Nominal and verbal affixation in the Caucasus

A morphological and phonological approach

Larsson, Filip

2024

Document Version:
Publisher's PDF, also known as Version of record

Nominal and verbal affixation in the Caucasus

A morphological and phonological approach

FILIP LARSSON
CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

Joint Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-89874-51-0

Nominal and verbal affixation in the Caucasus

Nominal and verbal affixation in the Caucasus

A morphological and phonological approach

Filip Larsson

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Joint Faculties of Humanities and Theology at Lund University to be publicly defended on the 28th of November at 10:15 in LUX:C121.

Faculty opponent
Prof. Diana Forker

Organization: LUND UNIVERSITY, Centre for Languages and Literature, Box 201, SE-221 00 LUND Sweden

Document name: Doctoral thesis **Date of issue: 2024-11-28**

Author(s): Filip Larsson **Sponsoring organization: -**

Title and subtitle:

Nominal and verbal affixation in the Caucasus: A morphological and phonological approach

Abstract:

This thesis aims to investigate the interaction between complex morphology and complex phonology in the languages of the Caucasus. The Caucasus is well-known for containing languages with exceptionally large case systems and complex polysynthetic verbal morphology, which is paired with some of the largest consonant inventories in the world outside Africa. The study focuses specifically on nominal and verbal affixation, the morphological process of adding bound morphemes to lexical stems, as the languages of the Caucasus present some of the most intricate affixation patterns in the world.

The underlying hypothesis of the thesis is that larger consonant inventories enable more complex morphology, which was operationalised as the number of grammatical functions expressed by affixation. A data set of more than 11,000 affixes was compiled, which enabled a comparison of the vast variety of grammatical functions expressed by affixation and the related phonological forms in 56 languages from the five language families of the Caucasus, i.e. Kartvelian, Nakh-Dagestanian, Northwest Caucasian, Indo-European and Turkic. The results indicated a significant positive correlation between the number of grammatical functions expressed by affixation and the size of a language's consonant phoneme inventory, which was also true for the combined inventories of both consonant and vowel phonemes.

It has previously been proposed that the three endemic language families of the Caucasus, i.e. Kartvelian, Nakh-Dagestanian and Northwest Caucasian, belong to a common linguistic area, known as the Caucasian Sprachbund. The thesis also intended to test whether the nominal and verbal affixation inventories could support the notion of a morphological Caucasian Sprachbund, and the results could not support such a morphological sprachbund.

A second hypothesis postulated that there are systematic phonological differences between affixes and lexical stems, which motivated a second data set of more than 21,500 lexical items from 52 of the 56 languages of the affixal data set. When the affixal data set and the lexical data set were compared, a significant difference could be observed between phonological distributions of combinations of place and manner of articulation. The results also demonstrated that voiceless consonants are significantly more common in lexical stems than in affixes. The phonological results also indicated that there are significant differences for certain combinations of place, manner and voicing, where particularly the various ejective consonants of the Caucasus all presented significantly different distributions in the affixal and lexical data sets. This suggests that the large inventories of ejectives in the Caucasus potentially facilitate the distinction between affixes and lexical stems in these languages.

Key words:

morphology, phonology, affixation, Caucasus, Kartvelian, Nakh-Dagestanian, Northwest Caucasian, Turkic, grammatical functions

Classification system and/or index terms (if any): - Supplementary bibliographical information: -

Language: English **ISSN and key title**: -

ISBN:

978-91-89874-51-0 (print)

978-91-89874-52-7 (pdf)

Recipient's notes: - **Number of pages**: 376 **Price:** Free

Signature Date 2024-10-10

Nominal and verbal affixation in the Caucasus

A morphological and phonological approach

Filip Larsson

LUND
UNIVERSITY

Coverphoto by Filip Larsson

Joint Faculties of Humanities and Theology
Centre for Languages and Literature

ISBN (print): 978-91-89874-51-0
ISBN (pdf): 978-91-89874-52-7

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

May the mountains never lose their tongues!

Table of Contents

Acknowledgements

There are many people who I would like to thank for making this thesis possible, as I am deeply grateful for their help and support. This linguistic journey through the mighty and daunting Caucasus would never have begun if Gerd Carling had not supported my initial idea in 2018. I would like to thank you from the bottom of my heart, as your support and encouragement have given me so much over the years we have worked together. Without you I would likely not have given academia another go, so I am eternally grateful for leading me back.

I would also like to sincerely thank Karina Vamling and Revaz Tchantouria for generously guiding me into the field of Caucasiology and Kartvelology, both in Sweden and in Georgia. I would particularly like to thank Karina for sharing her indispensable knowledge of the Caucasus and the Russian-language literature, for which I am deeply grateful.

Life does not always go according to plan, and when it felt like the Four Horsemen of the Apocalypse had descended upon the world, I had to revise my original plan of doing fieldwork in the North Caucasus. One person in particular turned this issue from a lethal blow to a minor wound, as the data collection would have been extremely difficult during the pandemic without the generosity of Harald Hammarström. I am indebted to you, as my thesis would not be the same without your help.

The pandemic came and went, but as change is the only constant, I found myself with a new main supervisor. I am grateful that Arthur Holmer accepted this role, and although we do not always agree, our discussions and your feedback have been of utmost importance to the final phase of this process.

I want to thank Niklas Erben Johansson and Sandra Cronhamn for all the transformative discussions we have had over a span of more than ten years, and for offering advice and guidance along the journey of this thesis.

I would also like to thank Erich Round, Chundra Cathcart, Manana Kock Kobaidze, Johan Frid and Henriette Arndt as you all have given me important feedback in my writing process.

I would also like to thank Merab Chukhua for welcoming me to the Circassian Cultural Centre in Tbilisi, and Larisa Tuptsokova for providing me with valuable insights into the Adyghe language. I am also grateful that I was welcomed by Sirarpi Karapetyan and Susanna Grigoryan in Yerevan.

Finally, I am grateful for all the support and inspiration I have received from my family, my colleagues Kajsa Gullberg, Felix Ahlner, Mechtild Tronnier, Victor Bogren Svensson, Eline Visser, Karolin Obert, Tuğba Lulaci, Ante Petrović, and my friends Josefin Waldenström, Teri Schamp-Bjerede, Lisa Samuelsson, Veronika Enekvist, Anna Lindberg, Markus Wendén and many more.

Abstract

This thesis aims to investigate the interaction between complex morphology and complex phonology in the languages of the Caucasus. The Caucasus is well-known for containing languages with exceptionally large case systems and complex polysynthetic verbal morphology, which is paired with some of the largest consonant inventories in the world outside Africa. The study focuses specifically on nominal and verbal affixation, the morphological process of adding bound morphemes to lexical stems, as the languages of the Caucasus present some of the most intricate affixation patterns in the world.

The underlying hypothesis of the thesis is that larger consonant inventories enable more complex morphology, which was operationalised as the number of grammatical functions expressed by affixation. A data set of more than 11,000 affixes was compiled, which enabled a comparison of the vast variety of grammatical functions expressed by affixation and the related phonological forms in 56 languages from the five language families of the Caucasus, i.e. Kartvelian, Nakh-Dagestanian, Northwest Caucasian, Indo-European and Turkic. The results indicated a significant positive correlation between the number of grammatical functions expressed by affixation and the size of a language's consonant phoneme inventory, which was also true for the combined inventories of both consonant and vowel phonemes.

It has previously been proposed that the three endemic language families of the Caucasus, i.e. Kartvelian, Nakh-Dagestanian and Northwest Caucasian, belong to a common linguistic area, known as the Caucasian Sprachbund. The thesis also intended to test whether the nominal and verbal affixation inventories could support the notion of a morphological Caucasian Sprachbund, and the results could not support such a morphological sprachbund.

A second hypothesis postulated that there are systematic phonological differences between affixes and lexical stems, which motivated a second data set of more than 21,500 lexical items from 52 of the 56 languages of the affixal data set. When the affixal data set and the lexical data set were compared, a significant difference could be observed between phonological distributions of combinations of place and manner of articulation. The results also demonstrated that voiceless consonants are significantly more common in lexical stems than in affixes. The phonological results also indicated that there are significant differences for certain combinations of place, manner and voicing, where particularly the various ejective consonants of the Caucasus all presented significantly different distributions in the affixal and lexical data sets. This suggests that the large inventories of ejectives in the Caucasus potentially facilitate the distinction between affixes and lexical stems in these languages.

Abbreviations

1	1st person
2	2nd person
3	3rd person
1PL	1st person plural
1SG	1st person singular
2PL	2nd person plural
2SG	2nd person singular
3PL	3rd person plural
3SG	3rd person singular
A	Agent
ABESS	Abessive
ABL	Ablative
ABS	Absolutive
ACC	Accusative
AD	Ad(essive)
ADD	Additive
ADDR	Addressive
ADJ	Adjective
ADV	Adverb(ial)
AFF	Affirmative
AFFT	Affective
ALL	Allative
AN	Agentive noun
ANT	Anterior
ANTE	Orientation Ante-
ANTIC	Anti-causative
ANTIP	Antipassive
AOR	Aorist
APPL	Applicative
APPR	Apprehensive
APPROB	Approbative
APUD	Orientation Apud-
ASS	Assumed/Assumptive
ATTR	Attributive
AUGM	Augment
AUX	Auxiliary
B	Nakh noun class B
BEN	Benefactive
C	Consonant
CAUS	Causative
CAUSL	Causal
CIRC	Circumlative
CM	Noun class marker
COM	Comitative
COMMIS	Commiserative
COMP	Complementiser
COMPR	Comparative
CONC	Concessive
COND	Conditional
CONJ	Conjunction
CONT	Orientation Cont-
CONTENT	Contentive
COP	Copula
CRD	Coordinative
CUM	Orientation Cum-
CVB	Converb
D	Nakh noun class D
DAT	Dative
DE	Delative
DEB	Debitive
DEF	Definite
DELIB	Deliberative
DEM	Demonstrative
DESID	Desiderative
DETR	Detransitive
DIR	Direction/Directional
DIST	Distal
DUBIT	Dubitative
DUR	Durative
DYN	Dynamic
ELA	Elative
EMPH	Emphatic
EQTEMP	Equitemporal
EQU	Equative
ERG	Ergative
ESS	Essive
EVID	Evidential
EVT	Eventual
EXC	Excessive
EXCL	Exclusive
EXH	Exhaustive
EXIST	Existential

EZ	Ezafe
F	Feminine
FIN	Finite
FORL	Forlative
FUT	Future
GEN	Genitive
GNOM	Generic/Gnomic
GRAD	Gradual/Graduative
H	Human
HAB	Habitual
HAND	Orientation ‘hand’
HOME	Orientation ‘home’
HORT	Hortative
HPL	Human plural
HSAY	Hearsay Evidential
I	Noun class I
II	Noun class II
III	Noun class III
IE	Indo-European
ILOC	‘Inanimate’ location
IMANT	Immediate Anterior
IMP	Imperative
IMPOSTR	Immediate Posterior
IN	Orientation In-
INAN	Inanimate
INCEP	Inceptive
INCH	Inchoative
INCL	Inclusive
IND	Indicative
INDEF	Indefinite
INF	Infinite
INFER	Inferential
INS	Instrumental
INT	Intentional
INTER	Orientation Inter-
INTERJ	Interjection
IO	Indirect object
IOV	Indirect objective version
IPF	Imperfect
IPFV	Imperfective
IRR	Irrealis
IRR.COND	Irrealis Conditional
ITER	Iterative
ITR	Intransitive
IV	Noun class IV
J	Nakh noun class J
Kart.	Kartvelian
LAT	Lative
LC	Local case
LOC	Locative
LV	Light verb
M	Masculine
MAL	Malefactive
MIR	Mirative
MNR	Manner
MOD	Modal
MOM	Momentane
MP	Medio-passive
MSD	Masdar
NC	Noun class
ND	Nakh-Dagestanian
NEC	Necessitative
NEG	Negative
NFIN	Non-finite
NH	Non-human
NOM	Nominative
NULL	Null/Empty
NUM	Numeral
NV	Neutral version
NWC	Northwest Caucasian
NZ	Nominaliser
O	Direct Object
OBL	Oblique
OPT	Optative
OR	Orientation
OV	Objective version
P	Patient
PART	Partitive
PASS	Passive
PERMIS	Permissive
PFV	Perfective
PL	Plural
PLUP	Pluperfect
POSS	Possessive
POST	Orientation Post-

POSTR	Posterior
POT	Potential
PRET	Preterite
PRF	Perfect
PROB	Probabilitive
PROG	Progressive
PROH	Prohibitive
PROSP	Prospective
PROX	Proximal
PRS	Present
PrV	Preradical vowel
PST	Past
PTCP	Participle
PTCL	Particle
PURP	Purposive
PV	Preverb
Q	Interrogative
QUOT	Quotative
RE	Retrolative
REC	Reciprocal
REDUP	Reduplication
REFL	Reflexive
REP	Repetitive
RES	Resultative
S	Subject
SEM	Semelfactive
SEQ	Sequential
SG	Singular
SIM	Simultaneous
SIMIL	Similative
SM	Series marker
STAT	Stative
SUB	Orientation Sub-
SUBJ	Subjective
SUBST	Substitutive
SUPER	Orientation Super-
SUS	Suslative
SV	Subjective version
TEMP	Temporal
TERM	Terminative
TERMIN	Orientation ‘end’
TR	Transitive
TRANS	Translative
TS	Thematic suffix
Turk.	Turkic
UNINT	Unintentional
UWPST	Unwitnessed past
V	Noun class V
VN	Verbal noun
VOC	Vocative
VPL	Verbal pluraliser
WPST	Witnessed past

1. Introduction

The Caucasus is a linguistically diverse region on the border between Europe and Asia, as more than 60 languages from five language families are spoken in these mountainous areas ranging from the Black Sea to the Caspian Sea. The Caucasus contains some of the world's most elaborate affixation patterns (Daniel & Ganenkov 2009: 671), which is paired with some of the world's largest consonant inventories outside Africa (Beguš 2020: 699). The languages of the Caucasus therefore constitute a highly interesting case to examine the possible correlations between morphology and consonant inventories from a phonological perspective. The Caucasus contains three endemic language families, i.e. the Kartvelian, Nakh-Dagestanian and Northwest Caucasian language families, all of which have rich morphology but clear differences in affixation patterns. The aim of this thesis is twofold, as it examines both the phonology and morphology of affixation in the languages of the Caucasus.

I begin this thesis by introducing the research questions and thereafter I give a brief theoretical background in chapter 2 to introduce the most important concepts, how I define affixation, and why it is relevant for our understanding of the relationship between grammatical processes and the lexicon. I conclude chapter 2 by presenting my hypotheses. In chapter 3, I introduce the Caucasus and its linguistic diversity. The methodology and the data sets are thereafter presented in chapter 4. In chapters 5 and 6, I describe and exemplify in detail the various grammatical functions expressed by affixation in the languages of the Caucasus, in order to define and demonstrate the basis of the subsequent morphological results. Morphological and phonological results are presented in chapter 7, with the purpose of answering the research questions presented in the introduction. The final chapter concludes the discussions in the Results chapter and the wider conclusions drawn from this thesis are discussed.

1.1. Research questions

Considering the presence of complex affixation patterns and rich phoneme inventories of the languages of the Caucasus, the subsequent question is whether these similarities are merely superficial and coincidental or whether there are structural and systematic similarities. Since the notion of a *Caucasian Sprachbund* has been thoroughly debated and fairly recently reiterated by Chirikba (2008), it is relevant to quantitatively test how many morphological similarities are shared between the three endemic language families. The central morphological research questions are therefore:

(1) Do the affix inventories of the languages of the three endemic Caucasian language families display sufficient morphological similarities to support the notion of a Caucasian Sprachbund, as claimed by e.g. Chirikba (2008)?

(2) Do the affixation patterns of the Turkic and Indo-European languages spoken in the Caucasus exhibit morphological similarities with the three endemic language families of the Caucasus?

(3) Are there hierarchical patterns in the distribution of grammatical functions expressed by affixation across the five language families of the Caucasus?

The phonological part of this study aims to investigate both the interaction between affixation and phonology and the effect of phoneme inventory size on affixation inventories. I therefore intend to answer the following phonological research questions:

(4) Are there significant phonological differences between affixes and lexicon in the languages of the Caucasus, based on the articulatory variables of place, manner and voicing?

(5) Are there significant differences in the distribution of certain places and manners of articulation between affixes and lexicon in the languages of the Caucasus, as these languages are well-known for their large inventories of places and manners of articulation?

(6) Is there a significant correlation between phoneme inventory size, i.e. the number of consonant and vowel phonemes, and the number of grammatical functions expressed by affixation in the languages of the Caucasus?

2. Background

2.1. Morphology and the lexicon

The most salient component of language is most likely the word, which in itself is a concept that is inherently difficult to define. The notion of the grammatical word in opposition to the phonological word presents the fundamental premise of this thesis, as the grammatical word in non-isolating languages generally 'must include at least one inflectional morpheme' (Aikhenvald 2007: 2). The grammatical word therefore consists of a lexical stem and one or more additional morphemes, where the latter typically belong to a closed class of formatives, which are morphological entities at the heart of inflectional morphology (Bickel & Nichols 2007: 172). In this framework, formatives are by definition different from words as they cannot govern or be governed by other words, require agreement nor can they head phrases according to the definition of Bickel and Nichols (2007: 173). A precise demarcation between words and formatives is at best theoretical, which will become apparent later in this thesis. Formatives can either occur on their own as phonologically unbound units, e.g. as particles, or as phonologically bound units, i.e. affixes (Bickel & Nichols 2007). The phonologically bound units are contained within the category of phonological affixes, which includes bound formatives, lexical affixes and incorporated stems (Bickel & Nichols 2007: 192), and the grammatical affixes, which exclusively refers to bound formatives (Bickel & Nichols 2007: 174).

Stump postulates two defining distinctions of morphological theory, i.e. lexical vs. inferential and incremental vs. realisational (Stump 2001). Lexical theories treat inflectional morphemes as any other part of the lexicon by inserting them into the hierarchical structure of the syntax, while inferential theories assume that morphology is a process which is independent from syntax (Stump 2019: 286). The incremental theories analyse morphology as the sum of the 'inflectional exponents' associated with a particular word form, whereas realisational theories assume that the morphosyntactic properties of a word form are defined by its position in a paradigm (Stump 2019: 286).

In non-isolating languages, the lexical stem and one or more affixes thereby constitute the grammatical word, which suggests that if words and formatives are intrinsically different it is reasonable to make the assumption that inflectional morphology and the lexicon operate on separate yet parallel levels, cf. the inferential

approach above. The morphological distinction between stems and formatives should also be compared to the notions of phonological contrastiveness and distinctiveness, where contrastiveness is 'characterized by the dominant faithfulness constraints at the stem level, the innermost layer of the lexicon' (Kiparsky 2018: 63). Distinctive segments on the other hand 'enhance lexical feature contrast by redundant features' which can 'appear on the contrastive segments themselves or – what is more relevant here – on neighbouring segments' (Kiparsky 2018: 63).

Stump's *Paradigm Function Morphology* (PFM) is in its many forms explicitly inferential-realisational (Stump 2019: 286), and is as such categorised by Blevins et al. (2019: 267) as a *realisational Word and Paradigm* (WP) approach to morphology, which is fundamentally similar to the approach of this thesis. My approach is ostensibly inferential, as it assumes that the lexicon and morphology operate on parallel levels and that they likely differ phonologically, as affixes are not listed in the lexicon. Stump discourages lexical and incremental theories, while specifically arguing against an incremental approach due to the problems of 'extended exponence' and the tendency to 'underdetermine' morphosyntactic properties (Stump 2001: 3-9). Stump defines 'extended exponence' as a given property being expressed by more than one morphological marking, which is explicitly 'precluded' in incremental theories such as *Articulated Morphology* (Steele 1995: 280). This claim is highly problematic, as it would fundamentally question the presence of circumfixes and transfixes, cf. section 2.3.1.

The second problem of underdetermining is primarily exemplified by TAM form syncretisms and partial system levelling (Stump 2001: 8). These arguments support the realisational importance of analysing inflectional morphology from a paradigmatic perspective, as Stump gives numerous examples of how particular morphological forms can primarily be described by its position in relation to other forms in the word-specific paradigm. The approach of this thesis is therefore implicitly *inferential-realisational*, thereby not following any form of the PFM or other related frameworks, while Stump's concept of *property* is largely synonymous with the *grammatical functions* discussed below.

2.2. Grammatical functions

The concept of *grammatical functions* is best understood in relation to *meaning*, as grammatical functions can only have meaning if combined with lexical stems, i.e. the grammatical word. The distinction between meaning and grammatical function goes back to Sapir's distinction between 'material content', i.e. meaning, and 'relational concepts', i.e. grammatical function, where he defines relational concepts as being 'normally expressed by affixing non-radical elements to radical elements' (Sapir 1921: 106-107). Bybee (1985) makes a similar distinction between 'lexical meaning' and 'grammatical meaning' (Bybee 1985: 7). Grammatical functions are

not only expressed by formatives and affixes, as grammatical functions can be expressed by syntactic words, i.e. function words. This is easily exemplified by English, where *the* and *–ing* both lack lexical meaning and carry grammatical functions, while the former is a syntactic word and the latter is a suffix. As already observed by Sapir (1921: 107), there is no definite boundary between grammatical functions and meaning, as a universal distinction between these concepts would likely become exceedingly abstract. Assumptions can be made about the nature of grammatical functions, as they typically form a closed class in most languages. I have principally followed Bybee's approach to grammatical function as expressing a certain value within a grammatical category (Bybee 1985: 28). The first step is therefore to define a set of relevant grammatical categories as variables that can be assigned categorical values, i.e. specific functions. Although all grammatical functions must be given *labels*, Bybee stresses the importance of reanalysing morphemes based on their grammatical function and not the language-specific label used in previous descriptions (Bybee 1985: 28).

