25. <https://doi.org/10.1080/09296174.2021.1877004>
- van Hout, R., & Muysken, P. (1994). Modeling lexical borrowability. *Language Variation and Change*, 6, 39-62.

Appendix 1

The table showcases every concept used in this thesis together with its LWT ID number, NorthEuraLex ID and the lexical concept in English. The digits preceding the decimals indicate which LWT semantic domain the concept belongs to, consult the first table.

<i>ID</i>	<i>Semantic domain</i>
1	The physical world
2	Kinship
3	Animals
4	The body
5	Food and drink
6	Clothing and grooming
7	The house
8	Agriculture and vegetation
9	Basic actions and technology
10	Motion
11	Possession
12	Spatial relations
13	Quantity
14	Time
15	Sense perception
16	Emotions and values
17	Cognition
18	Speech and language
19	Social and political relations
20	Warfare and hunting
21	Law
22	Religion and belief
23	Modern world
24	Miscellaneous function words

LWT ID number	NEL ID	lexical concept	LWT ID number	NEL ID	lexical concept
13.105	hundert::num	a hundred	4.45	leber::n	liver
13.106	tausend::num	a thousand	7.23	schloss::n	lock
12.081	oben::adv	above	12.57	lang::a	long
14.35	wieder::adv	again	2.21	mann::n	man
14.12	alter::n	age	19.41	meister::n	master
1.71	luft::n	air	5.41	essen::n	meal
13.331	allein::adv	alone	5.61	fleisch::n	meat
14.31	immer::adv	always	4.88	arznei::n	medicine
17.51	und::cnj	and	12.37	mitte::n	middle
3.11	tier::n	animal	5.86	milch::n	milk
3.817	ameise::n	ant	5.87	melken::v	to milk
4.31	arm::n	arm	17.11	verstand::n	mind

20.25	pfeil::n	arrow	6.96	spiegel::n	mirror
1.84	asche::n	ash	16.77	fehler::n	mistake
2.52	tante::n	aunt	14.63	montag::n	monday
4.19	rücken::n	back	11.43	geld::n	money
16.72	schlecht::a	bad	14.71	monat::n	month
8.75	rinde::n	bark	1.53	mond::n	moon
3.73	bär::n	bear	14.44	morgen::n	morning
4.142	bart::n	beard	2.36	mutter::n	mother
16.81	schön::a	beautiful	4.24	mund::n	mouth
7.42	bett::n	bed	8.98	pilz::n	mushroom
12.04	vorher::adv	before	9.5	nagel::n	nail
12.011	hinter::prp	behind	4.99	nackt::a	naked
4.431	bauch::n	belly	18.28	name::n	name
6.57	gürtel::n	belt	12.62	schmal::a	narrow
12.55	groß::a	big	4.43	navel::n	navel
8.63	birke::n	birch	12.43	nah::a	near
3.581	vogel::n	bird	4.28	hals::n	neck
15.37	bitter::a	bitter	3.58	nest::n	nest
15.65	schwarz::a	black	14.34	niemals::adv	never
7.422	decke::n	blanket	14.13	neu::a	new
4.97	blind::a	blind	23.6	zeitung::n	newspaper
4.15	blut::n	blood	14.42	nacht::n	night
15.67	blau::a	blue	13.09	neun::num	nine
15.79	stumpf::a	blunt	12.47	norden::n	north
7.57	brett::n	board	4.23	nase::n	nose
10.83	boot::n	boat	24.06	nicht::adv	not
4.11	körper::n	body	14.18	jetzt::adv	now
4.16	knochen::n	bone	10.85	ruder::n	oar
18.61	buch::n	book	14.32	oft::adv	often
6.52	stiefel::n	boot	5.79	öl::n	oil
20.24	bogen[waffe]::n	bow	14.15	alt::a	old
2.25	junge::n	boy	13.01	eins::num	one
4.203	gehirn::n	brain	13.33	nur::adv	only
5.51	brot::n	bread	17.54	oder::conj	or
4.41	brust::n	breast	24.11	anderer::a	other
10.74	brücke::n	bridge	3.596	eule::n	owl
15.57	hell::a	bright	16.31	schmerz::n	pain
9.38	besen::n	broom	9.88	farbe::n	paint
2.44	bruder::n	brother	9.89	malen::v	to paint
3.21	bulle::n	bull	2.37	eltern::n	parents
5.89	butter::n	butter	13.23	teil::n	part
3.92	schmetterling::n	butterfly	3.16	weide::n	pasture
6.62	knopf::n	button	10.72	pfad::n	path
1.322	ruhe::n	calm	19.21	leute::n	people
3.62	katze::n	cat	13.231	stück::n	piece
1.28	höhle::n	cave	3.35	schwein::n	pig
7.43	stuhl::n	chair	7.421	kissen::n	pillow
11.89	billig::a	cheap	8.64	kiefer[baum]::n	pine