The typical issue with language-specific descriptions is that they contain *multifunctional labels*, where two or more underlying functions are subsumed under one grammatical label. This is a well-known phenomenon in linguistics, e.g. case syncretism, cf. section 2.3.1.1. Comparative morphology can therefore be analysed as the attempt to identify the enormous yet finite amount of grammatical functions that are distinguished in the world's languages. If a grammatical function is expressed by inflectional morphology in one language, then it means that it is potentially a grammatical function in any other language, either as a bound morpheme or a syntactic word.

The difference between grammatical labels and grammatical functions can be exemplified by comparing affixes indicating core case functions in Georgian, Lezgian and Kabardian.

Table 2.1: Examples of affixes expressing the core case functions absolutive/nominative, dative, ergative and genitive in Georgian, Lak and Kabardian.

Case function	Georgian (Hewitt 1995)	Lak (Friedman 2020)	Kabardian (Kumakhov & Vamling 2009)
Absolutive/Nominative	*-i*	*-Ø*	*-r*
Dative	*-s(a)*	*-n*	*-m*
Ergative	*-ma*	*-l*	*-m*
Genitive	*-is(a)*	*-l*	*-m*

Table 2.1 demonstrates how languages express grammatical functions differently, which in itself is unproblematic. The grammatical labels used for these languages present a different picture however, as the suffix –l in Lak is labelled as simply 'genitive' by Friedman (2020: 210-211). The Kabardian suffix –m is either divided into the two categories 'ergative' for the ergative function and 'oblique ergative' for the dative, 'the owner in possessive constructions' and some other functions

(Kumakhov & Vamling 2009: 22-23), or simply labelled as 'oblique' regardless of function (Arkadiev & Lander 2020: 393). The case labels in Lak and both of the Kabardian case label systems exhibit multifunctional labelling. This shows why it is notoriously difficult to compare languages based on grammatical labels alone, since labels tend to mean different things depending on which language or language family it describes.

Although Haspelmath argues that cases are 'language-particular entities', where case labels are 'valid only for particular languages' (Haspelmath 2009: 510), some of these issues can be avoided by making the distinction between grammatical labels and grammatical functions as proposed by Bybee (1985). Grammatical functions are not found in some sort of generative notion of 'deep-structure' as proposed in Fillmore's Case Grammar (Fillmore 1968) however, as he is rather describing semantic roles (Haspelmath 2009). Grammatical functions should instead be understood as being present in the surface morphosyntax, realised either explicitly or implicitly by morphology, function words and word order.

Optimally, a linguistic meta-language should therefore be able to distinguish as many grammatical functions as possible in order to avoid multifunctional labels. If one perceived grammatical function is consistently realised by two different surface forms in one language, then it is important to consider the possibility that there is a functional distinction that has not yet been identified. This demonstrates that grammatical functions cannot be defined by labels or form alone, as allomorphs and syncretism are inherent parts of human language. The combination of function, form and position must therefore all be taken into consideration when analysing and comparing grammatical categories.

2.3. Affixes and clitics

2.3.1. Affixes

Affixes are by definition 'syntactically and phonologically dependent units' (Bickel & Nichols 2007: 174), usually bearing an abstract meaning or function and they cannot occur on their own as independent words (Haspelmath & Sims 2010: 19). Affixes are typically divided into various categories based on either form or function, where the most important distinction for this thesis is the functional division between *inflectional* and *derivational affixes* (Aikhenvald 2007: 35) as this study intends to primarily investigate inflectional affixes, following Sapir's distinction between 'material content', defined as basic and derivational concepts, and 'relational concepts', i.e. grammatical functions (Sapir 1921: 106-107), cf. section 2.4. for further discussion.

Affixes can be used for various functions and they are not evenly distributed among the languages of the world. Isolating languages practically lack affixation altogether

while non-isolating languages typically have a fairly restricted inventory of affixes, but this is not the case for most languages in the Caucasus. Affixes tend to exclusively attach to their associated word class, e.g. nominal affixes only attach to nominals, etc. (Bickel & Nichols 2007: 175), which makes it important to distinguish nominal affixes from verbal affixes. Affixes are formally divided into suffixes that follow the stem, prefixes that precede the stem, infixes that are entered into the stem and circumfixes that occur both before and after the stem (Haspelmath & Sims 2010: 20). There is also a fifth highly unusual category of transfixes, which is *de facto* a combination of an infix with a prefix, suffix or even a circumfix (Aikhenvald 2007: 45).

2.3.1.1. Affixal syncretism

Affixal syncretism is a frequent linguistic phenomenon in which one morpheme has two or more grammatical functions within an inflectional paradigm (Zwicky 1985: 373; Stump 2001: 212) and it complicates morphological analyses. In some languages, e.g. Latin, the presence of affixal syncretism becomes apparent primarily while comparing the declension patterns (Bennett 1908; Baldi 2002) (cf. table 2.2). It would therefore not be possible to assert the distinction between dative, genitive and locative in Latin if it only had the first declension pattern, while the second and third declensions do not differentiate the nominative and accusative cases. The conventional approach regarding affixal syncretism has largely been to categorise all phonologically and syntactically identical affixes as instances of the same grammatical category, which becomes rather problematic in many contexts. I therefore argue that the best solution to these issues is to apply a functional approach by considering affix syncretism when separate grammatical functions are realised with the same affix.

Table 2.2: Examples of affixal syncretism of the vestigial locative case in some irregular Latin declension patterns.

Case	**Feminine (1st declension)**	**Neuter (2nd declension)**	**Neuter (3rd declension)**	**Feminine (4th/2nd declension)**
Nominative	*mīliti-a* ‘war’	*bell-um* ‘war’	*rū-s* ‘countryside’	*dom-us* ‘house’
Accusative	*mīliti-am*	*bell-um*	*rū-s*	*dom-um*
Genitive	*mīliti-ae*	*bell-ī*	*rū-ris*	*dom-ūs*
Dative	*mīliti-ae*	*bell-ō*	*rū-rī*	*dom-uī*
Locative	*mīliti-ae*	*bell-ī*	*rū-re*	*dom-ī*

2.3.2. Clitics

Clitics or affixed particles are bound morphemes that are similar to affixes as they also tend to convey grammatical and syntactical functions but they are conventionally differentiated from affixes as they do not attach to only one word class and are thereby technically *categorially unrestricted bound formatives* (Bickel & Nichols 2007: 174). Clitics also differ from affixes in that they can often attach to whole phrases and not just stems, where the English genitive *=s* is a well-known example as it typically attaches to the last element of the NP (Bickel & Nichols 2007: 175). A syntactically similar yet grammatically distinct group is *cliticised adpositions*, since they can often occur both as affixes and as unbound words, but they commonly grammaticalise, e.g. into nominal case affixes, which also indicate a close affiliation to conventional affixes (Bickel & Nichols 2007: 174).

2.3.3. Affixes vs. clitics – separating the attached

The categorisation of bound morphemes into affixes and clitics is not an easy task, as it is notoriously difficult to precisely define these two categories. Zwicky and Pullum present some of the most cited criteria that distinguish affixes from clitics (Zwicky & Pullum 1983: 503-504):

A. Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.
B. Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.
C. Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups.
D. Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups.
E. Syntactic rules can affect affixed words, but cannot affect clitic groups.
F. Clitics can attach to material already containing clitics, but affixes cannot.

It is important to bear in mind that these criteria were developed for English, and many of them are noticeably biased towards Indo-European morphology. The first criterion of host selectivity is clearly a defining feature of affixes due to their intimate relationship with their host words (Spencer & Luís 2012: 108), while it is problematic to define the lack of host selectivity as indicative of cliticisation. The Turkic plural suffixes *–lAr* would likely be defined as clitics by this criterion, as they can attach to nouns, possessives, pronouns, adjectives, adverbs and verbs (Göksel & Kerslake 2005), while the exact function differs depending on the context, as it sometimes marks the third person plural. The second criterion of arbitrary gaps is primarily relevant for fusional languages, as it makes little sense for highly regular agglutinative languages. This is also true for the criteria of

‘morphophonological idiosyncrasies’, since it is largely a feature of fusional languages with irregular morphology. The Turkish example given by Spencer and Luís (2012: 109) of the allomorphs of the Turkish genitive suffix *–(n)In* can hardly be categorised as idiosyncratic, since its allomorphy is regular, as it is completely driven by phonotactic restrictions.

The criterion of ‘semantic idiosyncrasy’ runs a risk of becoming circular, as a stable semantic meaning would indicate that a morpheme is a clitic and vice versa, while both affixes and clitics typically convey grammatical functions rather than semantic meaning, cf. section 2.2. Lexical affixes pose an interesting threat to this criterion, as Northwest Caucasian preverbs (cf. section 6.11) can both have a concrete semantic meaning and a grammaticalised spatial function (Arkadiev, Lander & Bagirokova 2024: 883-884), e.g. the preverb *ҫhe-* which has the semantic meaning ‘head’ while also indicating the spatial function ‘on top of’ (Kumaxov 2006: 81).

The syntactic criterion is perhaps the most relevant, as lexical stems and affixes are typically ‘treated as units by syntactic operations’ (Zwicky & Pullum 1983: 506), while clitics do not. Spencer and Luís (2012: 110) refer to this criterion as ‘lexical integrity’, while also claiming that ‘host=clitic combinations’ behave like syntactic words in the same manner as combinations of lexical stems and affixes. This seems to contradict the original definition, while Zwicky and Pullum do not give any actual examples of criterion E. Object clitics in Swedish clearly show this difference, as *jag såg=na igår* (1SG saw=3SG.F.O yesterday) ‘I saw her yesterday’ but *igår såg jag=na* (yesterday saw 1SG=3SG.F.O) ‘yesterday, I saw her’. This indicates that the syntactic criterion is by definition connected to host selection, while both must be considered, as host selection alone will likely yield numerous ‘false positive’ clitics.

The last criterion claims that affixes cannot attach to clitics, which would mean that everything that follows a clitic is also a clitic. The ‘clitic-affix ordering’ entails that the ‘natural relative order’ is for the affix to be placed first and the clitic afterwards (Spencer & Luís 2012: 110). This criterion has two very different outcomes, as it either promotes an analysis where almost anything can be a clitic or the opposite, that chains of bound morphemes ought to be analysed as affixes. The bound copula in Turkish is a good example, as e.g. *hastaydık* ‘we were ill’ would then either be analysed as *hasta=y=dı=k* or *hasta-y-dı-k* (ill-COP-PST-1PL), where the Zwicky-Pullum criteria would be inconclusive, as the former is supported by criteria A and F while the latter is supported by criteria C and E. This indicates that although the Zwicky-Pullum criteria have been influential for the understanding of affixes and clitics, they also prove to be difficult to apply to languages that have more complex morphology than English.

Zwicky (1994) later describes clitics as an ‘umbrella term’, as he argues that it is not a genuine grammatical category, but rather a phenomenon that ‘present[s] “mixed” properties’ (Zwicky 1994: xiii). Spencer and Luís also conclude their introduction to clitics by stating that traditional typology cannot conclusively

identify ‘a universal category of clitic’ (Spencer & Luís 2012: 321), while they describe clitics as a useful descriptive construct ‘that has the distribution of a function word and the phonological properties of an affix’ (Spencer & Luís 2012: 328). The relationship between affixes and clitics is therefore best described as that of a continuum, where true affixes are found at one end of the spectrum and independent particles are found at the other, while clitics are found somewhere in the middle. I have decided to include both affixes and clitics in this thesis, partly because of the phonological approach of this study but also because the lack of a precise distinction between these categories makes any cross-linguistic attempt to systematically exclude clitics from affixes arbitrary at best.

2.4. Inflection vs. derivation

Inflection and derivation are fundamental concepts within morphology, while the task to differentiate the two categories is not trivial. Aikhenvald states that the primary factor that sets derivational affixes apart from inflectional affixes is that derivation is typically a pre-final process which in itself is subject to the often obligatory addition of inflectional affixes (Aikhenvald 2007: 36). Haspelmath & Sims define inflectional morphology as ‘the relationship between word-forms of a lexeme’ and derivational morphology as ‘the relationship between lexemes of a word family’ (Haspelmath & Sims 2010: 18). The concept of *word family* lies at the heart of the difficulty of separating inflection from derivation, as it is not self-evident which morphological forms belong to the lexeme or to the word family. This question can either be approached as a dichotomy, where words can be ‘neatly divided into two disjoint classes’, or as a continuum, where morphology is analysed as a continuum from ‘the most clearly inflectional patterns to the most clearly derivational patterns’ (Haspelmath & Sims 2010: 81). I follow Haspelmath & Sims in arguing that morphology is best understood as a continuum, since there are numerous examples of borderline cases and overlap between inflectional and derivational morphology, which I will discuss below.

Haspelmath & Sims list eleven properties that set inflection apart from derivation, and I demonstrate why the dichotomy approach becomes problematic based on the most relevant of these properties. The first property is that inflection is relevant to syntax, while derivation is not, defining syntax as agreement and government. However, defining syntax in such a way is not optimal, while acknowledging that this thesis does not investigate case government. It is generally true that derivational morphology will trigger syntactic differences while inflectional morphology will not, e.g. *the person walk-s*, *the person walk-ed* but **the person walk-er*. The syntactic criterion also demonstrates how certain forms can be both inflectional and derivational, as *I am walk-ing* should be analysed as inflectional while *the walk-ing person* is rather derivational, suggesting that the English gerund is positioned

somewhere between inflection and derivation on the morphological continuum. These are primarily examples of the ninth property, which states that 'canonical inflection does not change the word-class of the base; derivational affixes may change the word-class of the base' (Haspelmath & Sims 2010: 96).

The second property states that 'inflectional features are obligatorily expressed', while derivation is not (Haspelmath & Sims 2010: 92). This is relevant for the category of preverbs that are found in all endemic language families of the Caucasus (cf. section 6.11), as they are not conventionally analysed as inflection. Preverbs are however obligatory in many Kartvelian verb forms (Tuite 1998: 19), and preverbs are also obligatory for spatial reference in many languages of these language families. This demonstrates that obligatoriness as a criterion is not unproblematic, as the functional context determines whether something is obligatory or not, cf. Nakh-Dagestanian local cases which are obligatory for nominal spatial referencing.

The fourth property states that inflection should 'express the same concept as the base' while derivation should 'express a new concept' (Haspelmath & Sims 2010: 93). This is an inherently problematic criterion, because how do we define a 'new concept'? The spatial preverbs mentioned above pose this question, as is there a conceptual difference between 'to go', 'to go up' and 'to go down'? Similarly, Haspelmath & Sims classify *reread* as derivation, while functionally identical constructions in some Northwest Caucasian languages could just as well be analysed as inflection. This suggests that the definition of what qualifies as a new concept has the potential of arbitrarily dividing morphology into inflection and derivation, which is less problematic if morphology is understood as a continuum.

The seventh property claims that 'canonical inflection is expressed at the periphery of words; canonical derivation is expressed close to the root' (Haspelmath & Sims 2010: 95), which they stress is not an absolute property. The examples given indicate that they analyse causatives and reflexives as derivation, which is conventional but still problematic. The status of reflexives as derivation can be questioned by testing this property on the Northwest Caucasian languages, as they all have reflexive prefixes at the periphery, i.e. in the same positions as the clearly inflectional absolutive and indirect objects (Arkadiev & Lander 2020: 404-408). It is also counter-intuitive to claim that, e.g. 'to dress oneself' and 'to dress someone else' are two different concepts, particularly since they often just include a reflexive object instead of a direct/indirect object marker, e.g. French *je me lave* 'I'm washing myself' and *je te lave* 'I'm washing you'.

Causative affixes are somewhat different, as they are generally found closest to the root in the Caucasus, which would indicate that it should be analysed as derivation. However, there are two counterarguments, the first of which is connected to the position of the causative suffix in Turkic languages. If the post-radical position of the causative suffix in Turkic languages is used as an argument for it being derivational, then the other suffixes in the same position ought to be derivational as well, i.e. the reflexive, reciprocal and passive suffixes (Johanson 2022a: 37). This becomes problematic, not only because of the reflexive discussed

above, but also due to the passive voice having clearly inflectional properties in, e.g. Classical Greek (Morwood 2002). If passives are derivational by definition, then this should ideally apply to passives in all languages. Since passive and causative suffixes occupy the same slot in most Turkic languages, the inflectional properties of passives should apply to causatives as well. The second counterargument is causative constructions in Georgian where the causative suffix *–in* follows the thematic suffix (Hewitt 1995: 411-416), but this is a weak argument for a property that is not absolute.

The tenth and eleventh properties are both relevant for the discussion above, as they postulate that inflection may be cumulative while not iterative, whereas derivation cannot be cumulative but iterated (Haspelmath & Sims 2010: 98). Cumulative in this sense refers to the tendency for single affixes to convey multiple grammatical functions simultaneously, which is a well-known feature of inflectional morphology. This further supports the analysis that passives are inflectional, as e.g. passive forms in Classical Greek (Morwood 2002: 64) are arguably cumulative. The iterative property refers to the iteration of affixes, which is a counterargument for causatives being inflectional, as double causatives do occur in many languages, e.g. Huallaga Quechua (Haspelmath & Sims 2010: 98). Double causatives do occur in the Caucasus as well, but they often have a slightly different meaning and function than single causatives, cf. the description of the causative function in section 6.8.2.

I have decided to adopt an unconventionally wide definition of inflection, which will inevitably affect the results. This is primarily motivated by the phonological approach of this thesis, since my definition of affixes as phonological affixes (cf. section 2.3) makes a more inclusive definition of inflectional affixes more relevant.

2.5. Morphological complexity

Morphological complexity is a well-studied phenomenon, yet it lacks both a general definition and a universally accepted methodological framework. What is morphological complexity and how do we measure it? Affixation and stem alternation are the primary representations of morphology, and as affixes are 'considered the canonical exponent of inflection' (Baerman, Brown & Corbett 2017: 8), the relevance of affixation as an operationalisation of morphology becomes apparent. Sagot (2013) proposes a distinction between *counting-based*, *measures*, *description-based* and *entropy-based complexity measures*, which all lead to highly different investigations. The counting-based complexity measure has been frequently used in previous research on morphological complexity, and involves counting the number of features that are distinguished in a specific morphological system (Baerman, Brown & Corbett 2015: 5).

Although the counting-based approach is a familiar methodology, there are some important disadvantages, as a counting-based approach presupposes that the data

analysed are commensurable. This is highly important, as we have to be certain that we are comparing the same features and functions, otherwise the validity of the results will decrease. Miestamo discusses commensurability or 'comparability' with the implicit purpose of comparing different grammatical categories (Miestamo 2008: 30), while commensurability is even more important within the same grammatical category from a functional perspective.

The description-based and entropy-based complexity measures examine the morphological complexity of actual word forms and how they can be predicted (Sagot 2013), which unfortunately make them irrelevant for this thesis as actual word forms are not considered. I have therefore decided to use the conventional counting-based approach, while applying it to the entire nominal and verbal affixation systems and analysing the affixes binarily by the grammatical functions they express. This will counteract some of the issues described by Sagot (2013), as only totalling the counts for each grammatical category, e.g. the number of cases or genders, is only relevant if all categories are investigated on a systemic and not a paradigmatic level. Chapters 5 and 6 are therefore devoted to the systematic description and reanalysis of all observed grammatical functions expressed by affixation in the Caucasus, which is complicated by the great variation between how the five language families of the Caucasus have been conventionally and historically described.

Nichols (2020) proposes two measures of linguistic complexity, i.e. *enumerative complexity* (EC) and *canonical complexity* (CC), where EC is 'is based on assessing the number of elements in an inventory or values in a system' (Nichols 2020: 163), which makes EC similar to the counting-based complexity measure described above. Nichols also describes the EC as having disadvantages, while citing the same issues with commensurability and simply comparing the sizes of grammatical inventories. Using a binary approach to compare grammatical functions will largely avoid these issues, as it requires a thorough definition of all functions and enables comparisons of actual features. Nichols relates the CC to *descriptive complexity* and *Kolmogorov complexity* (Nichols 2020: 164), which aligns CC with the *entropy-based complexity measures*, as Kolmogorov complexity relates to information entropy. Nichols defines CC as 'determining the central, or ideal, position' in a logical space, i.e. the canonical position, while 'any departure from that ideal is non-canonical' (Nichols 2020: 164). Nichols also compares CC and EC in ten Nakh-Dagestanian languages (Nichols 2020: 180), while exactly how these complexity measures have been obtained and what they refer to remains somewhat opaque. As both EC and CC, like the description-based and entropy-based complexity measures, appear to largely deal with predictability and non-transparency, they are not applicable as complexity measures for the approach of this thesis.

2.6. The Caucasus – the perfect case?

The languages of the Caucasus are extraordinary in two ways as they have the world's most elaborate affixation patterns (Daniel & Ganenkov 2009: 671) and they employ some of the largest consonant inventories outside Africa (Beguš 2020: 699). Two Caucasian languages are in the top ten of the largest consonant inventories listed in PHOIBLE 2.0 (Moran & McCloy 2019), i.e. Lezgian (Nikolaev et al. 2015) and Archi (Kodzasov 1977). However, the largest consonant inventory in the Caucasus famously belonged to Ubykh at between 80 and 84 consonant phonemes (Vogt 1963: 13; Fenwick 2011: 16), whereas the consonant inventories described in PHOIBLE for Lezgian and Archi should be problematised as I demonstrate below.

Haspelmath (1993: 34) only lists 54 consonant phonemes in Lezgian, while mentioning the potential of adding 'more than a dozen' of palatalised consonants as effects of vowel syncope (Haspelmath 1993: 38) which are included in PHOIBLE (Nikolaev et al. 2015). Kodzasov (1977) describes 81 consonants in Archi, but this includes pharyngealised uvulars which should rather be analysed as a suprasegmental feature indicating stress (Kibrik 1994a: 303; Chumakina, Bond & Corbett 2016: 20-21). Nevertheless, if all 56 languages investigated in this thesis had been included in PHOIBLE, they would comprise 21 of the 50 largest consonant inventories in the world, cf. appendix C.

The area is furthermore an excellent case to study as it contains three endemic language families with a remarkable linguistic diversity between the language families. The presence of non-endemic Turkic and Indo-European languages is also interesting, as they constitute a good basis for comparison and indications of pan-Caucasian tendencies and language contact. The Caucasus is also an interesting area to investigate from a morphological perspective since numerous grammatical categories are shared across the language families although they are expressed by means of highly different affixation patterns. This is important as it enables fine-grained morphological comparisons that would be difficult with a global sample, since that would limit any comparative study to the most core aspects of morphology, e.g. case, tense and person marking. The complexity of Caucasian morphology can also provide us with a better understanding of the oppositional nature of grammar in relation to lexicon. The languages of the Caucasus cannot possibly tell us anything about the limitations of morphology, but since they convey linguistic information through affixation that is most often expressed analytically in the rest of the world, they can manifest some basal functions of morphology.

The combination of complex phonology and complex morphology is exceptional in the Caucasus, but there are other areas of the world that would be equally interesting to investigate. The world's largest phoneme inventories are found in the Tuu, the Kx'a, and to a lesser degree the Khoe–Kwadi language families of southern Africa. While many Tuu and Kx'a languages have limited morphology (Güldemann 2013), the Khoe–Kwadi languages have complex patterns of tonal morphology (Voßen 1997). The Southern Bantu Nguni languages are an interesting group to

study in this regard, as they include, e.g. Xhosa, which has an extremely rich phoneme inventory and complex morphology (Mzamane 1962).

The various endemic language families of the North American Pacific coastal region are also relevant, as it is a region with a general presence of languages with complex morphology and large phoneme inventories, from the northern Na-Dené languages, e.g. Tlingit (Story 1979), to the Salishan and Wakashan languages, e.g. Nuu-chah-nulth (Nakayama 2001). There are also large phoneme inventories and complex morphology further inland in the non-coastal Southern Athabaskan languages, e.g. Navajo (Sapir & Hoijer 1967).

Northern North America, apart from the Pacific region, is of great interest in relation to affixation as notoriously polysynthetic language families such as Inuit–Yupik–Unangan, Wakashan and Iroquoian (Anderson 2015; Koenig & Michelson 2015) are highly relevant to investigate. However, these language families do not have particularly rich phoneme inventories, which makes them less interesting than the languages of the Caucasus and the North American Pacific if the intention is to investigate the interaction between morphology and large phoneme inventories. The Oto-Manguean languages are also interesting in this context due to their morphological and phonological complexity, while the morphology in, e.g. the Chinantec languages, rely heavily on tone (Baerman, Brown & Corbett 2017: 139-144), which is not considered in this thesis. The central Andes could also constitute a promising case in which to investigate this phenomenon, as particularly the Quechuan, Aymaran and Uru-Chipaya languages all have complex affixation patterns and large phoneme inventories (Adelaar & Muysken 2004).

2.6.1. The Caucasian Sprachbund

The notion of a Caucasian linguistic area, i.e. the *Caucasian Sprachbund*, has been thoroughly debated within Caucasiology, primarily driven by observed phonological and morphological similarities. The concept of *Sprachbund* was coined by Trubetzkoy, who lists four positive criteria of strongly similar *syntax*, similar *morphological structures*, a large amount of *shared cultural vocabulary* and (superficial) *phonological similarities* (Trubetzkoy 1928: 18). Klimov (1965) and Catford (1977) present some of the earlier comparisons of both phonology and morphology across the Caucasus, but the notion of the Caucasian Sprachbund is complicated by the longstanding effort to link the three endemic language families into one Caucasian language family, of which Uslar (1888: 35) is one of its earliest proponents (Tuite 2008b). Čikobava later introduced the concept of the *Ibero-Caucasian* language family inspired by Uslar (Čikobava 1965), which lives on in contemporary Georgian linguistics, e.g. Kurdiani (2016). In more recent literature, Chirikba (2008) systematically investigates the proposed Caucasian Sprachbund, largely based on Klimov (1978), by postulating 34 diagnostic features divided into phonological, morphological, syntactical, 'lexical semantic' and lexical features

(Chirikba 2008: 41). Twenty-three of these features are phonological and morphological, which makes his investigation highly relevant for this thesis.

Table 2.3: Chirikba's phonological and morphological diagnostic features (Chirikba 2008: 41).

Phonological	**Morphological**
Rich consonantism	Agglutination
Ternary contrast of stops and affricates	Polysynthetism
Glottalisation	Predominace of prefixal conjugation
Rich sibilant systems	Predominance of postpositional constructions
Rich postvelar (uvular, pharyngeal and laryngeal) systems	Masdar (verbal noun)
Similarly built harmonic clusters	Morphological marking of causative
Presence of schwa	Category of evidentiality
Lack of phonemic diphthongs	Category of potential
Lack of vocalic clusters	Attachment of coordination markers to each conjunct
Ablaut	Directional and orientational preverbs
	Group inflection
	A three-grade deictic distinction
	Vigesimal numeral system

The syntactic features compromise 'identical word order (SOV, Attr-N)', ergative constructions, inversive constructions and 'the possessor constituent precedes the possessed one' (Chirikba 2008: 41). Most of these syntactic features are evidently feeble, as a SOV word order combined with attribute-noun and possessor-possessed order are among the most common word order patterns in the world (Dryer & Haspelmath 2013). The lexical semantic features cover a stative/dynamic verb distinction, inversive verbs, ambitransitive or labile verbs and suppletive verbs for singular and plural arguments, and the final lexical features deal with common cultural terms, common phraseology and common semantic patterns, but this goes beyond the scope of this thesis.

The phonological features are worth discussing as Chirikba mainly gives quantitative arguments such as the number of sibilants and postvelars or the presence of glottalised consonants, i.e. ejectives. The ternary contrast of stops and affricates, e.g. the three-way distinction between voiced [b], aspirated voiceless [p^h] and ejective [p'] in Georgian (Bolkvadze & Kizira 2023: 16), is generally found in the Caucasus, but a non-ejective ternary contrast was also present between [b], [p] and [p^h] in Classical Greek (Woodard 2008: 16), and still is in Eastern Armenian (Dum-Tragut 2009: 17). Ejectives are one of the most salient features of the languages of the Caucasus, but the actual inventories of ejectives differ noticeably between the language families. Alveolar lateral ejective affricates [tɬ'] and uvular ejective affricates [qχ'] are only found in the Nakh-Dagestanian languages while

particularly the Circassian languages stand out, as they are among few languages in the Caucasus to have ejective fricatives, e.g. [fʼ], [ɬʼ], [sʼ], [ɕʼ] and [ʃʼ].[1]

His comparisons of Caucasian and various Eurasian sibilant inventories indicate that the Caucasian languages generally have larger sibilant systems, but many Indo-European languages are on the same level as most of the non-Northwest Caucasian languages (Chirikba 2008: 45-47). The Northwest Caucasian languages unsurprisingly stand out, while the high number in Tabasaran is clearly linked to consonant lengthening and labialisation (Babaliyeva 2013: 17). The rich postvelar systems are a shared feature of the three endemic language families, but it is slightly misleading to link it to postvelars in general since the high frequency of uvulars is the only feature found in all three language families. Pharyngeals are completely absent in the Kartvelian languages, and glottal consonants are quite common in the languages of the world (Moran & McCloy 2019).

The diagnostic features connected to vowels are somewhat problematic, as e.g. the presence of schwa can hardly constitute a Sprachbund feature since it is the most 'neutral' vowel, and it is not a phoneme in many if not most languages of the Caucasus. The lack of phonemic diphthongs and vocalic clusters, i.e. hiatus (Chirikba 2008: 50), are also problematic diagnostic features as they are negative features that only are valid if diphthongs and vocalic clusters are common in the rest of the world. Ablaut is an interesting diagnostic feature however, but for it to be a sound Sprachbund criterion it has to be compared to the global presence of ablaut, e.g. Indo-European ablaut patterns.

The morphological features also need to be discussed further, as many of them are not optimal or even misleading. The first features of agglutination and polysynthesis are connected, while only the Northwest Caucasian and Kartvelian languages can be classified as truly polysynthetic (Polinsky 2020: 12). Using agglutination as a Sprachbund feature is not optimal, since it holds true for many Eurasian language families and is therefore not sufficiently distinctive. The claim that there is a predominance of prefixal conjugation in the Caucasus is questionable, whereas person agreement is fully prefixal in the Northwest Caucasian languages, it is both prefixal and suffixal in Kartvelian. The situation in Nakh-Dagestanian is complicated by the widespread presence of noun class marking prefixes, which constitute a part of the conjugation systems, while person-marking is more often suffixal in Nakh-Dagestanian languages with person agreement.

The predominance of postpositional constructions is not a distinctive Sprachbund feature due to the general SOV word order pattern, as the combination of postpositions and SOV is the most common pattern worldwide (Dryer & Haspelmath 2013). The masdars, i.e. verbal nouns, could constitute a distinctive feature, but their relationship with verbal nouns in general is somewhat undefined. The morphological categories of causative, evidentiality and potential are

[1] There are also ejective fricatives in the Nakh-Dagestanian languages Bagvalal and Chamalal, i.e. [sːʼ] and [ʃːʼ] (Kodzasov 2001: 35; Magomedova 2004: 4).

interesting diagnostic features, but they need further scrutiny. Causatives affixes are found in most Caucasian languages, but morphological marking of causatives is also a very common feature globally (Song 2013). Although evidentiality is a somewhat uncommon feature in western Eurasia, it is by no means uncommon in Eurasia in general nor worldwide (De Haan 2013). The category of potential is interesting, since it is only universally marked by explicit morphological forms in the Northwest Caucasian languages, whereas it is marked in some but not all Nakh-Dagestanian languages and by using passive constructions in Kartvelian (Chirikba 2008: 52).

Chirikba claims that the attachment of coordination markers to each conjunct is found in all three language families of the Caucasus, as he gives the examples Megrelian *ma-ti si-ti* and Svan *m-i s-i* (Chirikba 2008: 53), both meaning 'me and you'. However, this is not true for nominal conjuncts in either Megrelian (Harris 1991b: 364) nor Svan (Gudjedjiani & Palmaitis 1986). Directional and orientational preverbs are found in all three language families, and it is an interesting pan-Caucasian feature. The feature Chirikba refers to as 'group inflection' concerns the morphological marking of only the last constituent of the NP, which can hardly be a distinctive Caucasian feature as this is a widespread phenomenon, e.g. phrasal affixes and clitics. The final two features, three-grade deictic distinction and vigesimal numeral system, go beyond the scope of this thesis, as I do not include deictic distinctions and numeral systems in my investigation, but I question whether deictic distinctions and numeral systems should be considered morphological features unless they are expressed morphologically.

If we summarise the discussion above, it becomes apparent that only *ternary contrast of stops and affricates*, the *presence of ejectives*, the *high frequency of uvulars*, *ablaut*, *masdars*, *causatives*, *evidentiality*, *directional and orientational preverbs*, *ergative constructions* and *inversive constructions* should be kept as possible phonological, morphological and syntactical diagnostic features that apply to all three language families. Many of these features are not optimal as distinctive Sprachbund features, since some of them are rather common worldwide.

2.7. Hypotheses

At the core of this thesis lie two hypotheses:

(1) Larger phoneme inventories enable larger inventories of affixes, as each new phoneme adds new potential distinctive affixes. Consequently, larger affix inventories enable more grammatical functions to be expressed through affixation.

(2) There is a distributional difference of segments between affixes and lexical stems, as affixes form a closed class, the lexical stems will inevitably contain more types of segments, while certain segments will be more common in the affixes and the lexical stems respectively.

The first hypothesis is based on a purely mathematical assumption that the number of potential mono-syllabic (and mono-consonantal) morphemes in a language can be calculated by the following equation, where *C* is the number of all consonant phonemes (including secondary articulation), *V* is the number of all vowel phonemes (including secondary articulation, diphthongs and tones), *ωCL* is the number of all consonant phonemes and licit consonant clusters allowed in the onset, *κCL* is the number of all consonant phonemes and licit consonant clusters in the coda and *ν* is the number of phonemes allowed as nuclei:

$$\Sigma(MORPH) = C + V + (\omega CL * \nu) + ((\omega CL * \kappa CL) * \nu) + (\nu * \kappa CL)$$

The sum of potential mono-syllabic morphemes naturally applies to both lexical stems and affixes, but does not apply to all potential syllable types in languages that do not allow consonantal nuclei. The relevance of this equation for affixation is based on another assumption, i.e. that the optimal affix from a productional perspective is mono-syllabic or mono-consonantal, which are tested in section 7.2. For most languages $\nu = V$ as only vowels are allowed as nuclei in the vast majority of the world's languages, while Donohue et al. (2013) found that out of a sample of 2181 languages, consonantal nuclei of some sort were permitted in 11.3% of these languages (Gordon 2016: 109).

The equation above assumes that the number of consonants should be the primary factor in the number of potential mono-syllabic morphemes, as $C = 1$, even with *CVC* syllable structure, would only yield $\Sigma(MORPH) = C + V + V + V + V$. The lack of codas in many of the world's languages is not an issue for the equation at hand, since this would just yield $\Sigma(MORPH) = C + V + V + 0 + 0$, which further supports the assumption that consonant phonemes are fundamental to the number of syllable types. However, the licit consonant clusters cannot be calculated by this equation, as both *ωCL* and *κCL* will be affected by factors such as the sonority hierarchy, and licit consonant clusters are largely unpredictable in the world's

languages, as e.g. the highly restrictive *κCL* in Mandarin cannot be predicted by its *C*. The equations above do not take into account the interaction between consonants and vowels, as many languages do not allow certain consonant-vowel combinations depending on the syllable structure.

The first hypothesis of this thesis contradicts an 'assumed truism in linguistics', i.e. the *trade-off hypothesis* (Bentz et al. 2023), that linguistic complexity has a tendency to balance itself out, i.e. increased complexity for one variable leads to decreased complexity in another variable and vice versa (Hockett 1958: 180-181; Moran & Blasi 2014: 217). Bentz et al. (2023) recently tested trade-offs and 'equi-complexity' between morphological and syntactical complexity, which showed that the languages included in their study 'turn out equally complex overall' (Bentz et al. 2023: 16). Although this evidently supports the trade-off hypothesis, the need to differentiate relative complexity, i.e. the 'cost' of using a system, from absolute complexity, i.e. 'the number of parts of a linguistic subsystem', becomes relevant (Moran & Blasi 2014: 218). These are two fundamentally different definitions of complexity, and the term complexity is in itself problematic, as it has a long history of evaluative associations. However, the terms complex and simplex still have an explanatory value, particularly in relation to affixation. This thesis only investigates absolute complexity, as we cannot *prima facie* assume a relationship between morphological complexity and cognitive complexity (Baerman, Brown & Corbett 2017: 3).

The second hypothesis is based on a long-standing observation that 'affixes, particularly inflectional suffixes, in the languages where they exist, habitually differ from the other morphemes by a restricted and selected use of phonemes and their combinations' (Jakobson 1965: 29). Bybee (2005) tested this hypothesis quantitatively by investigating affixes expressing tense, aspect and mood in 23 genealogically unrelated languages from across the world, of which Abkhaz was included from the Caucasus. The investigation revealed that the number of consonants used in affixes was smaller than the entire inventory of consonants in all 23 languages, but the difference was only significantly different from what would be predicted by chance for six languages in the study while Bybee acknowledges that 'the phenomenon in question may be said to represent a tendency, albeit a rather weak one' (Bybee 2005: 175).

Bybee also investigated whether 'highly marked segments' and 'highly complex segments' are excluded from affix inventories, and whether certain classes of segments are excluded altogether through pattern exclusion, while she inconclusively concluded that her hypotheses were generally weakly supported although 'neither hypothesis provides a good fit with the data' (Bybee 2005: 191-192). This indicates that further research into the phonology of affixation is warranted, as Bybee only examined TAM affixes in a small number of randomly sampled languages, and the hypotheses were primarily concerned with exclusion which runs a risk of becoming overly binary. It is therefore highly relevant to

investigate distributions of segments in both nominal and verbal affixes rather than just investigate the presence or absence of certain segments.

3. The Caucasus

The Caucasus region coincides with the Greater and Lesser Caucasus mountain ranges between the Black Sea and the Caspian Sea, and it is conventionally described as constituting the geographical and political border of Europe and Asia. The region is primarily known for its ethnic and linguistic diversity, as well as political tensions due to its long-standing position as a marchland between various culture spheres and empires. The Caucasus is conventionally divided into the South Caucasus or Transcaucasia) and the North Caucasus (or Ciscaucasia). The South Caucasus more or less coincides with the independent states of Georgia, Armenia and Azerbaijan. The North Caucasus is entirely within the Russian Federation and comprises the seven republics of Adygea, Karachay-Cherkessia, Kabardino-Balkaria, North Ossetia-Alania, Ingushetia, Chechnya and Dagestan. The wider North Caucasus region sometimes also includes the krais of Krasnodar and Stavropol, the Rostov oblast and the republic of Kalmykia. I will present a brief historical outline before delving into the complex and highly interesting linguistic diversity of the Caucasus.

3.1. A brief history of the Caucasus

The ancient history of the Caucasus is dominated by various indigenous political entities such as Colchis, Caucasian Iberia, Urartu, Armenia and the later encroachment of various Iranian, Hellenic and Roman empires. The presence of ancient Iranian peoples in the Caucasus can be divided into two groups, the various East Iranian peoples in the north such as the Cimmerians, Scythians and Sarmatians, and the North-West Iranian Medians in the south-east. The East Iranian Alans, i.e. the ancestors of modern Ossetians, founded a long-lasting kingdom in the North Caucasus already in late antiquity (Coene 2010: 111; Belyaev 2020: 574). In the first centuries AD the region was amongst the earliest in the world to adopt Christianity, which resulted in early Christian texts in Armenian, Georgian and Caucasian Albanian[2], a Nakh-Dagestanian language that was rediscovered in the 1930's but not truly deciphered until the breakthrough findings of Caucasian Albanian palimpsest in the early 21st century (Gippert et al. 2008). The first Turkic

[2] Not to be confused with Albanian, which is an Indo-European language spoken in the Balkans.

peoples started to migrate into the steppes north of the Caucasus from Central Asia in the 5th century (Károly 2022: 145), where the Khazar khaganate was an early Turkic political entity in the North Caucasus, and the Khazars likely spoke an Oghur Turkic language (Károly 2022: 150). The Oghur migrations were later followed by Kipchaks in the 11th century (Coene 2010: 111), i.e. the ancestors of modern Kumyks, Karachays and Balkars, and the Kipchak language Cuman was historically spoken across the steppes to the north of the Caucasus.

The most important migration into the region was the arrival of the Oghuz Turks, i.e. the precursors of the Azerbaijanis and the Turkish, who migrated from Central Asia as a result of the expansion of the Turco-Persian Seljuk Empire in the 10th and 11th centuries (Coene 2010: 111-112; Forsyth 2013: 95-97). The Caucasian Middle Ages were characterised by frequent shifts in the power balance between the Persian Empire, various Turco-Mongol empires such as the Timurid Empire, and the kingdom of Georgia. which experienced its largest extent in the 11th to 13th centuries. This period is often described as the golden age of Georgia (Coene 2010: 113), particularly under the reign of Queen Tamar (1184-1212), as she exerted influence over almost the entire Caucasus (Coene 2010: 114; Forsyth 2013: 143-146). Medieval Dagestan was divided into various independent political entities primarily ruled by Kumyks, Avars, Laks and Nogais (Forsyth 2013: 166-167). Despite the difficult terrain of the region, Dagestan was one of the principal markets of the Caucasus and consequently in extensive contact with the surrounding regions, whereas the Avars were generally more isolated (Forsyth 2013: 167-169).

The Mongol Empire invaded the Caucasus in the 13th century, which ended Georgia's dominant status in the region, as it became a vassal of the Mongol khan (Coene 2010: 120). The Mongol Empire eventually fractured into the Golden Horde ruling the North Caucasus and the Ilkhanate controlling much of the South Caucasus, which was followed by the disruptive invasions of the Timurids in the late 14th century (Coene 2010: 117). The South Caucasus was divided in the 15th century into various Georgian kingdoms and the Oghuz states of Qara Qoyunlu and later Aq Qoyunlu (Coene 2010: 118). The situation stabilised in the 16th century, as the South Caucasus was effectively divided between the Ottoman Empire in the west and the Persian Empire in the east (Henze 1992: 64; King 2008: 23; Coene 2010: 120), with the Circassians, Ossetians, Chechens, Ingush in the North Caucasus being largely self-governing (Forsyth 2013: 200).

This would drastically change in the 19th century[3] with the Russian Empire's subsequent conquest of the Caucasus ending in 1864 (Forsyth 2013: 285), thus initiating the Russian hegemony which largely prevails to the present day. The Russian conquest of the North Caucasus was particularly devastating as it was marred by ethnic cleansing, displacement and mass deportations of almost the entire

[3] The Russian Empire briefly annexed large swathes of the Caucasus during the reign of Peter the Great, while the enduring annexation process was initiated under Catherine the Great in the late 18th century (King 2008: 25-26).