4.208	wange::n	cheek	8.691	rohr::n	pipe
3.55	huhn::n	chicken	12.11	ort::n	place
4.209	kinn::n	chin	12.352	spitz::a	pointed
22.131	kirche::n	church	11.52	arm::a	poor
12.82	kreis::n	circle	5.26	topf::n	pot
4.345	klaue::n	claw	16.79	loben::v	praise
9.73	lehm::n	clay	11.87	preis::n	price
15.87	sauber::a	clean	1.75	regen::n	rain
16.84	klug::a	clever	1.59	regenbogen::n	rainbow
6.21	stoff::n	cloth	5.122	roh::a	raw
1.73	wolke::n	cloud	15.66	rot::a	red
15.86	kalt::a	cold	11.51	reich::a	rich
6.45	kragen::n	collar	6.73	ring::n	ring
6.91	kamm::n	comb	5.123	reif::a	ripe
12.76	ecke::n	corner	10.71	straße::n	road
19.11	land::n	country	7.51	dach::n	roof
3.23	kuh::n	cow	8.54	wurzel::n	root
12.77	kreuz::n	cross	12.81	rund::a	round
3.593	krähe::n	crow	5.81	salz::n	salt
5.35	tasse::n	cup	1.215	sand::n	sand
15.63	dunkel::a	dark	14.68	samstag::n	saturday
2.42	tochter::n	daughter	1.32	meer::n	sea
4.95	taub::a	deaf	13.36	zweiter::a	second
15.88	schmutzig::a	dirty	8.311	saat::n	seed
5.31	speise::n	dish	13.07	sieben::num	seven
3.61	hund::n	dog	15.78	scharf::a	sharp
12.03	hinab::adv	down	3.25	schaf::n	sheep
5.9	trinken::v	drink	7.47	regal::n	shelf
15.84	trocken::a	dry	6.44	hemd::n	shirt
3.57	ente::n	duck	6.51	schuh::n	shoe
1.213	staub::n	dust	1.27	ufer::n	shore
3.584	adler::n	eagle	12.59	kurz::a	short
4.22	ohr::n	ear	4.3	schulter::n	shoulder
12.45	osten::n	east	8.24	schaufel::n	shovel
12.353	kante::n	edge	12.36	seite::n	side
5.97	ei::n	egg	9.65	silber::n	silver
13.08	acht::num	eight	23.51	versinken::v	sink
4.32	ellenbogen::n	elbow	2.45	schwester::n	sister
13.101	elf::num	eleven	13.06	sechs::num	six
13.22	leer::a	empty	1.51	himmel::n	sky
19.52	feind::n	enemy	1.83	rauch::n	smoke
14.46	abend::n	evening	8.69	rauchen::v	to smoke
11.88	teuer::a	expensive	15.77	glatt::a	smooth
4.21	auge::n	eye	3.85	schlange[tier]::n	snake
4.204	gesicht::n	face	1.76	schnee::n	snow
2.82	familie::n	family	15.75	weich::a	soft
12.44	fern::a	far	14.33	manchmal::adv	sometimes
2.35	vater::n	father	2.41	sohn::n	son