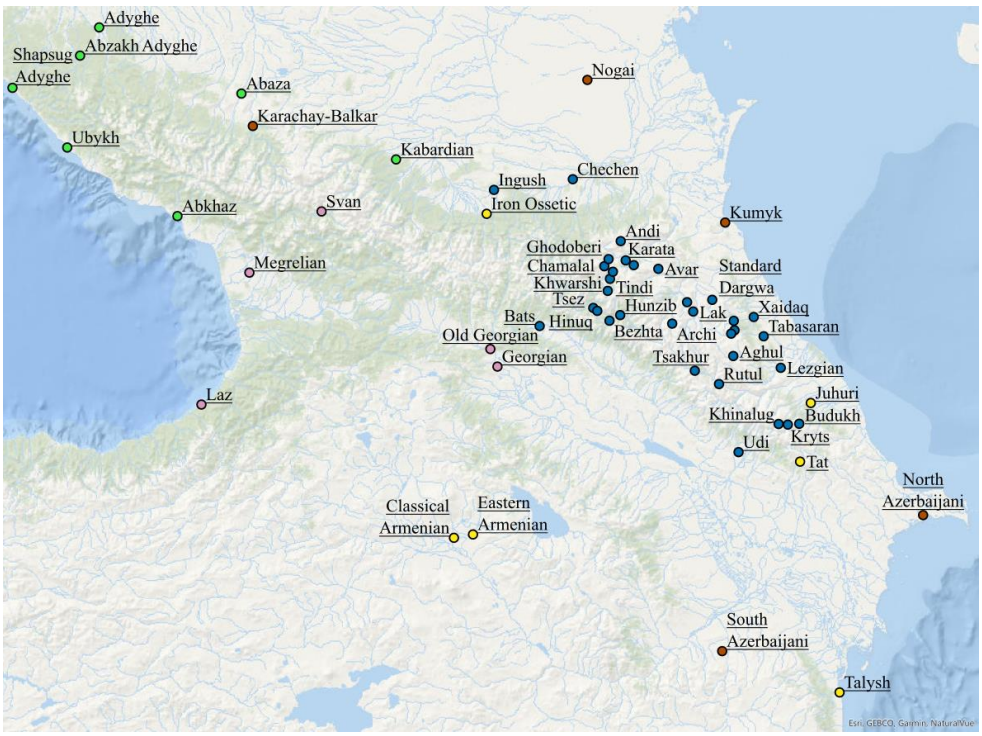
Circassian nation, sometimes described as the *Circassian genocide* (Richmond 2013). The Nogais were also displaced in these wars (Forsyth 2013: 293), which explains why the Nogais today are mainly found in north-eastern Dagestan and Karachay-Cherkessia (Dobrushina, Daniel & Koryakov 2020: 34). Other peoples of the North Caucasus, such as the Chechens, also fiercely resisted the Russian invasion, which resulted in mass expulsions of Chechens to the Ottoman Empire in the 19th century (Forsyth 2013: 295). The Russian Revolution followed some fifty years after the full annexation of the Caucasus, and it resulted in the emergence of the independent republics of Georgia, Armenia and Azerbaijan in 1918 (Coene 2010: 132). The *Republic of the Union of Mountain Peoples of the Caucasus* (or the 'Mountain Republic'), which was a union of Circassians, Chechens, Ingush, Dagestanians, Ossetians, Karachay-Balkars and Nogais, was however the first to proclaim full independence from the recently proclaimed Russian Soviet Republic in 1918 (Forsyth 2013: 360-365). These independent republics were short-lived, as the entire Caucasus would become annexed by Soviet Russia in 1920-1921 (Forsyth 2013: 424-426).

The Soviet Union was declared in December 1922, and the early Soviet period saw an increase in the documentation of the languages of the Caucasus, which was an effort to increase the status and literacy of the various languages of the region (Forsyth 2013: 441; Polinsky 2020: 5). The Chechens, Ingush, Karachays and Balkars would later be severely oppressed during World War II, as the entire populations of these ethnic groups were deported to Central Asia in 1944 (Coene 2010: 137) and they were not allowed by Soviet authorities to return until 1956 (Kazenin 2020: 72, 78-80). During the Chechen and Ingush exile many Dagestanians were forcibly relocated to Chechnya and Ingushetia by Soviet authorities, but they all returned after 1956 (Comrie, Xalilov & Xalilova 2015: 16-17; Isakov & Xalilov 2012: 10; Forker 2013: 6). The fall of the Soviet Union in 1991 led to the re-emergence of Georgian, Armenian and Azerbaijani independence, but it also ignited numerous violent conflicts such as the Chechen wars, the Abkhaz war, the conflict in South Ossetia, the Nagorno-Karabakh conflict[4] and the Ingush-Ossetian conflict over Prigorodnyj raion (Coene 2010: 141-159), many of which remain latent and unsolved to this day.

[4] Which began already in 1988 (Coene 2010: 145-147).

3.2. A linguistic overview of the Caucasus

The more than 60 languages that are indigenous to the Caucasus can be divided into five language families: Kartvelian, Nakh-Dagestanian, Northwest Caucasian, Indo-European and Turkic languages. The first three of these language families are endemic to the Caucasus, and they are conventionally regarded as constituting three separate language families. Although numerous efforts have been made in the past to merge them into one or two families, any genealogical relationship between these language families has not been convincingly demonstrated (Polinsky 2020), although the grouping of two or all three language families are proposed by e.g. Chirikba (2008), Kurdiani (2016) and Chukhua (2019).



Map 3.1: All 56 languages included in the data, represented by coordinates and colour-coded according to language family.

The Kartvelian or South Caucasian family has been the most spoken of these endemic language families, with an estimated 4.4 million speakers (Testelets 2020: 492)[5]. However, according to Russian and Azerbaijani census data from 2019-2020,

[5] This number is potentially too high, since the latest census states 3.25 million native speakers of Georgian in Georgia, which would also include Megrelian and Svan (National Statistics Office of Georgia 2014).

the Nakh-Dagestanian languages are today spoken by more than 4.5 million speakers (State Statistical Committee 2019; Federal State Statistic Service 2020), and they are clearly the most numerous as they number potentially more than 40 languages (Dobrushina, Daniel & Koryakov 2020). The two most spoken languages in the region are Russian and Azerbaijani, both belonging to the non-endemic language families Indo-European and Turkic respectively.

3.2.1. Kartvelian languages

The Kartvelian (or South Caucasian) language family is population-wise the largest of the three indigenous language families of the Caucasus, and consists of the four languages Georgian, Megrelian, Svan and Laz, which are all spoken in or around Georgia in the western South Caucasus. Georgian is the most spoken language of the family at somewhere between 3.25 and 4 million speakers (National Statistics Office of Georgia 2014; Testelets 2020: 492), and the culturally and politically most important as the official language of the Republic of Georgia. Georgian has a long written record, stretching back to the advent of Christianity in Georgia in the 4th century AD (Tuite 2008a: 145), and it has since been the liturgical language of the Georgian Orthodox Church.

Georgian has been written in the well-known Georgian alphabet known as *Mxedruli*, the 'knightly' script, since the 11th century (Shanidze 1982: 12; Tuite 2008a: 147), while it was originally written in the *Mrglovani* 'rounded' or *Asomtavruli* 'majuscule' script between the 5th and 9th centuries and later the *K'utxovani* 'angular' or *Nusxuri* 'minuscule' script between the 9th and 11th centuries (Shanidze 1982: 11-12; Hewitt 1995: 4; Tuite 2008a: 146-147). The Georgian language area is almost entirely within the borders of the modern Republic of Georgia, while there are also Georgian-speaking communities in Russia, Turkey and Azerbaijan (Testelets 2020: 491; Bolkvadze & Kiziria 2023: 6). The Fereydani Georgian dialect is spoken in central Iran by descendants of Georgians who were deported from Georgia by Shah Abbas I in 1614 (Bakuradze, Beridze & Pourtskhvanidze 2020). There is also a Judeo-Georgian language, which is still spoken by the Georgian Jewish community, but due to mass emigration from the 1970s to the 1990s it is mainly spoken in Israel today (Lomtadze & Enoch 2019: 23).

The second largest Kartvelian language is Megrelian, which is spoken by more than 300,000 speakers in the region of Samegrelo in western Georgia and in south-eastern Abkhazia (Rostovtsev-Popiel 2020: 530). The exact number of speakers is unknown, as Megrelian has no official status in Georgia. Even though Megrelian is generally not a written language it has an important body of literature and Megrelian newspapers were published in the early Soviet period (Rostovtsev-Popiel 2020: 530-531). Megrelian also has a long history of linguistic description as was described already in the late 19th century by Tsagareli (1880). Megrelian and Georgian both belong to the Karto-Zan branch of the Kartvelian languages, where

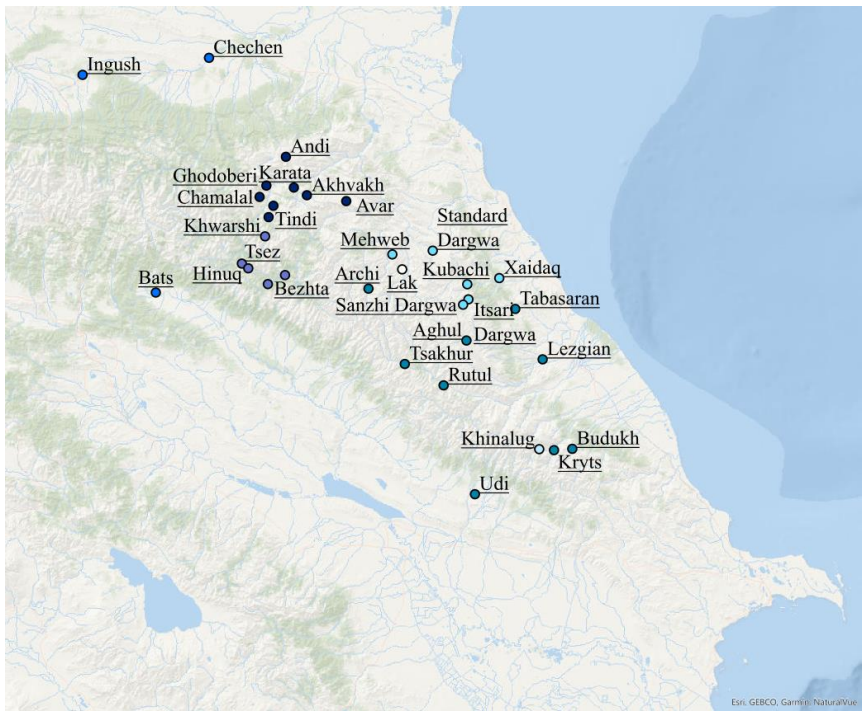
Megrelian belongs to the sub-branch Zan together with Laz (Rostovtsev-Popiel 2020: 529). Laz is primarily spoken in north-eastern Turkey close to the Georgian border and in the Georgian village of Sarpi (Holisky 1991) by an unknown number of speakers, with figures varying from 50,000 (Testelets 2020) to 250,000 (Kutscher 2008: 83). Laz has no official status in Turkey and the use of Laz has been severely limited by the restrictive Turkish minority language policies of the 20th century (Lacroix 2009: 5), but Laz has increasingly become a written language since the introduction of a Latin alphabet orthography in 1984 by Lazoğlu and Feuerstein (Lacroix 2009: 6).

Map 3.2: All Kartvelian languages included in the data, represented by coordinates.

The fourth Kartvelian language is Svan, which is spoken by approximately 50,000 speakers (Tuite 2017: 226) in the mountains of the northwest Georgian region of Svaneti. Svan is the most divergent of the Kartvelian languages and forms a branch of its own separated from the other Karto-Zan languages (Tuite 1997) and there are considerable dialectal differences between Upper Svan and Lower Svan (Schmidt 1991). Svan is generally not a written language, although a sizeable Svan corpus was produced in the early 20th century (Tuite 1997: 3) and an increasing use of written Svan has been observed with the advent of social media in the 21st century (Tuite 2017: 233-239).

3.2.2. Nakh-Dagestanian languages

The Nakh-Dagestanian or Northeast Caucasian language family is the most numerous language-wise of the indigenous language families of the Caucasus, as it contains more than 40 languages (Dobrushina, Daniel & Koryakov 2020). The Nakh-Dagestanian languages are conventionally and geographically divided into two main groups, the Nakh languages and the Dagestanian languages. This division has recently been demonstrated to not be a valid genealogical classification though, as the Dagestanian languages do not seem to form a true taxon (Ganenkov & Maisak 2020: 88). The Dagestanian languages form a highly diverse geographical grouping that comprises more than 40 languages spread across six branches primarily within the borders of the North Caucasian republic of Dagestan, i.e. Avar-Andic, Dargic, Khinalug, Lak, Lezgic and Tsezic (Ganenkov & Maisak 2020).



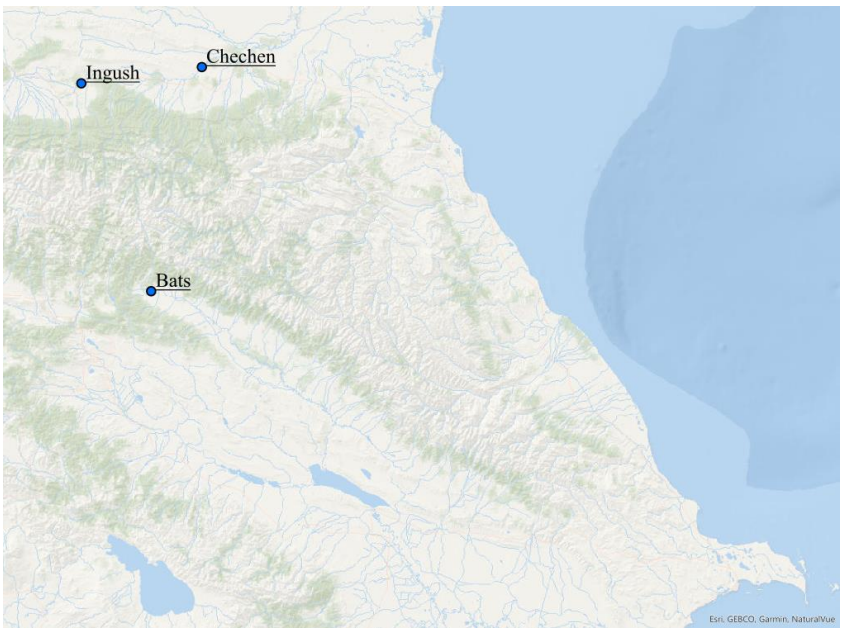
Map 3.3: All Nakh-Dagestanian languages included in the data, represented by coordinates, colour-coded according to branch.

The Nakh branch

The Nakh languages comprise the three closely related languages Chechen, Ingush and Bats. Chechen is the largest of all Nakh-Dagestanian languages with almost 1.65 million speakers (Federal State Statistic Service 2020), and it is the official

language of the North Caucasian republic of Chechnya. Ingush is the official language of the neighbouring republic of Ingushetia with more than 500,000 speakers (Federal State Statistic Service 2020). Both Chechen and Ingush have written standards, and Chechen has a long literary tradition since the early 19th century (Komen, Molochieva & Nichols 2020: 318).

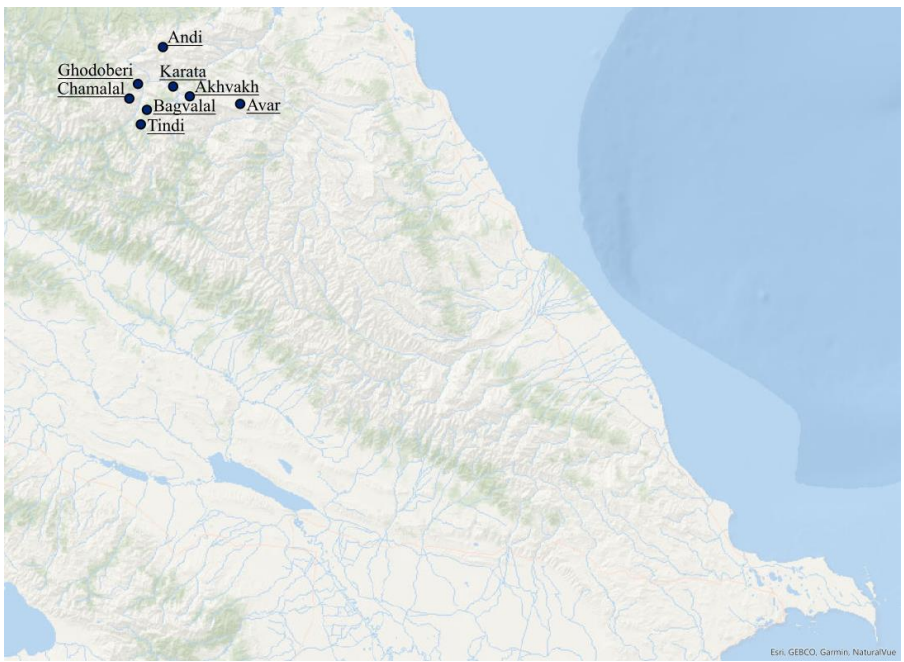
The third Nakh language is Bats (also Tsova-Tush or Batsbi), which is spoken by 275-800 speakers in the village of Zemo Alvani in the Georgian region Kakheti (Holisky & Gagua 1994; Hauk 2020; Wichers Schreur 2024). Bats was previously spoken in the Tusheti region bordering Chechnya, but the Bats relocated to the lowlands in the 19th century due to natural disasters (Hauk 2020). Bats has been heavily influenced by Georgian due to centuries of persistent language contact (Dešeriev 1953: 5; Wichers Schreur 2024) and all speakers of Bats have been bilingual in Georgian since at least the Soviet era (Hauk 2020: 4). Bats has been written in various scripts since the 19th century and there is a recent ongoing initiative to introduce a standardised orthography (Hauk 2020).



Map 3.4: All Nakh languages included in the data, represented by coordinates.

The Avar-Andic branch

The most spoken Dagestanian language is Avar with an ethnic population of at least 1 million (Forker 2020b: 243). As one of the official languages of Dagestan it functions as a lingua franca and is taught in schools in western Dagestan amongst Avars and various Andic and Tsezic peoples (Nichols 2020: 182), including Archis (Alekseev et al. 2012; Forker 2020b) and Mehwebs (Dobrushina 2019a). Avar is mainly spoken in Dagestan but there are also significant communities in northern Azerbaijan, where there are more than 48,000 ethnic Avars (State Statistical Committee 2019), and smaller Avar communities in Georgia (Forker 2020b) and around Güneyköy in western Turkey (Alekseev et al. 2012: 24). Avar is one of few Dagestanian languages to have a literary tradition predating the Russian period, as Avar has a significant body of texts written in the Arabic script from the 15th century up until the successive adoptions of first a Latin orthography in 1928 and later a Cyrillic orthography in 1938 (Forker 2020b: 243). Since the Soviet period, there has also been a rich literary tradition in Avar, and it is widely used in media and in public life (Alekseev et al. 2012). Avar forms the Avar-Andic branch together with the Andic languages described below. It is also important to consider the significant dialectal variation that has been observed between the various dialects of Avar, as some of these dialects might qualify as distinct languages (Dobrushina, Daniel & Koryakov 2020: 30), particularly Zaqatala Avar spoken in Azerbaijan (Forker 2020b: 279).



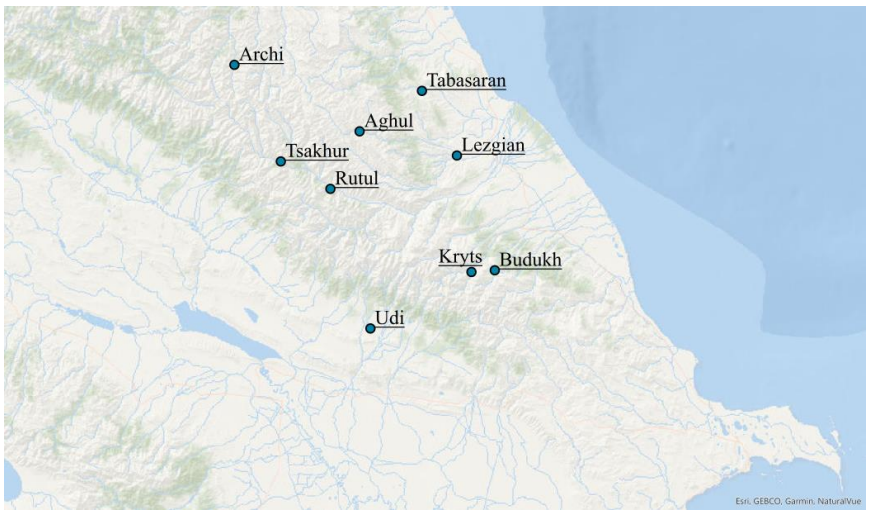
Map 3.5: All Avar-Andic languages included in the data, represented by coordinates.

The Andic languages are a diverse group of languages spoken in a small mountainous area on the border to Chechnya in western Dagestan. Most of these languages are only spoken in a handful of remote villages in a complex patchwork where multilingualism has historically been the norm (Chirikba 2008: 30) and multilingualism is still prevalent among speakers of the Andic languages (Ganenkov & Maisak 2020: 92). The Andic languages all have relatively small speaker communities, and there is no clearly dominant Andic language, as this role is held by Avar (Alekseev et al. 2012). The largest language of the group is Andi in the north, with somewhere around 22,500 speakers (Dobrushina, Daniel & Koryakov 2020: 30). Like Avar, Andi was first written in the Arabic script and a Latin orthography was adopted in 1928, followed by a Cyrillic orthography in 1937, which has been used to produce folklore and literary works (Aglarov 1994).

If considered as one language, Akhvakh would be the second most spoken Andic language, with approximately 20,000 speakers (Creissels 2010: 105), but it is nowadays often divided into Northern Akhvakh with 9,500 speakers and Southern Akhvakh with 8,000 speakers (Dobrushina, Daniel & Koryakov 2020). The Akhvakh results in this thesis refer exclusively to Northern Akhvakh. Northern Akhvakh is also spoken in the village of Axaxdərə in Azerbaijan (Creissels 2010). The remaining Andic languages are found in a continuous area in some of the most inaccessible parts of Dagestan. They are in descending order of speakers: Karata (11,000), Chamalal (9,600), Tindi (9,300), Bagvalal (5,500), Botlikh (7,400) and Ghodoberi (3,200) (Dobrushina, Daniel & Koryakov 2020: 30-31). Most of these languages are fairly well described in Russian, e.g. Bagvalal (Kibrik et al. 2001) and Chamalal (Magomedova 2004), while the only thorough description of Botlikh is a grammar in Georgian by Gudava (1962).

The Lezgic branch

The second largest Dagestanian language is Lezgian or Lezgi, with approximately 546,000 speakers (Dobrushina, Daniel & Koryakov 2020: 31), and it is spoken in southernmost Dagestan and on the other side of the border in northern Azerbaijan (Haspelmath 1993; Babaliyeva 2007). Lezgian is one of the official languages of Dagestan, and has had an important body of literature since the introduction of its first Latin orthography in the 1920s, which was subsequently replaced by a Cyrillic orthography in the 1930s (Haspelmath 1993: 23-24). The influences from neighbouring Azerbaijani are noticeable in Lezgian (Haspelmath 1993: 26), as Azerbaijani has been an important lingua franca in southern Dagestan (Dobrushina, Daniel & Koryakov 2020: 55). Lezgian is the main language of the diverse Lezgic branch, which consists of nine languages on the border between Dagestan and Azerbaijan. Tabasaran is the second largest Lezgic language, with approximately 140,000 speakers (Federal State Statistic Service 2020), just north of the Lezgian language area, and Tabasaran is one of the official written languages of Dagestan (Dobrushina, Daniel & Koryakov 2020).