16.76	schuld::n	fault	23.64	lied::n	song
4.393	feder::n	feather	14.331	bald::adv	soon
8.16	zaun::n	fence	5.64	suppe::n	soup
4.841	fieler::n	fever	15.38	sauer::a	sour
4.34	finger::n	finger	12.48	süden::n	south
4.344	fingernagel::n	ingernail	8.23	spaten::n	spade
8.65	tanne::n	fir	18.222	rede::n	speech
1.81	feuer::n	fire	3.818	spinne::n	spider
13.34	erster::a	first	5.37	löffel::n	spoon
3.65	fisch::n	fish	3.869	eichhörnchen::n	squirrel
13.05	fünf::num	five	1.54	stern::n	star
12.71	flach::a	flat	11.54	geizig::a	stingy
7.26	fußboden::n	floor	4.46	magen::n	stomach
8.57	blume::n	flower	7.32	ofen::n	stove
3.83	fliege::n	fly	17.22	dumm::a	stupid
10.37	fliegen::v	to fly	14.76	sommer::n	summer
1.324	schaum::n	foam	1.52	sonne::n	sun
1.74	nebel::n	fog	14.62	sonntag::n	sunday
5.12	nahrung::n	food	1.38	sumpf::n	swamp
4.37	fuß::n	foot	15.35	süß::a	sweet
4.205	stirn::n	forehead	7.44	tisch::n	table
5.39	gabel::n	fork	4.18	schwanz::n	tail
13.04	vier::num	four	23.9	tee::n	tea
3.74	fuchs::n	fox	17.27	lehrer::n	teacher
14.67	freitag::n	friday	16.38	träne::n	tear
19.51	freund::n	friend	9.28	zerreißen::v	to tear
13.21	voll::a	full	13.1	zehn::num	ten
6.28	fell::n	fur	24.08	das::prn	that
2.26	mädchen::n	girl	24.10	dort::adv	there
9.74	glas::n	glass	2.96	sie::prn	they
9.56	kleben::v	glue	12.63	dick[gegenstand]::a	thick
22.12	gott::n	god	4.351	oberschenkel::n	thigh
9.64	gold::n	gold	12.65	dünn::a	thin
16.71	gut::a	good	11.18	sache::n	thing
3.56	gans::n	goose	13.42	dritter::a	third
8.42	korn::n	grain	24.07	dies::prn	this
2.46	großvater::n	grandfather	6.38	faden::n	thread
2.47	großmutter::n	grandmother	13.03	drei::num	three
8.51	gras::n	grass	4.29	kehle::n	throat
4.79	grab::n	grave	24.05	durch::prp	through
15.68	grün::a	green	1.56	donner::n	thunder
16.32	kummer::n	grief	14.66	donnerstag::n	thursday
20.471	bewachen::v	guard	14.11	zeit::n	time
19.56	gast::n	guest	14.47	heute::adv	today
20.28	gewehr::n	gun	4.38	zeh::n	toe
4.14	haar::n	hair	14.48	morgen::adv	tomorrow
13.24	hälfte::n	half	4.26	zunge::n	tongue
4.33	hand::n	hand	4.27	zahn::n	tooth

15.74	hart::a	hard	6.82	handtuch::n	towel
3.863	hase::n	hare	19.15	stadt::n	town
8.52	heu::n	hay	20.64	falle::n	trap
2.931	er::prn	he	8.6	baum::n	tree
4.2	kopf::n	head	6.48	hose::n	trousers
4.44	herz::n	heart	16.66	wahr::a	TRUE
15.81	schwer::a	heavy	14.64	dienstag::n	tuesday
4.372	ferse::n	heel	13.102	zwölf::num	twelve
24.09	hier::adv	here	13.104	zwanzig::num	twenty
12.85	loch::n	hole	13.02	zwei::num	two
5.84	honig::n	honey	2.51	onkel::n	uncle
12.75	haken::n	hook	12.07	unter::prp	under
4.17	horn::n	horn	12.08	hinauf::adv	up
3.41	pferd::n	horse	19.16	dorf::n	village
15.85	heiß::a	hot	18.11	stimme::n	voice
7.12	haus::n	house	15.851	warm::a	warm
2.31	ehemann::n	husband	1.31	wasser::n	water
2.91	ich::prn	i	1.35	welle::n	wave
1.77	eis::n	ice	2.94	wir::prn	we
12.041	vor::prp	in front of	4.82	schwach::a	weak
9.67	eisen::n	iron	1.78	wetter::n	weather
1.25	insel::n	island	14.65	mittwoch::n	wednesday
4.207	kiefer[anatomie]::n	jaw	14.61	woche::n	week
5.27	kessel::n	kettle	12.46	westen::n	west
19.32	könig::n	king	15.83	nass::a	wet
4.36	knie::n	knee	15.64	weiß::a	white
9.192	knoten::n	knot	12.61	breit::a	wide
7.37	leiter::n	ladder	2.32	ehefrau::n	wife
1.33	see::n	lake	1.72	wind::n	wind
1.21	festland::n	land	7.25	fenster::n	window
18.24	sprache::n	language	4.392	flügel::n	wing
13.35	letzter::a	last	14.74	winter::n	winter
4.92	faul::a	lazy	3.71	wolf::n	wolf
9.68	leiten::v	lead	2.22	frau::n	woman
8.56	blatt::n	leaf	1.43	holz::n	wood
6.29	leder::n	leather	6.22	wolle::n	wool
12.42	linker::a	left	18.26	wort::n	word
4.35	bein::n	leg	9.12	arbeit::n	work
23.42	brief::n	letter	1.1	welt::n	world
4.741	leben::n	life	3.84	wurm::n	worm
1.61	licht::n	light	14.73	jahr::n	year
1.86	anzünden::v	to light	15.69	gelb::a	yellow
12.84	linie::n	line	14.49	gestern::adv	yesterday
4.25	lippe::n	lip	14.14	jung::a	young

Appendix 2

Complete dataset for the languages, Georgian, Northern Kurdish, Turkish, and Western Farsi. The dataset is too large to fit here and the reader is referred to the csv-file Appendix2. Some columns have been removed for ease of reading, but if the full xlsx is required, please contact me.