Map 3.6: All Lezgic languages included in the data, represented by coordinates.

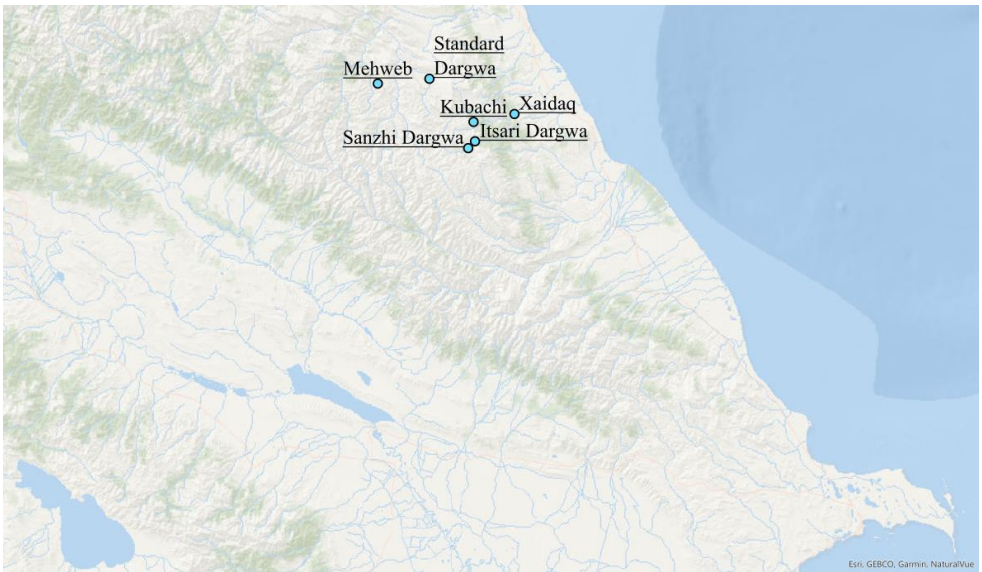
The third and fourth largest languages are Rutul and Aghul, spoken by approximately 33,000 speakers each (Federal State Statistic Service 2020), both to the west of the Lezgian language area. Both Rutul and Aghul are official languages of Dagestan (Dobrushina, Daniel & Koryakov 2020). The fifth Lezgic language with official status in Dagestan is Tsakhur, which is mainly spoken to the west of Rutul in north-western Azerbaijan and along the Dagestanian border by more than 24,000 speakers (State Statistical Committee 2019; Federal State Statistic Service 2020). Two of the remaining Lezgic languages are spoken entirely in Azerbaijan,

i.e. Kryts or Kryz with estimates ranging from 300 to maximally 2,000 speakers (Authier 2009; State Statistical Committee 2019) and the severely endangered Budukh with approximately 200 speakers (Dobrushina, Daniel & Koryakov 2020).

The Udi language has historically been spoken in north-western Azerbaijan but, due to the pervasive Armenian-Azerbaijani conflict, the Udis, who are traditionally Christians, have largely emigrated since the 1980s to Russia, with the exception of the Udi villages Nic in Azerbaijan and Zinobiani in Georgia (Alekseev et al. 2008: 5-6). Udi is spoken by more than 5,000 speakers, of which the majority live in Azerbaijan (State Statistical Committee 2019; Federal State Statistic Service 2020). The relatively recent discovery of numerous early Christian texts in Caucasian Albanian (or Aghwan) suggests that it is a relative and a potential ancestor of modern Udi (Gippert et al. 2008). The last Lezgic language is Archi, which is spoken to the northwest of the rest of the group by approximately 1,200 speakers in Archib and surrounding villages in Dagestan (Chumakina 2020). As Archi is spoken in an area completely surrounded by Avar and Lak, Archis have traditionally been proficient in at least Avar but also Lak, and since Archi is not recognised by Dagestani authorities, Archi children are taught Avar in school (Chumakina 2020). Archi is amongst the most well-described Lezgic languages, and it is known for its rich consonant inventory (Chumakina, Bond & Corbett 2016).

The Dargic branch

The third largest Dagestanian branch is the Dargic languages, spoken by more than 580,000 speakers in eastern Dagestan (Federal State Statistic Service 2020). Standard Dargwa or Literary Dargwa, is one of the official languages of Dagestan, and although the standard is based on Northern Dargwa, closest to Aqusha and Urakhi Dargwa, it is used as the written language for all Dargins (Dobrushina, Daniel & Koryakov 2020: 31). Dargwa was written in the Arabic script until 1928, when a Latin orthography was adopted, and like most other Dagestanian languages it adopted a Cyrillic orthography in 1938 (Forker 2020b: 8). Before the adoption of Standard Dargwa the lingua franca in the Dargic speaking areas of Dagestan was Kumyk (Forker 2020b: 9). Dargwa has since the Soviet era been officially and conventionally treated as one language (Sumbatova 2020), and the number of Dargic languages and their exact internal classification is not universally agreed upon. According to some linguists, e.g. Koryakov, there could be as many as 17 individual Dargic languages (Koryakov 2002).



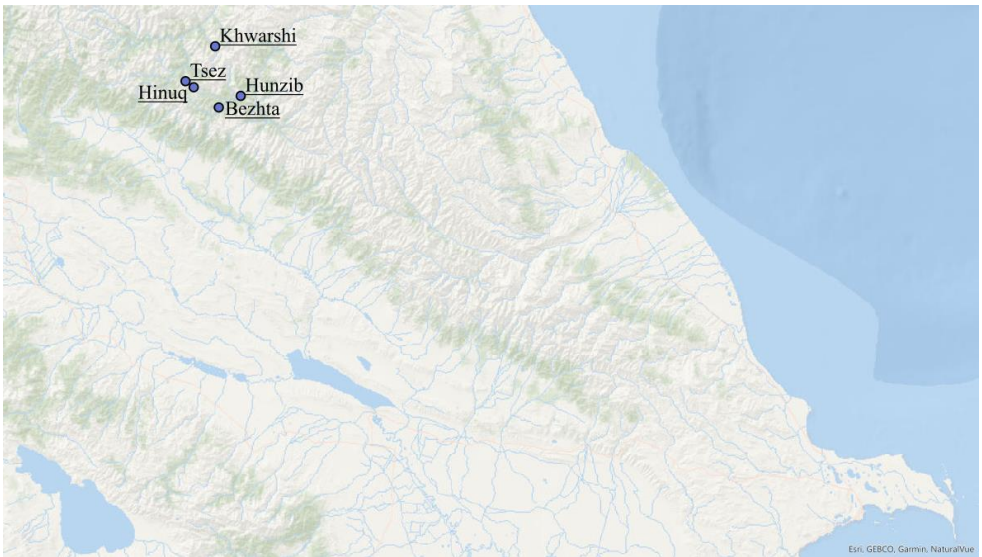
Map 3.7: All Dargic languages included in the data, represented by coordinates.

The Dargic languages are typically divided into a northern group and a southern group, but the overall internal classification is unclear (Ganenkov & Maisak 2020: 89). The northern group comprises, e.g. Northern Dargwa, i.e. the basis of Standard Dargwa, which has approximately 133,000 speakers, Muira with 34,500 speakers and Mehweb (or Megeb) with approximately 800-900 speakers (Dobrushina 2019a). The southern group comprises, e.g. the closely related Itsari Dargwa and Sanzhi Dargwa, which are included in this thesis and they have together 2,000

speakers (Dobrushina, Daniel & Koryakov 2020). Xaidaq (or Kaytag), Kubachi and Chirag are generally classified as separate Dargic sub-branches (Ganenkov & Maisak 2020). The Mehwebs live surrounded by Avar and Lak communities as they are geographically separated from the other Dargins, and Mehweb children are taught Avar and not Standard Dargwa in school (Dobrushina 2019a).

The Tsezic branch

The Tsezic (or Didoic) branch contains five relatively small languages that are spoken in the most mountainous regions of western Dagestan. Tsezic is, together with their northern neighbours the Andic languages, one of two branches not having any language with the status of official written language of Dagestan (Dobrushina, Daniel & Koryakov 2020), as Avar has been the historical lingua franca of the Tsezic peoples (Nichols 2020: 183).



Map 3.8: All Tsezic languages included in the data, represented by coordinates.

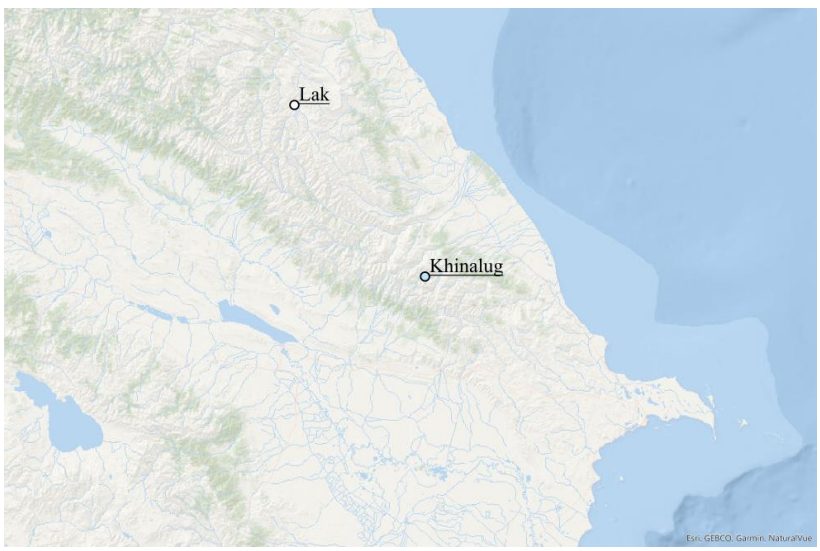
Tsez or Dido is the largest Tsezic language with between 12,300 and 17,000 speakers (Dobrushina, Daniel & Koryakov 2020; Federal State Statistic Service 2020) and it is spoken in westernmost Dagestan on the border to Georgia. The second largest Tsezic language is likely Khwarshi which is spoken to the north of Tsez, but the estimated number of speakers varies from approximately 3,300 to 8,500 speakers (Khalilova 2009; Federal State Statistic Service 2020). The closely related languages Bezhta, with more than 8,000 speakers, and Hunzib, with almost 3,500 speakers (Federal State Statistic Service 2020), are spoken to the southeast of

Tsez and in a few communities in north-eastern Georgia (Isakov & Xalilov 2012: 10; Comrie, Xalilov & Xalilova 2015: 16-17).

The last Tsezic language is Hinuq (or Hinukh) with about 600 speakers (Forker 2013; Federal State Statistic Service 2020). The Tsezic languages are generally not written and Tsez has the official status of an unwritten language in Dagestan, as Avar has historically been the main lingua franca and language of instruction for the Tsezic communities (Bokarev 1959). The Tsezic languages have also been in close contact with Georgian, due to their geographical location and economic ties, for millennia (Comrie & Khalilov 2009: 418). Most Tsezic communities have therefore had some proficiency in Georgian up until the fall of the Soviet Union, as the Georgian-Russian conflicts have largely limited Georgian-Dagestanian contacts (Comrie & Khalilov 2009: 418; Isakov & Xalilov 2012: 12; Forker 2013). It is worth mentioning that there are also Tsezic speaking communities in central and northern Dagestan as a result of Soviet re-localisation programmes in the 1960s and 1970s (Forker 2013: 6).

Lak

The Lak language forms its own branch of the Dagestanian languages, and is one of the official languages of Dagestan (Friedman 2020) with approximately 144,000 speakers (Federal State Statistic Service 2020). Lak is spoken between Avar and the Dargic languages in southern Dagestan, and it was described by Uslar already in the late 19th century (Uslar 1890). Lak is renowned for having one of the largest case systems in the world (Friedman 1992).



Map 3.9: The location of Lak and Khinalug, represented by coordinates.

Lak was first written in the Arabic script and the first attested extant text in Lak is from the 16th century, although the earliest attestation of written Lak is from the 10th century (Friedman 2020: 202). The Arabic script was used up until the adoption of a Latin orthography in 1928, which was subsequently used until the adoption of its current Cyrillic orthography in 1938 (Žirkov 1955: 5). Lak was an important lingua franca in central Dagestan before the Soviet period (Friedman 2020).

Khinalug

The last Dagestanian language is Khinalug, which also forms a branch of its own, and it is spoken by around 2,300 speakers in the village Xınalıq in northern Azerbaijan (Rind-Pawlowski 2023). Khinalug has previously been considered a part of the Lezgic branch, but it is today generally classified as belonging to a separate branch (Khvtisiashvili 2013), as the similarities might rather be due to Lezgic influences since Khinalug is surrounded by Lezgic languages. The lingua franca of the Khinalugs is Azerbaijani however, and almost all speakers of Khinalug also speak Azerbaijani (Khvtisiashvili 2013: 13; Kibrik 1994b: 369). Khinalug is generally not written and it was not until the 1980s that the first texts in Khinalug were produced in a Cyrillic orthography by local poet Rahim Alhas (Dobrushina, Daniel & Koryakov 2020: 44), which was later followed by a slightly different Cyrillic orthography by Ganieva (2002). Kibrik and Rind-Pawlowski have since been involved in introducing a Latin orthography that is closer to the Azerbaijani orthography (Khvtisiashvili 2013).

3.2.3. Northwest Caucasian languages

The Northwest Caucasian (NWC) or Abkhaz-Adyghe language family is a group of at least four extant languages, historically spoken along the north-eastern shores of the Black Sea extending into the western North Caucasus in a continuous area from the Kuban river and the Sea of Azov to Chechnya and Ossetia (Kuipers 1960; Smeets 1984; Colarusso 1992). The Northwest Caucasian languages were severely affected by the Russian conquest in the 19th century, as large numbers were massacred and forcibly displaced from their former homelands (Henze 1992: 96; Kuipers 1960). An estimated 1.2 to 1.5 million Caucasian refugees were driven out of the Northwest Caucasus in the 1860's and 1870's as a result of the Russian conquest, of which many perished en route (Henze 1992: 96-97; Coene 2010: 128).

Today, the Northwest Caucasian languages are therefore significantly decimated in their original homelands and only spoken in four non-contiguous areas in the three North Caucasian republics of Adygea, Kabardino-Balkaria, Karachay-Cherkessia and in the *de facto* independent republic of Abkhazia (Arkadiev & Lander 2020). There is however still an important Northwest Caucasian diaspora as Circassians, Ubykhs, Abkhazians and Abazins were expelled *en masse* in the 19th century to the then Ottoman Empire (Colarusso 1992; Henze 1992; Kumakhov & Vamling 2009: 19), resulting in Northwest Caucasian languages still being spoken in Turkey, Syria, Jordan and Israel (Arkadiev & Lander 2020). The Northwest Caucasian languages are conventionally divided into two main branches, the Circassian branch and the Abkhaz-Abaza branch, where the extinct Ubykh forms a separate branch as its exact position within the family is disputed (Fenwick 2011: 9).

Map 3.10: All Northwest Caucasian languages included in the data, represented by coordinates. The coordinates for Abzakh Adyghe, Shapsug Adyghe and Ubykh reflect their historical language areas.

The Circassian branch

The Circassian branch comprises the two closely related languages Kabardian and Adyghe, which are sometimes described as one macro-language, and they are the most spoken languages of the family by far. Kabardian, East Circassian or Kabardino-Cherkess is spoken by 515,000 to 600,000 speakers in Russia alone (Arkadiev & Lander 2020: 370; Federal State Statistic Service 2020) and as much as a third of all Kabardian speakers live in Turkey (Applebaum 2013: 5), while the number of ethnic Kabardians is even higher in Turkey than in Russia (Arkadiev & Lander 2020). Kabardian is an official language of the republics of Kabardino-Balkaria and Karachay-Cherkessia (Arkadiev & Lander 2020).

Adyghe or West Circassian is spoken by between 114,000 and 117,500 speakers in Russia alone (Arkadiev & Lander 2020: 370; Federal State Statistic Service 2020), while there are no reliable figures for the number of Adyghe speakers in the diaspora, which are primarily scattered across Turkey, Jordan and Syria (Jaimoukha 2001: 101-112). There are also two Circassian villages in Israel, primarily of Shapsug Adyghe descent with a population of approximately 3,400 people (Jaimoukha 2001: 114). Adyghe or Standard Adyghe is an official language in the Republic of Adygea (Arkadiev & Lander 2020). Adyghe is the most heterogeneous language of the family and there are four major dialects that persist to this day, namely Abzakh or Abadzekh Adyghe, Shapsug Adyghe, Temirgoy Adyghe and Bzhedug Adyghe (Smeets 1984: 41; Arkadiev & Lander 2020). The Temirgoy dialect is the basis of Standard Adyghe (Rogava & Keraševa 1966), and together with Bzhedug Adyghe are the two most important dialects in Adygea (Smeets 1984: 55). The Abzakh Adyghe and Shapsug Adyghe are today more or less only spoken in the diaspora, and they are the two main Adyghe dialects spoken in Turkey (Smeets 1984: 51).

The Abkhaz-Abaza branch

The Abkhaz-Abaza branch consists of the two closely related languages Abkhaz and Abaza. Abkhaz is the largest of the two and is spoken by somewhere between 129,000 and 190,000 speakers, primarily in Abkhazia (O'Herin 2020). Abkhaz is the official language of Abkhazia, and it has a strong written tradition, as there is an important body of Abkhaz-language literature and Abkhaz-language media (O'Herin 2020). The modern history of written Abkhaz is a multitude of different orthographies, as Uslar proposed the first Cyrillic orthography in 1862, but it was followed by a new Cyrillic orthography by Machavariani and Gulia in 1892, which in turn became widely used in Abkhaz schools in Chochua's revised version from 1909 (Chirikba 2003a: 15). In the late 1920s, first Marr and later Jakovlev introduced their own Latin orthographies, which were replaced by a Georgian orthography in 1938, used until Chochua's Cyrillic orthography was reintroduced in 1954 (Chirikba 2003a: 15).

Due to the mass exodus of almost 60% of all Abkhazians after the Russian conquest in 1864 (Chirikba 2003a: 6), there are even larger Abkhazians populations

in the diaspora of Turkey and Syria where there are numerous Abkhaz villages, but many of these communities are today fully Turkish- or Arabic-speaking (Chirikba 2003a; O'Herin 2020). The Abkhazians have been in longstanding contact with the neighbouring Megrelians, which has resulted in numerous Megrelian loanwords in Abkhaz (Chirikba 2003a: 14), and Abkhaz-Megrelian bilingualism has been reported in southern Abkhazia (Dobrushina, Daniel & Koryakov 2020). The early Soviet era was characterised by efforts to introduce Georgian as the main language of Abkhazia, e.g. by replacing Abkhaz with Georgian as the language of instruction in Abkhazian schools (Chirikba 2003a: 15). The lingua franca has been Russian since at least the 1950s (Chirikba 2003a: 15) and Russian is the main language of urban Abkhazia, especially among younger Abkhazians (Dobrushina, Daniel & Koryakov 2020).

The Abaza language is spoken by approximately 37,800 speakers, primarily in the Republic of Karachay-Cherkessia, and like the Abkhazians there has also been an important Abaza diaspora in Turkey since the 1870s (O'Herin 2002; O'Herin 2020). Abaza is one of the official languages of Karachay-Cherkessia and it is taught as an elective subject in schools (O'Herin 2020). The first efforts to write in Abaza were made in the Arabic script in the late 19th and early 20th centuries, and in 1932 a Latin orthography was adopted, which was later replaced by the current Cyrillic orthography in 1938 (Tabulova 1976: 9).

Ubykh

The last Northwest Caucasian language of Ubykh is perhaps the most famous, as it prominently had one of the largest consonant inventories ever recorded outside Africa (Fenwick 2011: 17). Ubykh was historically spoken in the area of modern Sochi (Kumakhov & Vamling 2009: 19), but as a result of the Ubykhs' fierce resistance to the Russian invasion, the entire Ubykh nation was deported to the Ottoman Empire in 1864 (Fenwick 2011: 12). The Ubykhs relocated to modern day western Anatolia where they formed several villages up until the end of the 20th century. The last known speaker of Ubykh, Tevfik Esenç, died in 1992 (Fenwick 2011: 11). The exact relationship between Ubykh and the other Northwest Caucasian languages is not agreed upon, and Chirikba argues that it is a transitional branch between Circassian and Abkhaz-Abaza (Chirikba 1996: 7-8).

3.2.4. Indo-European languages in the Caucasus

Indo-European languages have been spoken in the South Caucasus for millennia, but it is difficult to determine when they first arrived in the region. Armenian and various Iranian languages have been present in the Caucasus since at least classical antiquity (Coene 2010: 93-94; Belayev 2020). Russian has however been the most important lingua franca and most spoken Indo-European language in the region since the late 19th century (Chirikba 2008: 30), with more than one million reporting Russian as their first language in the South Caucasus and the republics of the North Caucasus (Dobrushina, Daniel & Koryakov 2020; Federal State Statistic Service 2020).

Map 3.11: All Indo-European languages included in the data, represented by coordinates.

Armenian

The second largest Indo-European language in the Caucasus after Russian is Eastern Armenian, which together with Western Armenian is one of the two extant main varieties of Modern Armenian. Eastern Armenian is spoken by approximately 4 million speakers, primarily in Armenia, but also in the previously *de facto* independent Nagorno-Karabakh region of Azerbaijan, Georgia, Abkhazia, Iran and in the Krasnodar krai in Russia (Dum-Tragut 2009; Belayev 2020: 575). Eastern

Armenian is the official language of Armenia, and has a long-standing literary tradition reaching back to the translation of the Bible into Classical Armenian in the 5th century AD (Dum-Tragut 2009). Classical Armenian, i.e. Grabar, is still used as the liturgical language of the Armenian Apostolic Church, and it was used as the literary language for both Eastern and Western Armenian until the 18th century (Dum-Tragut 2009).

The historical Armenian language area was drastically decimated in the early 20th century as a result of the Armenian genocide in the aftermath of the dissolution of the Ottoman Empire (Belayev 2020), which has essentially displaced the entire Western Armenian population. Western Armenian is now more or less only spoken in the diaspora and by a minority of the small remaining Armenian community in Turkey (Ajello 1997: 197). There is however one Western Armenian dialect that is still widely spoken in the Caucasus, Homshetsma, which is spoken by Hamshen Armenians along the Black Sea coast in north-eastern Turkey and Abkhazia (Vaux 2007).

The Iranian languages

The contemporary Iranian languages of the Caucasus belong to three fairly distantly related branches, North-East Iranian, North-West Iranian and South-West Iranian (Belyaev 2020). The North-East Iranian language of Ossetic (or Ossetian) is the most important Iranian language of the region and it is spoken by approximately 500,000 speakers in the North Caucasian republic of North Ossetia-Alania and in the *de facto* independent republic of South Ossetia (Belayev 2020). Ossetic is likely a descendant of the poorly attested Alanic language[6], and there is important dialectal differences between the two main dialects of Iron and Digor as they are not mutually intelligible (Erschler 2020: 641). Ossetic is the official language of both North Ossetia-Alania and South Ossetia, but the main language of instruction in Ossetian schools is Russian, and communication between Digor and Iron speakers is carried out in Russian (Erschler 2020: 642).

The first attested texts in modern Ossetic were written in various Cyrillic and Georgian orthographies from 1798 and onwards (Bagaev 1965: 10). In the late 19th century a Cyrillic orthography was adopted (Belyaev 2020: 575), which was later followed by the introduction of a Latin orthography in 1923 (Erschler 2020). In 1938, two separate orthographies emerged, as a Cyrillic orthography was introduced in North Ossetia and a Georgian orthography was introduced in South Ossetia, a situation which lasted until 1954, when the Cyrillic orthography was adopted in South Ossetia as well (Erschler 2020).

The North-West Iranian language of Talysh or Talyshi is spoken on the edge of the South Caucasus along the Caspian coast (Stilo 2015). Talysh is mainly spoken in Iran but there is an important Northern Talysh community in southern Azerbaijan,

[6] Alanic has been attested in inscriptions from the 10th-12th century and a manuscript from the 14th or 15th century, all in the Greek alphabet (Erschler 2020).

and the estimated total number of Talysh speakers varies considerably, from official Azerbaijani figures of approximately 43,000 speakers (State Statistical Committee 2019), to 200,000 (Belyaev 2020: 575) and possibly more than 500,000 speakers in Azerbaijan alone (Stilo 2015: 415). The Talysh autonomist movement proclaimed the short-lived Talysh-Mughan Autonomous Republic in 1993, which possibly affects the official figures, as this is still a politically sensitive issue in Azerbaijan (Coene 2010: 161-162).

Talysh is interesting from a linguistic point of view, as it likely descends from the North-West Iranian languages that were spoken in pre-Turkic Azerbaijan, i.e. Median and Parthian (Schulze 2000; Stilo 2015). Talysh is generally not a written language, but a Latin orthography was introduced in the late 1920s (Asatrian & Borjian 2005), which was later abandoned in 1935 for political reasons (Schulze 2000: 7). During this brief period the language was used in schools and a number of texts and newspapers were published in Talysh (Asatrian & Borjian 2005; Schulze 2000). Talysh is officially recognised as a language in Azerbaijan (Stilo 2015: 412) and efforts have been made to reintroduce the Latin orthography (Schulze 2000) with sporadic publications in Talysh in the post-Soviet era (Asatrian & Borjian 2005). There are also Northern Kurdish or Kurmanji-speaking communities scattered across the South Caucasus, with as many as 60,000 speakers, with an important Yezidi population in Armenia (Dobrushina, Daniel & Koryakov 2020: 32).

The third Iranian branch of South-West Iranian languages in the Caucasus has two members, Juhuri (or Judeo-Tat) and Tat (or Muslim Tat). These two closely related languages are spoken by descendants of Persian migrations in late antiquity (Belyaev 2020: 574), and the Tat languages were once spoken in urban settlements of southern Dagestan and northern Azerbaijan (Borjian 2015). The Jewish Tats, historically known as Mountain Jews, are today mainly found in Israel due to mass emigration (Belyaev 2020: 575), but there is still an important Jewish Tat community in Qırmızı Qəsəbə in Azerbaijan, where their language Juhuri is still spoken (Borjian 2015: 269). The Muslim Tat community is declining due to assimilation, and their language has between 9,000 and 20,000 speakers, primarily in Azerbaijan (State Statistical Committee 2019; Belyaev 2020).

Greek

Greek-speaking communities have also been present in the Caucasus since at least the 1820s, particularly in southern and western Georgia but also in Abkhazia (Loladze 2016: 178). The majority of these Greek-speaking communities have historically spoken Pontic Greek, which was previously also widely spoken in the region of Pontus in modern day Turkey (Loladze 2016: 178). However, the vast majority of these Pontic Greeks have migrated to Greece or Russia since the fall of the Soviet Union (Loladze 2016: 176; Dobrushina, Daniel & Koryakov 2020: 33), and the latest Georgian census reported only around 5,500 Greeks in Georgia (National Statistics Office of Georgia 2014).

3.2.5. Turkic languages in the Caucasus

Map 3.12: All Turkic languages included in the data, represented by coordinates.

The Oghuz languages

The Turkic languages of the Caucasus belong to two different branches, Oghuz and Kipchak. Oghuz is by far the largest of the two and contains the closely related languages Azerbaijani and Turkish. Azerbaijani or Azeri has historically been an important lingua franca in the Caucasus, and it was a court language of the Safavid Empire from the 16th century onwards (Ragagnin 2022) and did not lose this status in Iran until the fall of the Qajar dynasty in 1925 (Bulut 2022: 288). The Russian conquest of the South Caucasus in the mid-19th century changed the situation, but Azerbaijani remained an important lingua franca particularly in southern Dagestan up until the Soviet era (Dobrushina, Daniel & Koryakov 2020: 55). Azerbaijani is conventionally divided into North and South Azerbaijani along the 19th century border between the Russian Empire and the Persian Empire after the treaty of Gülistan in 1813 (Forsyth 2013: 307-308). North Azerbaijani is the official language of the Republic of Azerbaijan, and it is spoken by more than 10 million speakers, primarily in Azerbaijan, eastern Georgia and southern Dagestan (National Statistics Office of Georgia 2014; State Statistical Committee 2019; Federal State Statistic Service 2020). South Azerbaijani is spoken in north-western Iran, i.e. Iranian

Azerbaijan, and the number of speakers of South Azerbaijani is uncertain, but assessed to be more than 12 million (Bulut 2022: 289).

Before the partition of Azerbaijan in the 19th century Azerbaijani was universally written in the Arabic script, which is still used for South Azerbaijani (Bulut 2022), and there is an important body of Azerbaijani literature and poetry from as early as the 15th century (Johanson 2022b: 86). The Soviet era ended this as North Azerbaijani transitioned to a Latin orthography in 1929, which was later replaced by a Cyrillic orthography in 1939, only to revert to a Turkish-inspired orthography after the Republic of Azerbaijan regained its independence in the 1990s (Ragagnin 2022). Turkish was historically spoken by large numbers of Meskhetian Turks in south-western Georgia (Dobrushina, Daniel & Koryakov 2020: 33), which changed in 1944 when they were deported *en masse* to Central Asia (Kazenin 2020: 79), while most Meskhetian Turks today reside outside of Georgia in various Turkic-speaking post-Soviet republics and Turkey (Boeschoten 2022: 4). There is also a Greek community speaking Caucasian Urum in the region of Kvemo Kartli in southern Georgia (Loladze 2016: 178). Caucasian Urum is a variety of Anatolian Turkish (Höfler et al. 2016: 172), and should not be confused with the Kipchak language[7] Crimean Urum, primarily spoken around Mariupol in Ukraine (Smolina 2008: 8).

The Kipchak languages

The second branch of Turkic languages in the Caucasus are the three Kipchak languages Kumyk, Karachay-Balkar and Nogai. Kumyk is the largest of these with more than 520,000 speakers primarily residing in the lowlands of central Dagestan (Federal State Statistic Service 2020). Kumyk is an official language of Dagestan, and Kumyks are the third largest ethnic group in the republic after Avars and Dargins (Kazenin 2020: 68). Kumyk has historically also been an important lingua franca for various ethnic groups in central Dagestan (Dobrushina, Daniel & Koryakov 2020; Forker 2020b). Karachay-Balkar is the second largest Kipchak language of the Caucasus, as it is spoken by more than 340,000 speakers (Federal State Statistic Service 2020) and it is an official language of Kabardino-Balkaria and Karachay-Cherkessia (Dobrushina, Daniel & Koryakov 2020). Karachay-Balkar consists of the two closely related dialects Karachay and Balkar, which share one written standard. Parts of the Karachay community fled to the Ottoman Empire after the Russian conquest in the 19th century and there is still a small Karachay community in Turkey today (Seegmiller 1996). Karachays and Balkars were mass-deported by Soviet authorities in the 1940s to Central Asia, but the majority returned when they were allowed to repatriate in 1957 (Kazenin 2020: 78-80).

The last Kipchak language spoken in the Caucasus is Nogai or Noghay, which is most closely related to Kazakh and Karakalpak (Karakoç 2022). The Nogai Horde ruled the steppes north of the Caspian Sea between the Volga and the Ural during

[7] Crimean Urum has elements of both Kipchak and Oghuz however (Smolina 2008: 8).

the 15th and 16th centuries (Karakoç 2005), but they later split up into the Lesser Nogai Horde and the Greater Nogai Horde. The Greater Horde migrated southwards to settle along the Terek River in the northeast Caucasus, and the Lesser Horde migrated westwards to settle east of the Crimean Khanate by the Sea of Azov and north of the Kuban river (Forsyth 2013: 239). The Lesser Horde was weakened by the Russian annexation of the Crimean Khanate in 1783, and the Nogais of the Kuban plain were almost completely exterminated by Russian massacres in the late 18th century (Forsyth 2013: 246), while the extant Nogai community are descendants of the Greater Horde. Nogai is today spoken by between 77,000 and 104,000 speakers in Dagestan, Chechnya, Karachay-Cherkessia and the Stavropol region (Federal State Statistic Service 2020; Boeschoten 2022: 6), and it is an official language of Dagestan and Karachay-Cherkessia (Dobrushina, Daniel & Koryakov 2020: 40).

4.Methodology

4.1. Affixal data

The Caucaffix data set[8] is a compilation of 11,016 nominal and verbal affixes from 56 languages from the five language families spoken in the Caucasus. These affixes have been coded for affix type, i.e. prefix, suffix, infix, circumfix and transfix, grammatical function and phonological form. The phonological form of each affix has been coded according to its phonotactics, i.e. syllable structure, and thereafter its consonants and vowels. Allomorphs and allophones were included whenever they were mentioned in the grammatical descriptions, in order to capture all available variation, as segmental *morphologically conditioned phonology* is a well-known phenomenon in, e.g. Turkish (Inkelas 2011: 69).

I decided not to collect any prosodic information, as my focus has been solely segmental, which is worth problematising (Kiparsky 2018: 57). The interaction between affixation and, e.g. accent and stress patterns is undeniably central in many languages (Inkelas 2011: 70) and even more so with morphology that is only distinguished on a tonal level, such as Hausa imperatives (Inkelas 2011: 97). The decision not to consider this was mainly based on the nature of the available literature, as it does not always offer satisfactory prosodic descriptions (Borise 2020: 758), and since affixes seldom map exclusively to certain suprasegmental processes (Inkelas 2011: 76). I have also encountered morphologically motivated partial reduplication, which has been extensively discussed within the frameworks of, e.g. Prosodic Morphology and Optimality Theory (Inkelas 2011: 93), but this could not be stored in the data as the phonological form of these reduplications is fully reliant on the lexical context.

4.1.1. Data structure

The data are structured according to the principle that each allophone of an affix with a particular grammatical function constitutes one entry which in turn is connected to 91 variables. These 91 variables can be divided into three categories: 12 metadata variables, 33 grammatical variables and 46 phonological variables. The first variable is a unique six-digit identification number where the first digit

[8] https://zenodo.org/records/13902383

indicates language family, the second digit indicates language family sub-branch and the third digit is unique for the individual language, e.g. Georgian has identification numbers beginning with 100XXX, Megrelian has 101XXX and Chechen has 200XXX. The second variable is the phonological representation of the specific affix, which in itself is mainly used for referential purposes. Each affix is thereafter attributed to its language and the two metadata variables of associated language family branch and language family.

The two subsequent variables are the most fundamental as they encode fix type, i.e. suffix, prefix, infix, circumfix or transfix, and word class, i.e. if the affix is nominal or verbal. The variables *Fix type* and *Word class* are followed by 31 grammatical variables, which are presented in the section Grammatical categories below. These are followed by the variable *Phonotactics*, which encodes the syllable structure of the affix, and the 44 phonological variables. The phonological variables encode the consonantal and vocalic segments of each affix. Each affix can consist of up to five consonants and five vowels, encoded as C_1, C_2, C_3, C_4, C_5, V_1, V_2, V_3, V_4 and V_5. The consonants were coded according to place of articulation, manner of articulation and voicing, i.e. C_1 *Place*, C_1 *Manner*, C_1 *Voicing*, etc., thereby making each consonant a combination of three variables, which I henceforth refer to as the *trivariate consonant*.

Secondary features such as aspiration, labialisation and consonant length were also included in the data but not as separate variables, as this was only encoded by the IPA symbol. The reason for not including secondary features as separate variables was largely due to presence of consonants with multiple secondary features in the Caucasus, e.g. [kːʷ] and [kʷʰ], which would require encoding e.g. aspiration, labialisation and consonant length as separate variables, thus leading to at least a *hexavariate* consonant model. An alternative would be to encode multiple features as one value, which is not optimal from a comparative perspective.

The vowels were coded according to vowel height, vowel backness and rounding, thereby generating three variables for each vowel, which subsequently forms the *trivariate vowel*. The last seven columns contain metadata in the form of a *Comments* field in which potentially relevant information is added that cannot be captured by the other variables, and three possible slots for sources and the page on which the affix was found.

4.1.2. Languages

The aim of the affixal data set was to compile affixes from all endemic languages of the Caucasus, which presents an obvious question, i.e. what qualifies as an endemic language of the Caucasus? This is clearly a matter of discussion, but I could distinguish at least 65 languages or sufficiently divergent language varieties, of which 56 of these were sufficiently described to constitute a representative language-specific affix inventory (see table 4.1 for further details).

Table 4.1: The 56 languages of the Caucaffix data set according to genealogy.

Language family	Family branch	Language (ISO-639-3)	Language family	Family branch	Language (ISO-639-3)
Nakh-Dagestanian	Nakh	Chechen (che)	Kartvelian	Karto-Zan	Georgian (kat)
		Ingush (inh)			Old Georgian (oge)
		Bats (bbl)			Megrelian (xmf)
	Avar-Andic	Avar (ava)			Laz (lzz)
		Andi (ani)		Svan	Svan (sva)
		Tindi (tin)	Northwest Caucasian	Circassian	Kabardian (kbd)
		Bagvalal (kva)			Adyghe (ady)
		Chamalal (cji)			Abzakh Adyghe (-)
		Karata (kpt)			Shapsug Adyghe (-)
		Akhvakh (akv)		Abkhaz-Abaza	Abkhaz (abk)
		Ghodoberi (gdo)			Abaza (abq)
	Lezgic	Lezgian (lez)		Ubykh	Ubykh (uby)
		Tabasaran (tab)	Indo-European	Armenian	Eastern Armenian (hye)
		Rutul (rut)			Classical Armenian (xcl)
		Aghul (agx)		Iranian	Iron Ossetic (oss)
		Tsakhur (tkr)			Talysh (tly)
		Udi (udi)			Tat (ttt)
		Kryts (kry)			Juhuri (jdt)
		Budukh (bdk)	Turkic	Oghuz	North Azerbaijani (azj)
		Archi (aqc)			South Azerbaijani (azb)
	Dargic	Standard Dargwa (dar)		Kipchak	Karachay-Balkar (krc)
		Xaidaq (xdq)			Kumyk (kum)
		Kubachi (ugh)			Nogai (nog)
		Itsari Dargwa (-)			
		Sanzhi Dargwa (-)			
		Mehweb (-)			
	Tsezic	Tsez (ddo)			
		Khwarshi (khv)			
		Hinuq (gin)			
		Bezhta (kap)			
		Hunzib (huz)			
	Lak	Lak (lbe)			
	Khinalug	Khinalug (kjj)			

Since the compiled data comprised approximately 85% of the postulated goal of 65 endemic languages, the outcome of the data collection ought to be satisfactory as this was more than initially excepted. Among the languages that were not included in the data set are Botlikh, Caucasian Albanian or Aghwan, Sirhwa Dargwa, Chirag, Western Armenian, Homshetsi Armenian, Digor Ossetic and Caucasian Greek due to insufficient or inaccessible documentation. Tanti Dargwa and Pontic Greek could

have been included as it has been thoroughly described recently by Sumbatova & Lander (2014) and Berikashvili (2017), but it was excluded due to time limitations. The original aim was to also include Russian, Turkish and Persian as all three have been important linguae francae, but they fail to qualify as endemic languages of the Caucasus. Russian, Turkish and Persian would have been interesting from a language contact perspective and as non-Caucasian reference languages, which is why I have included them in the lexical data.

4.1.3. Grammatical categories

The selection of grammatical categories was primarily based on the criterion that each category had to be relevant for at least two language families, which yielded an initial list of 19 grammatical functions: case, number, person, tense, mood, aspect, voice, version, finiteness, noun class, causatives, local case, definiteness, transitivity, negation, possessives, intentionality, reflexivity and reciprocity. Case, number, tense and mood are universally expressed by affixation in all Caucasian language families (Klimov 1999: 16; Alekseev 1999: 157-159; Šagirov 1999: 82; Boeder 2005: 12-13). Person marking was included as it is found in all Kartvelian languages (Klimov 1999: 16), all Northwest Caucasian languages (Šagirov 1999: 82) and in some Nakh-Dagestanian languages such as Bats (Dešerieva 1999a: 169), Lak and Dargwa (Alekseev 1999: 159). Causative affixes were also included as they are found in Kartvelian (Tuite 1998: 91), Northwest Caucasian (Šagirov 1999: 82) in some Nakh-Dagestanian languages (Nichols 1996: 45-46) and Turkic (Johanson 2022a: 37).

Local cases and various locative affixes (e.g. affixed adpositions) were also included, as they are widespread in the region (Dešerieva 1999b: 178-181; Xajdakov 1999: 351-353; Boeder 2005: 14-16; Kumakhov & Vamling 2009: 30-31). Aspect and voice are found in Kartvelian (Klimov 1999: 16; Boeder 2005: 14-16) and in some Nakh-Dagestanian languages (Dešerieva 1999b: 178-181; Xajdakov 1999: 351-353). Applicatives, which have been traditionally referred to as ‘version’ within Caucasiologist linguistics, are shared by Kartvelian (Klimov 1999: 16; Tuite 2024) and Northwest Caucasian (Šagirov 1999: 82; Arkadiev, Lander & Bagirokova 2024). Noun classes or gender are found in most Nakh-Dagestanian languages (Alekseev 1999: 158-159) and in Abkhaz and Abaza (O’Herin 2020: 458). Intentionality marking is not widely found in the Caucasus, but is sporadically found in Nakh-Dagestanian (Daniel 2001b: 217; Sumbatova & Mutalov 2003: 105; Forker 2013: 199-202) and Northwest Caucasian (Smeets 1984: 260; Chiribka 2003a: 38; Fenwick 2011: 118-121).

The initial list of 19 grammatical functions was later expanded to also include preverbs, converbs, evidentiality and copular affixes. Preverbs have various grammatical functions in the languages of the Caucasus although they primarily encode the spatial direction and orientation of actions, and are as such present in Kartvelian (Boeder 2005: 32-34), Northwest Caucasian (Arkadiev & Lander 2020:

412), many Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 109) and in the Iranian languages of the Caucasus (Belayev 2020: 606). Converbs are widely used in Turkic (Johanson 2022a: 40) and Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 115), while the situation is somewhat more complicated for the three other language families, as particularly the Northwest Caucasian languages have affixes with nearly identical functions. Evidentiality marking is widespread in Nakh-Dagestanian languages (Forker 2018a) and it is also found in Kartvelian (Harris 1991a: 51) and in some Northwest Caucasian languages (Chirikba 2003a: 47; O'Herin 2020: 478). Copular affixes are highly common in Turkic languages, and they occur in some Nakh-Dagestanian languages such as Kryts (Authier 2009: 111) and Khinalug (Kibrik 1994b: 388), Eastern Armenian (Dum-Tragut 2009: 215) and Juhuri (Authier 2012: 133). Certain categories were also revised and split, e.g. the marking of person and number was divided into subject and object person-marking, finiteness was revised to only encode non-finite forms, and reflexivity and reciprocity were merged into one category.

Table 4.2: An overview of grammatical functions that are generally expressed by means of affixation in the five language families of the Caucasus.

	Kartvelian	**Nakh-Dagestanian**	**Northwest Caucasian**	**Caucasian Indo-European**	**Caucasian Turkic**
Case	Yes	Yes	Yes	Yes	Yes
Number	Yes	Yes	Yes	Yes	Yes
Tense	Yes	Yes	Yes	Yes	Yes
Mood	Yes	Yes	Yes	Yes	Yes
Aspect	Yes	Yes	Yes	Yes	Yes
Non-finite forms	Yes	Yes	Yes	Yes	Yes
Local case	Partly	Yes	No	No	No
Negation	Yes	Yes	Yes	Yes	Yes
Causative	Yes	Yes	Yes	Partly	Yes
Person (subject)	Yes	Partly	Yes	Yes	Yes
Person (object)	Yes	Partly	Yes	No	No
Converbs	No	Yes	Yes	Partly	Yes
Noun class/Gender	No	Yes	Partly	No	No
Evidentiality	Yes	Yes	Partly	Partly	Partly
Preverbs	Yes	Yes	Yes	Partly	No
Voice	Yes	No	No	Partly	Yes
Possessives	No	No	Yes	Partly	Yes
Reflexivity	No	No	Yes	Partly	Yes
Definiteness	No	No	Yes	Yes	No
Version/Applicatives	Yes	No	Yes	No	No
Copular affixes	Partly	Partly	No	Yes	Yes

The most important revision to the grammatical categories was to encode direction and orientation for local cases and spatial preverbs separately, as this is

conventionally done for Nakh-Dagestanian local cases and it enables a more fine-grained comparison. This is not typically done for spatial preverbs in the Caucasus, as the typology of spatial preverbs in these languages is generally not systematically described and a common terminology appears to be lacking. I have therefore applied the same descriptive approach and terminology for the local cases as for the spatial preverbs. I also decided to include various affixes and particles indicating clausal relationships, e.g. conjunctive affixes and particles, of which many are not technically affixes (see chapter 2.2 for further discussion).

4.1.4. Phonological form

The trivariate consonant

The consonantal data rest almost entirely on the concept of the '3-term label of phonetics' (Maddieson 1984: 163) which I choose to describe as the *trivariate consonant*, i.e. the categorisation of consonant segments as a combination of *place of articulation*, *manner of articulation* and *voicing*. This is an effective method of differentiating consonant segments while also enabling the analyses of larger groupings. The trivariate consonant does not consider any other phonological aspects of a specific consonant, which excludes information regarding secondary articulation such as aspiration, palatalisation, labialisation, consonant quantity and pharyngealisation. This approach can divide the places and manners of articulation into more or less fine-grained groupings. I tried to keep the groupings as fine-grained as possible, but there are some apparent exceptions that are worth discussing, e.g. the combined categories of alveolars/dentals and taps/flaps/trills (cf. table 4.3).

Table 4.3: Predefined values of place of articulation, manner of articulation and voicing.

Places of articulation:	Manners of articulation:	Voicing:
• *Alveolar/Dental* • *Alveolo-palatal* • *Bilabial* • *Epiglottal* • *Glottal* • *Labio-dental* • *Labio-velar* • *Palatal* • *Pharyngeal* • *Postalveolar* • *Retroflex* • *Uvular* • *Velar*	• *Affricate* • *Approximant* • *Ejective* • *Ejective affricate* • *Ejective fricative* • *Fricative* • *Lateral affricate* • *Lateral approximant* • *Lateral ejective affricate* • *Lateral ejective fricative* • *Lateral fricative* • *Nasal* • *Stop/Plosive* • *Tap/Flap/Trill*	• *Voiced* • *Voiceless*

The most important grouping that could have been divided further is the combined alveolar/dental place of articulation, thus following Maddieson (1984). The rationale to merge these two places of articulation was in part pragmatic, as alveolar and dental consonants are often not satisfactorily differentiated in older phonological descriptions. The merge was also theoretically motivated, as few languages outside Australia make a phonemic distinction between e.g. alveolar and dental stops (Maddieson 1984: 32-33), and most languages tend to have a set of consonants, i.e. /d/, /t/, /n/, /l/, that are realised as either alveolars or dentals. In Georgian a distinction is made between dental stops/ejectives and alveolar affricates/fricatives/sonorants (Shosted & Chikovani 2006), thereby presenting a more or less complementary distribution between the dental and alveolar consonants, which gives further support for merging alveolars and dentals into one category. A similar tendency motivated a merge of all taps, flaps and trills, i.e. rhotics (Gordon 2016: 49), into a joint category which has been labelled 'tap/trill' in the data. There are however numerous languages in the world that differentiate taps and trills phonemically, e.g. Castilian Spanish (Martínez-Celdrán et al. 2003), Albanian (Camaj 1984) and Eastern Armenian (Dum-Tragut 2009), but when no such distinction is made the phonological descriptions do not always specify whether the rhotic segments are taps, flaps or trills.

The trivariate vowel

The concept of the *trivariate vowel* is similar to the trivariate consonant and is based on the three variables of *height*, *backness* and *lip-rounding* employed by the UPSID survey (Gordon 2016: 50-51). The trivariate vowel does not include information regarding vowel length or whether the vowel is nasal or pharyngealised. The variables of the trivariate vowel are clearly less varied than for the trivariate consonant, as the number of possible vowel segments is naturally more restricted. The exact description of different vowel heights is not a trivial matter, and I have tried to keep the categories as close to the specific vowel as possible (cf. table 4.4).

Table 4.4: Predefined values of vowel height, vowel backness and vowel rounding.

Vowel height:	Vowel backness:	Vowel rounding:
• *Close* • *Near-close* • *Close-mid* • *Mid* • *Open-mid* • *Near-open* • *Open*	• *Front* • *Central* • *Back*	• *Rounded* • *Unrounded*

It would have been possible to follow Maddieson and group the vowel heights into only three values, i.e. close (close and near-close), mid (close-mid, mid and open-mid) and open (open and near-open) (Maddieson 1984: 167-168), but due to the structure of the data these groupings can still be merged for further analyses.

Phonotactics

The phonotactics of each affix has been coded as a separate variable, in order to analyse syllable structures and so that initial consonants could be distinguished from final consonants. The C_1 of a prefix with the syllable structure CV is hence initial and the C_1 of a suffix with the syllable structure VC is final. It is therefore possible to filter all affixes according to affix type and their phonotactics, thereby selecting only the consonants that are in either initial or final position.

4.2. Lexical data

Since this thesis assumes that morphology and the lexicon operate on parallel albeit different levels, it is important to compare affixal and lexical data. In order to investigate how the phonological results of the Caucaffix data relate to the lexicon in these languages, the second lexical data set of Caucalex[9] was compiled. The Caucalex data set contains 21,586 lexical items from 52 languages and the data were primarily collected from the Diachronic Atlas of Comparative Linguistics (DiACL) database (Carling 2024) and the Intercontinental Dictionary Series (IDS) (Key & Comrie 2023). The DiACL data consist primarily of culture words and Swadesh list items from 37 languages, totalling 7,423 lexical items after recent loanwords and duplicates were removed. Of these lexical items 5,540 were nouns and only 1,001 items were verbs. In order to increase the validity of the lexical data, I decided to collect semantically comparable data from the IDS chapters 'The physical world', 'Animals', 'The body', 'Food and drink', 'Clothing and grooming', 'The house', 'Agriculture and vegetation', 'Basic actions and technology' and 'Motion', which added 5,190 nouns and 5,929 verbs from 38 languages to the lexical data. Recent loanwords were again removed, while a small number of old loanwords were kept.

As the Kartvelian languages form a small yet well-studied language family, 315 Kartvelian nouns and 1,462 Kartvelian verbs were added from Fähnrich's Kartvelian etymological dictionary (2007). The Northwest Caucasian languages were the most difficult to find lexical data for, so data for these languages were also added from NorthEuralex (Dellert et al 2019), Kumakhov & Vamling (2009), Fenwick (2011) and O'Herin (2020). The Northwest Caucasian verbs are particularly difficult to collect data for, as verb roots typically only consist of one syllable or consonant, and the Northwest Caucasian lexical data would optimally need more data to increase the validity of the phonological analyses. Karachay-Balkar is not included in neither DiACL nor IDS, so the lexical data for Karachay-Balkar were collected from Savelyev & Robbeets (2020).

[9] https://zenodo.org/records/13903040

The second phase of the lexical data compilation was to remove all affixes from the lexical data, as verb forms typically were given in the infinitive or with various preverbs or derivational affixes. As the affixal data were already collected, many of the affixes in the lexical data could be detected and removed by comparing them to the affix data. Certain issues were language family specific, as the noun class affixes needed to be removed from the Nakh-Dagestanian verbs. The pronouns and adjectives present in the original DiACL data were not removed, even though they could potentially affect the phonological outcomes. A number of languages were not included in the analyses as the amount lexical data was too low, which was true for Turkish and Talysh, while Abaza has the smallest amount of lexical data of the languages included, as it only has 142 lexical items. Further lexical data could have been collected from dictionaries, but collecting data directly from dictionaries would take considerably longer time. Due to the relatively large amount of lexical data available in DiACL and IDS, I made the assessment that the 21,586 lexical items would be sufficient as a lexical control data set, since the mean number of words per language is 415, which is substantially more data than e.g. a Swadesh list would yield.

The lexical data were coded according to the same principles as the Caucaffix data, while including ten consonant slots instead of the five consonant slots for the affix data, to accommodate lexicon that contains more than ten consonants. The vowels of the lexical data were not coded, as this was not deemed feasible both due to the limited time frame of the thesis but also due to the low relevance for the research questions of this thesis, since the research questions primarily relate to consonant phonemes.

4.2.1. Languages

As it was difficult to find lexical data for all 56 languages of the Caucaffix, the languages of the Caucalex data differ somewhat from the affixal data. A few languages that were not included in the Caucaffix were however included in the Caucalex, i.e. Russian, Persian and Turkish, in order to use these languages as non-Caucasian reference points. The Turkish and Talysh lexical data were later excluded from the analyses, as the amount of data was too low

Table 4.5: The 52 languages of the Caucalex data set according to genealogy.

Language family	Family branch	Language
Nakh-Dagestanian	Nakh	Chechen (che)
		Ingush (inh)
		Bats (bbl)
	Avar-Andic	Avar (ava)
		Andi (ani)
		Tindi (tin)
		Bagvalal (kva)
		Chamalal (cji)
		Karata (kpt)
		Akhvakh (akv)
		Ghodoberi (gdo)
	Lezgic	Lezgian (lez)
		Tabasaran (tab)
		Rutul (rut)
		Aghul (agx)
		Tsakhur (tkr)
		Udi (udi)
		Kryts (kry)
		Budukh (bdk)
		Archi (aqc)
	Dargic	Standard Dargwa (dar)
		Xaidaq (xdq)
		Kubachi (ugh)
		Itsari Dargwa (-)
		Mehweb (-)
	Tsezic	Tsez (ddo)
		Khwarshi (khv)
		Hinuq (gin)
		Bezhta (kap)
		Hunzib (huz)
	Lak	Lak (lbe)
	Khinalug	Khinalug (kjj)
Kartvelian	Karto-Zan	Georgian (kat)
		Old Georgian (oge)
		Megrelian (xmf)
		Laz (lzz)
	Svan	Svan (sva)
Northwest Caucasian	Circassian	Kabardian (kbd)
		Adyghe (ady)
	Abkhaz-Abaza	Abkhaz (abk)
		Abaza (abq)
	Ubykh	Ubykh (uby)
Indo-European	Armenian	Eastern Armenian (hye)
		Classical Armenian (xcl)
	Iranian	Iron Ossetic (oss)
		Persian (pes)
		Juhuri (jdt)
	Slavic	Russian (rus)
Turkic	Oghuz	North Azerbaijani (azj)
	Kipchak	Karachay-Balkar (krc)
		Kumyk (kum)
		Nogai (nog)

4.3. Phonological coding issues

4.3.1. Uvular fricatives in Kartvelian

The presence of uvular fricatives in the Kartvelian languages is somewhat problematic, as different descriptions present different pictures. Chirikba uses it as an argument in his defence of the Caucasian Sprachbund, where he claims that [ʁ] and [χ] are present in Proto-Kartvelian, Svan and Old Georgian (Chirikba (2008: 47). Shanidze describes these consonants as velar in Old Georgian (Shanidze 1976), while Tuite describes them as uvular (Tuite 2008a: 148). Fähnrich presents a middleway, as he labels them as postvelar while he labels [q] and [q'] as 'pharyngeal', i.e. uvular (Fähnrich 2012: 51). The situation in Svan is similar, as Schmidt and Tuite describe these fricatives as uvular or 'dorso-uvular' (Schmidt 1991: 476; Tuite 1997: 7), while Gudjedjiani and Palmaitis describe them as 'dorso-velar', i.e. velar (Gudjedjiani & Palmaitis 1986: 17).

In modern Georgian these sounds are variably described as the velar fricatives [x] and [ɣ] (Butskhrikidze 2002: 87), as '(dorso-)velar/uvular' (Hewitt 1995: 19; Cherchi 1999: 1-2) and as postvelar (Aronson 1990: 16). Bolkvadze and Kiziria confusingly label them as velar but use the IPA symbols for the uvular fricatives (Bolkvadze & Kiziria 2023: 16). Shanidze describes them together with the velar stops rather than with the uvular ejective [q'], but he also adds that these fricatives are pronounced further back than the velar stops (Shanidze 1980: 15). Testelets analyse these fricatives as uvular, but also adds that they can be realised as velars (Testelets 2020: 497). Megrelian also has similar issues, as Rostovtsev-Popiel describe the corresponding fricatives as velar while using the symbols [χ] and [ʁ], and Harris labels them as dorso-velar (Harris 1991b: 317). In Laz, Holisky labels them as uvular (Holisky 1991: 399), Kutscher, Öztürk and Pöchtrager as velar (Kutscher 2001: 13; Öztürk & Pöchtrager 2011: 8) and Lacroix describes them as 'velar or uvular' (Lacroix 2009; Lacroix 2018).

This presents a difficulty in deciding how to encode these phonemes in the data sets. For the affixal data the problem is negligible for Georgian, Megrelian and Laz, as I have only found one affix in these languages containing [x] or [χ]. In Old Georgian and Svan, I have decided to follow Tuite and analyse these phonemes as uvulars.

4.3.2. Labio-velar approximants and labialisation in Kartvelian

The presence of labio-velar approximants [w] in the Kartvelian languages is worth discussing, as it appears to largely be in complementary distribution with the voiced labio-dental fricative [v] in all living Kartvelian languages (Butskhrikidze 2002: 88; Rostovtsev-Popiel 2020: 535; Harris 1991b: 317; Tuite 1997: 6), and the descriptions differ between [v] in Arhavi Laz and [w] in Pazar Laz (Öztürk &

Pöchtrager 2011; Lacroix 2018). The situation in modern Georgian and Old Georgian is further complicated by [w] being an allophone of /v/ when preceded by an obstruent (Butskhrikidze 2002: 88; Tuite 2008a: 148).[10] Although this phenomenon is generally indicated in the transcription of Old Georgian with /w/, it is not conventionally indicated in modern Georgian, even though consonant clusters with a final /v/ are realised as labialisation according to Butskhrikidze (2002: 88). This is also relevant for assessing the presence of labialisation in the Kartvelian languages, as none of the Kartvelian consonant inventories include labialised consonants, although Gudjedjiani & Palmaitis (1986: 20) describe [tʷ] and [lʷ] while also acknowledging that final /w/ in consonant clusters cannot be considered a 'sonant'. It is therefore worth discussing labialisation as a possible phonemic phenomenon in at least Georgian, Old Georgian and Svan.

4.3.3. Northwest Caucasian fricatives and affricates

The description of the Northwest Caucasian fricatives and affricates is not trivial, which is stressed by Arkadiev & Lander (2020: 372-373). It is notoriously difficult to determine the exact place of articulation for sibilants in these languages, as they can either be alveolar, dental, alveolo-palatal or retroflex, and the traditional descriptions often use unsatisfactory labels such as 'hissing-hushing' (Arkadiev & Lander 2020: 373). Hewitt (2004) and Fenwick (2011) have convincingly transcribed these segments into IPA, while Hewitt's Abkhaz self-tutor largely transcribes the Abkhaz phonemes into an IPA-hybrid, as he uses the Caucasianist degree symbol for labialisations, e.g. /tɕ°/ for /tɕʷ/ (Hewitt 2010: 10). Arkadiev & Lander mention palatalised postalveolars, e.g. /ʃʲ/, which should most likely be analysed as alveolo-palatal /ɕ/, while they use the same alveolo-palatal symbol /ɕ/ for alveolar fricatives (Arkadiev & Lander 2020: 372), which adds to the confusion.

A distinction is also found between laminal and apical sibilants in some descriptions, e.g. Paris (1989) and Colarusso (1992), which further complicates the description of fricatives and affricates in these languages. I have tried to map these descriptions onto the places of articulation found in the IPA, but it has been remarkably difficult to differentiate alveolo-palatal from retroflex sibilants, which in some cases might have affected the outcomes.

4.3.4. Ejectives vs. glottalised consonants in Northwest Caucasian

There are some inconsistencies in how an important category of Northwest Caucasian consonant phonemes are described, i.e. ejectives or glottalised/glottalic consonants. This issue is not unique to the Northwest Caucasian languages, as

[10] Another allophone of /v/ is [ɸ] followed by voiceless consonants (Butskhrikidze 2002: 88).

similar inconsistent descriptions are found for other language families, e.g. the Quechuan, Aymaran and Uru-Chipaya languages (Adelaar & Muysken 2004). Kuipers (1960: 19) and Smeets (1984: 74) use the term glottalic consonants while using the ejective apostrophe to describe them, and Paris (1989: 157) uses the term glottalised in a similar manner. Konuk (2022) even transcribes this category of glottalised phonemes with the IPA symbol ◌ˀ in Abzakh Adyghe (Konuk 2022: 63). I have analysed all consonant phonemes described as glottalic or glottalised as ejectives, since these labels appear to be synonymous and largely based on linguistic tradition rather than a meaningful phonological difference.

4.4. Analysis

The data were analysed in Spyder 5.5.1 (Python 3.9) and R (4.1.2) to concatenate, visualise and statistically test the data. The following Python packages were used: Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), pandas (McKinney 2010), scikit-learn (Pedregosa et al. 2011), SciPy (Virtanen et al. 2020) and seaborn (Waskom 2021). The data for the grammatical functions were concatenated in Spyder as a crosstab and then binarised, yielding binary data indicating the absence or presence of a certain grammatical function in each language. The binary grammatical data were used to determine the sum of all grammatical functions in each language for the calculation of the Pearson correlation coefficient (in Spyder), Spearman rank correlation coefficient (in Spyder), and the linear regression analysis (in R).

Principal Component Analysis (PCA), a linear dimensionality reduction technique that is conventionally used to explore and visualise multivariate data, was performed by means of scikit-learn in Spyder to generate the grammatical and phonological PCA plots in chapter 7. The results from the PCA plots were further analysed by implementing *k*-means clustering, which is an unsupervised machine learning technique to identify clusters of data objects, and it was calculated by using scikit-learn in Spyder. The appropriate number of clusters was evaluated by means of comparing results from the elbow method and the silhouette coefficient, also calculated by scikit-learn.

The affixal phonological data were concatenated in order to analyse all five consonant slots (C_1-C_5) in Spyder, and the concatenated phonological distributions of the various phonological variables were later equalised per language, as the occurrences of a certain phonological category were divided according to the sum of all occurrences in each language, yielding proportions instead of actual occurrences. All phonological data for a specific language therefore constitute 1, which is distributed across the actual occurrences as proportions, which was then used to calculate the mean proportion of all languages. If e.g. all affixes in a language are expressed by [n], the alveolar/dental nasals will have a proportion of 1, while a language where all affixes are equally divided between [n] and [d] the proportions will be 0.5 for [n] and 0.5 for [d], etc.

Equalising the data by language was motivated by the considerable differences between the languages in the number of affixes and consequently segments. If the data had not been equalised certain languages would skew the data, as e.g. the presence of vowel harmony in the Turkic languages mean that almost all affixes have minimally two or more forms. The equalised distributions were used to yield all affixal figures, including the affixal phonological PCAs, in chapter 7.

The distributional differences between the affixal and lexical data sets were tested by means of non-parametric Wilcoxon signed-rank tests. The wide format equalised affixal crosstabs were therefore melted in Spyder to produce long format affixal data for the Wilcoxon signed-rank tests (both Spyder and R) and the Shapiro-Wilk tests

(Spyder), where the long format affixal data were tested against the long format lexical data. The effect size of the Wilcoxon signed-rank tests was calculated as Pearson's *r* by dividing the *z*-value by the square root of the sample size (*N*) (Fritz, Morris & Richler 2012: 12). In order to counteract the multiple comparisons problem, the *p*-values of the Wilcoxon signed-rank tests were corrected using the Bonferroni correction.

The lexical phonological data were analysed in a similar way as the affixal data, while all ten consonant slots (C_1-C_{10}) were concatenated for the lexical data. The lexical data were equalised in the same way as the affixal data, and the equalised lexical distributions were used to yield all lexical figures and the lexical phonological PCAs in chapter 7. The wide format equalised lexical data were also melted in Spyder to produce the long format data used for the Wilcoxon signed-rank tests and Shapiro-Wilk tests mentioned above.

4.5. Glossing and transliteration

The affixes and their associated grammatical functions are exemplified with examples from the languages included in the data in chapters 5 and 6. I have decided to regloss examples when necessary, primarily to enable a coherent glossing but also to visualise when my analyses differ from previous research. Since all older sources and most Russian-languages sources lack glossing, I have tentatively glossed the examples from these sources to enable the inclusion of examples from all languages in the thesis. I have done this by searching through the available grammatical descriptions and dictionaries, while also using the affixal data set itself as a tool, since it already contains most if not all relevant affixes. I acknowledge that I am not an expert on these languages and I therefore stress that all glossings are based on the material that has been available to me, which means that some glossings are potentially problematic or even incorrect. As I have translated all Russian, German and French translations into English, I have included the original translations as footnotes to increase the transparency of the process. The glossing for each example will be categorised as either 'original glossing', 'gloss adapted from source', 'reglossed' or 'my glossing'. 'Original glossing' means that the example is given as it is glossed in the source, with only minor changes such as reglossing abbreviations, e.g. PRES to PRS. Examples categorised as 'gloss adapted from source' have smaller categorical changes, while 'reglossed' applies to examples that are glossed in the source, but they have been reanalysed either partially or considerably. The final category of 'my glossing' indicates that the example is unglossed in the source, which means that I have both segmented and glossed the example myself.

I have transcribed the examples according to the linguistic tradition of each language family to avoid diverging too far from previous descriptions. I have decided against transcribing all examples in IPA, as the material I have worked with is generally not detailed enough to enable a satisfactory phonematic representation in IPA for all languages, as e.g. secondary articulation is typically not indicated if it is not contrastive. Since this thesis deals with phonological typology, I have retransliterated certain consonant segments to disambiguate segments which are unsatisfactorily represented in the various traditions. The different representations of the velar fricatives [x] and [ɣ] and the uvular fricatives [χ] and [ʁ] have therefore been replaced in all language families, with the Kartvelian languages being a possible exception, cf. section 4.3.1. I have also generally retransliterated various lateral affricates and ejectives, e.g. [tɬ] and [tɬ'], labialised consonants, e.g. [χ^{w}] and [k^{w}'], and consonant lengthening, e.g. [k^{w}:] and [tɬ':] according to the principles of the IPA to avoid unnecessary confusion. The alveolar/dental and postalveolar voiced affricates [dz] and [dʒ] have also been consistently retransliterated as /dz/ and /dž/ to avoid the plethora of often confusing representations for these segments used within the various descriptive traditions.

Armenian and North Azerbaijani have been transliterated according to their own traditions, as Armenian has its own transliteration standard, and the North Azerbaijani Latin orthography has been used instead of the traditional Turkological transliteration. I have made three important digressions from the conventional Eastern Armenian transliteration however, as all aspirated consonants are transliterated with superscript [ʰ] instead of the diacritic /’/ to avoid potential confusion with the identical ejective marker [’], the synchronically confusing symbol /ł/ has been replaced by /ʁ/ since it represents a voiced uvular fricative (Dum-Tragut 2009: 13), and the alveolar/dental and postalveolar voiced affricates are transliterated as /dz/ and /dž/ instead of /j/ and /ǰ/.

I have decided to retransliterate the fricatives, affricates, ejective fricatives and ejective affricates in the Northwest Caucasian languages into IPA to avoid introducing the various representations used within, e.g. the Russian-language tradition (Arkadiev & Lander 2020: 373-377). The transcription of the Northwest Caucasian languages into IPA is not trivial, which is stressed by Arkadiev & Lander (2020: 372-373), and I acknowledge that I have made tentative transliterations and retransliterations based on the material I have available.

4.6. Maps

All maps were generated in ArcGIS Pro, where each language was given a coordinate based on its present or historical range. The base map used was NaturalVue, which is provided by Esri, GEBCO and Garmin. The Northwest Caucasian languages were therefore given coordinates reflecting their geographical distribution prior to the exodus caused by the Russian invasion in the 19th century. The coordinate symbols were later colour-coded according to language family. The maps do not include political borders as the map templates available in ArcGIS Pro do not properly reflect the current political landscape of the Caucasus.

5. Nominal affixation in the Caucasus

5.1. Case functions

Case is a category of nominal morphology that comprises a wide range of grammatical functions and it is one of the few categories found in all 56 languages of the affixal data set, and the data presented almost 30 case functions in these 56 languages. The exact definition and demarcation of cases are generally difficult, as one morpheme can often convey multiple case functions, e.g. the ergative/oblique case in Northwest Caucasian which encodes ergative, dative and genitive functions (Arkadiev & Lander 2020: 393). It is therefore important to differentiate descriptive cases from functional cases, as *case syncretism* is a well-known phenomenon where multiple case functions are expressed by one form (Baerman 2009). It also raises questions regarding the core function of a specific descriptive case, which is often more complicated than it might appear at first glance. One frequent example is the conflation of the dative and genitive case, e.g. in Classical and Eastern Armenian (Meillet 1936: 65; Dum-Tragut 2009: 83).

Should such a descriptive case primarily be analysed as a dative case, a genitive case or simply both? One possible approach is to carry out a diachronic analysis of the origin of the case, but this option is only available for a few languages in the region, as historical records of most Caucasian languages are scarce. I have instead chosen to apply a synchronic solution to this by coding these combined cases as having both a dative and a genitive function. A similar situation can be found in Kartvelian languages as the dative case is used to mark both the direct and indirect object in non-ergative constructions (Harris 1991b: 56-59; Hewitt 1995: 218), thus clearly having the function of an accusative case as well, cf. section 5.1.1.1.

Case should furthermore be positioned in relation to adpositions and adverbial derivation, as case functions can often be expressed either by affixes or adpositions. It is therefore relevant to discuss how to distinguish case affixes from affixed adpositions, as affixed adpositions are known to grammaticalise into case affixes (Heine 2009: 460-464; Lehmann 2015: 84-92) and Bickel and Nichols even describe the distinction between adpositions and case markers as being blurred in certain contexts (Bickel & Nichols 2007: 174). I have generally applied a functional approach as phonological affixes expressing a case function have been analysed as case, which therefore includes affixed adpositions in some languages as the distinction between affixed adpositions and case affixes is largely a matter of the

level of grammaticalisation. Case should be differentiated from adverbial derivation, as the latter by definition changes the word class of the noun, thus giving the derived adverb different morphosyntactic properties. Case forms and adpositional constructions typically behave morphosyntactically in a similar manner while adverbs do not, e.g. English *to our home-s* and Turkish *ev-ler-imiz-e* are grammatical while English **our home-s-ward* is ungrammatical. Cases thus often interact and combine with other nominal categories while adverbs do not.

The results for the case functions in figure 5.1 reveal that the dative and the genitive cases are the most common case functions expressed by affixation in the Caucasus. The uneven distribution between ergative and accusative languages in the Caucasus evidently skews the results, which likely explains why the dative and the genitive are the most common as these are shared by both ergative and accusative languages. These results do not code the multifunctional obliques in the Northwest Caucasian and Iranian languages as separate functions but simply as oblique, cf. section 5.1.1.1.

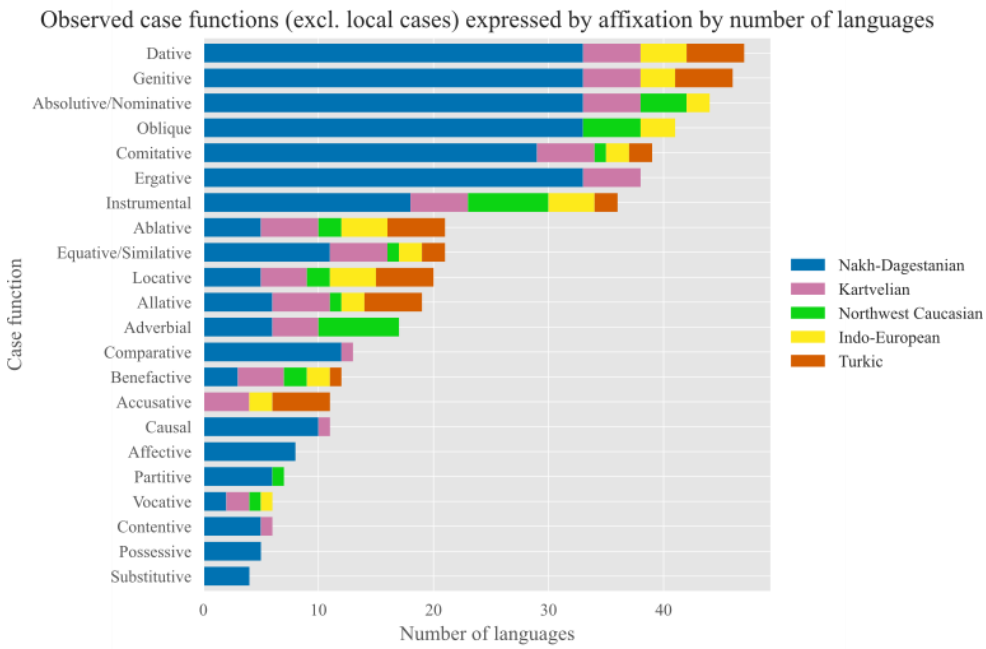


Figure 5.1: All observed case functions expressed through overt affixation in the data set in descending order according to number of languages.

Figure 5.2 below is an attempt to make the results somewhat more comparable by adding local cases by counting their essive, ablative/elative and lative affixes as instances of locative, ablative or allative. The most striking difference in the adjusted results in figure 5.2 is the prevalence of the local cases, which is not only caused by the Nakh-Dagestanian languages. The absolutive/nominative affixes are

furthermore mostly plural affixes, as the Nakh-Dagestanian languages generally have a formal distinction between absolutive and oblique plural affixes, while the Circassian languages and some Kartvelian languages such as Georgian and Old Georgian have explicit nominative/absolutive affixes (Tuite 1998: 50-51; Fähnrich 2012: 91; Arkadiev & Lander 2020: 393).

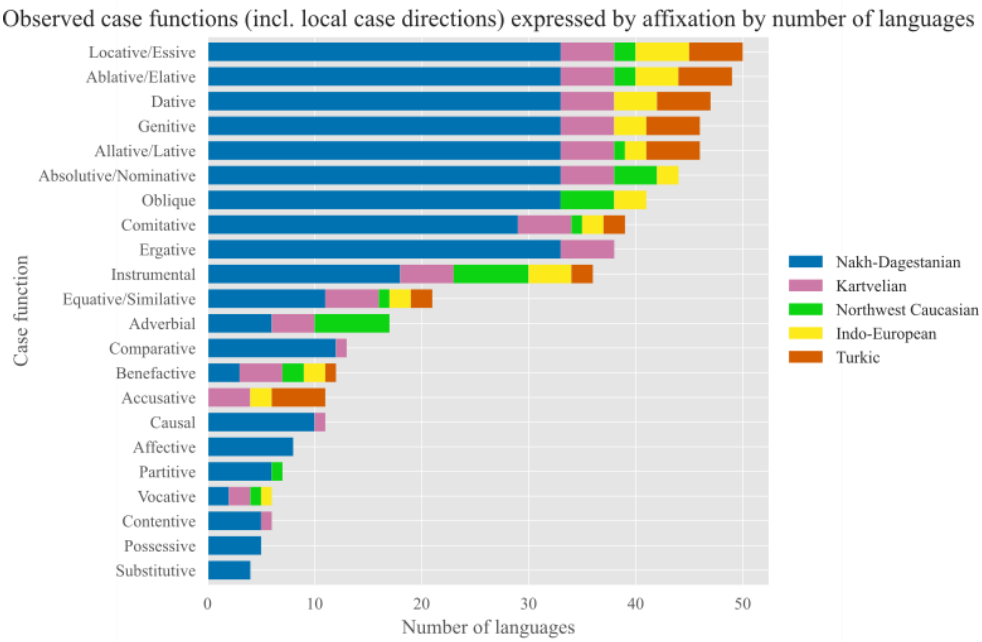


Figure 5.2: All observed case functions, including directional local case functions but excluding the multifunctional oblique cases, expressed through affixation in the data set according to number of languages.

It is important to remember that the results in figures 5.1 and 5.2 only concern affixes, which means that e.g. the nominative and absolutive cases are found in many more languages than the affix data will indicate and that the affixes are only nominal, which excludes pronominal affixes. Since I have not included zero-marked cases in my data, I can only discuss overtly expressed case functions in this thesis. These results rely on the assumption that case syncretism is present on a functional level even though there are no empirically observable traces of it, which is the general prerequisite to qualify as true case syncretism (Baerman 2009: 219). As some of the case functions mentioned in the results above require further explanation, the typology of the observed case functions in the data is presented below in order of their respective occurrence in the languages of the Caucasus.

5.1.1. Grammatical case functions

5.1.1.1. Core case functions

The Dative case function

The Dative case function is a core case function that generally marks the indirect object (Haspelmath 2009: 513), and is together with the genitive case the most common core case function in the data. The likely explanation is that it is a core case function that is valid for both accusative and ergative languages. The dative case can be found in all Kartvelian (Harris 1985: 72), all Nakh-Dagestanian (Ganekov & Maisak 2020: 102), all Turkic (Ragagnin 2022: 247; Dehghani 2000: 101; Berta & Csató 2022: 325; Karakoç 2022: 358) and all Indo-European languages of the Caucasus except Tat (Meillet 1936; Schulze 2000; Dum-Tragut 2009: 71-76; Authier 2012; Erschler 2020; Suleymanov 2020: 96-97).

(1) Georgian (Tuite 1998: 18) (gloss adapted from source)
švil-eb-ma *c'eril-i* *ga-Ø-u-gzavn-es* *mama-s*
child-PL-ERG letter-NOM PV-3.O-OV-send-AOR.3PL father-DAT
'The children sent a letter to (their) father.'

(2) Lezgian (Haspelmath 1993: 88) (gloss adapted from source)
ruš-a *gada-di-z* *cük* *ga-na*
girl-ERG boy-OBL-DAT flower give-AOR
'The girl gave a flower to the boy'

(3) North Azerbaijani (Širaliev 1971: 331) (my glossing)
uşaq *kağız-ı* *məşədi-yə* *ver-ir*
child letter-DEF.ACC Mashadi-DAT give-PRS.3SG
'The child gives the letter to Mashadi'[11]

(4) Eastern Armenian (Dum-Tragut 2009: 86) (gloss adapted from source, retransliterated)
dasaxos-ə *usanoʁ-i-n* *tv-ec^{h}* *girk^{h}-ə*
lecturer-DEF student-DAT-DEF give-AOR.3SG book-DEF
'The lecturer gave the book to the student'

Some languages such as Eastern Armenian have identical dative and genitive cases, making it theoretically problematic to label the suffix *–i* as simply a dative case. This is not a functional issue though, as Eastern Armenian consistently uses this

[11] Translated from Russian 'Паренек письмо Мешади отдает'.

case to indicate indirect objects. However, it would be inaccurate to claim that there is a dative case in the Northwest Caucasian languages, as these languages consistently employ the same affix to designate both ergatives, indirect objects and possessors (Rogava & Keraševa 1966: 65; Kumakhov & Vamling 2009: 22; Fenwick 2011: 33; Arkadiev & Lander 2020: 393-394), cf. the oblique case. This phenomenon is also similar to the situation in Kartvelian accusative constructions, where the dative case suffix is used for both the direct and indirect object (Hewitt 1995: 218).

(5) Temirgoy Adyghe (Arkadiev & Lander 2020: 431)
(original glossing, retransliterated)
tʃʼaɮe-m *pʂaʂe-m* *txəɬə-r*
boy-OBL girl-OBL book-ABS
Ø-Ø-r-j-e-tə
3SG.ABS-3SG.IO-DAT-3SG.ERG-DYN-give
'The boy is giving the book to the girl'

The Ergative case function

The Ergative case function is the third most common case, and it typically marks the agent in ergative constructions (Haspelmath 2009: 512). Case affixes indicating the ergative function are present in all languages of the three endemic language families (including the multifunctional ergative/oblique case in Northwest Caucasian), with the exceptions of Abkhaz and Abaza, as they lack core case affixes (Chirikba 2003a: 48; O'Herin 2020: 458). In Northern Talysh the oblique case covers both accusative and ergative functions (Schulze 2000: 17), but this is hardly surprising as identical split ergative patterns are found in other Northwestern Iranian languages, e.g. Kurmanji (Schulze 2000: 42; Haig 2018: 131).

(6) Georgian (Bolkvadze & Kiziria 2023: 29) (my glossing)
mama-m *bilet-eb-i* *uk've* *i-q'id-a*
father-ERG ticket-PL-NOM/ABS already SV-buy-AOR.3SG
'Father already bought the tickets'

(7) Chechen (Komen, Molochieva & Nichols 2020: 342) (original glossing)
beer-aš *šura* *molu*
child-ERG milk drink
'The child drinks milk'

(8) Adyghe (Kumakhov & Vamling 2009: 53)
(reglossed and retransliterated)
ʂʷəzə-m *sabəj-r* *Ø-j-e-hə*
woman-ERG/OBL child-ABS 3SG.ABS-3SG.ERG-DYN.PRS-carry
'The woman carries the child'

The Northwest Caucasian case systems are worth discussing, as they have traditionally been described as differentiating between the absolutive, ergative and oblique ergative cases (Kumakhov & Vamling 2009: 22), while the ergative and oblique ergative cases are identical. The same case thereby covers both an ergative function, indirect objects, adnominal possessors and locative and temporal adjuncts, which has led to a two-case distinction between the absolutive and oblique case instead (Arkadiev & Lander 2020: 393-394). I have therefore decided to analyse the Northwest Caucasian oblique case as a 'multifunctional' oblique case, by coding the various functions of the oblique case as separate functions. The rationale for this is that it will enable a phonological comparison of all case affixes expressing the same case function, cf. section 7.9.2, which would be impossible otherwise.

The Absolutive/Nominative case function

The Absolutive case function is the fourth most common case in the data, and it marks the subject in intransitive constructions and the patient in ergative transitive constructions (Haspelmath 2009: 512). The absolutive is typically a zero-marked case (Ganenkov & Maisak 2020: 101), which means that most ergative languages have an absolutive case but no absolutive case affixes. However, there are some languages in the Caucasus with overt absolutive affixes, e.g. most Karto-Zan languages (Harris 1991b: 326; Fähnrich 1994: 56; Hewitt 1995: 34) and the Circassian languages (Arkadiev & Lander 2020: 393).

The absolutive case is also implicitly marked in all Nakh-Dagestanian plural forms, as the final vowel of the plural affix typically changes from the absolutive to all other cases (with a few exceptions), e.g. Dargwa absolutive *ruz-bi* 'sisters' and *qul-ri* 'houses' become ergative *ruz-b-a-ni* and *qul-r-a-ni* (Musaev 2002: 60), etc. A similar pattern involving syncope of the vowel in the plural affix can be seen in Lezgian, e.g. absolutive *balk'an-ar* 'horses' becomes ergative *balk'an-r-i* (Haspelmath 1993: 75).

(9) Kabardian (Kumaxov 2006: 85) (my glossing)
ɬ'əʒə-m *žemə-r* *ʔeχʷe-m* *d-jə-xʷ-a-ɛ*
old.man-OBL cow-ABS pasture-OBL IN-3SG.ERG-drive-PFV-IND
'The old man drove the cow out (in)to the pasture'[12]

(10) Old Georgian (Tuite 2008a: 155) (gloss adapted from source)
mama-man *dze-y* *p'ov-a*
father-ERG son-ABS find-AOR.3SG
'The father found (his) son'

[12] Glossed and transliterated from *лIыжьым жэмыр Iэхъуэм дихуащ*, and translated from Russian '[С]тарик выгнал корову на пастьбу'.

The Nominative case indicates the subject in accusative constructions and it is often zero-marked (Haspelmath 2009: 512). Georgian, Old Georgian and Megrelian are some of the only languages in the Caucasus to have a distinct nominative affix (Harris 1991b: 326; Fähnrich 1994: 56; Hewitt 1995: 34), and the absence of nominative affixes in the remaining Kartvelian languages is mainly due to widespread apocope of the nominative in Svan (Tuite 1997: 16) and Laz case marking being fully ergative (Lacroix 2018: 852). Nominatives are completely absent in Nakh-Dagestanian and Northwest Caucasian, as these language families are fully ergative (Ganenkov & Maisak 2020: 102; Arkadiev & Lander 2020: 432). Classical Armenian exhibits the overt nominative plural suffix ending in *-kʿ* (Meillet 1936: 66-67; Clackson 2008: 132-133). The nominative plural suffix *–æ/-ɐ* in Ossetic is another potential candidate for overt nominative marking in the Caucasus (Abaev 1964: 19; Erschler 2020: 647).

(11) Classical Armenian (Clackson 2008: 140) (gloss adapted from source)
tes, *orpisi* *en* *kʿar-inkʿ=s*
see.AOR.IMP, what-sort be.PRS.3PL stone-NOM.PL=DEF
'Look, how wonderful the(se) stones are!'

(12) Iron Ossetic (Erschler 2020: 667) (gloss adapted from source)
χɐʃtɐg *qɐwu-t-ə* *irɐt-tɐ* *birɐ* *sɐr-ə*
nearby village-PL-OBL Ossetian-PL.NOM many live-PRS.3SG
'Many Ossetians live in nearby villages'

The Accusative case function

The Accusative case indicates the direct object or the patient in accusative constructions (Haspelmath 2009: 512), which means that is completely absent in fully ergative language families such as Nakh-Dagestanian and Northwest Caucasian. The accusative function is found in all Turkic and almost all Indo-European languages of the Caucasus, while it is absent in Eastern Armenian (Dum-Tragut 2009). The accusative case suffix only occurs with definite direct objects in Northern Azerbaijaini (Širaliev 1971: 45) and South Azerbaijani (Dehghani 2000: 100), which is also true for the syncretic accusative/dative or oblique case in Talysh (Schulze 2000: 17), Tat (Suleymanov 2020: 250) and Juhuri (Authier 2012: 47).

(13) South Azerbaijani (Dehghani 2000: 107) (reglossed)
Ali-nin *kitab-ı-nı* *oxu-du-m*
Ali-GEN book-3SG.POSS-ACC read-PST-1SG
'I read Ali's book'

The accusative and genitive case suffix *–nI* are identical in Kumyk and Karachay-Balkar (Abdullaeva et al. 2014: 183; Ulakov & Guseev 2016: 273; Berta & Csató 2022: 325), although the presence of the third person singular possessive suffix *–(s)I* indicates a case syncretism in Kumyk and Karachay, cf. table 5.1.

Table 5.1: Case syncretism of the accusative and genitive cases in Kumyk and Karachay.

	Kumyk (Abdullaeva et al. 2014: 135)	**Karachay** (Seegmiller 1996: 14)
Accusative/Genitive	*at-nï* horse-ACC/GEN	*ana-nï* mother-ACC/GEN
3SG Possessive Accusative	*at-ï-n* horse-3SG.POSS-ACC	*ana-sï-n* mother-3SG.POSS-ACC
3SG Possessive Genitive	*at-ï-nï* horse-3SG.POSS-GEN	*ana-sï-nï* mother-3SG.POSS-GEN

Berta & Csató give the specific genitive suffix *-nIŋ* in Balkar (Berta & Csató 2022: 325), but this suffix is completely absent in Filonenko's Balkar grammar (Filonenko 1940: 32-34). Iron Ossetic also has a similar pattern, where the definite accusative is identical to the genitive (Abaev 1964: 18; Bagaev 1965: 141-142). Classical Armenian used the prefix *z-*, i.e. the *nota accusativi*, to indicate definite direct objects (Schmitt 2007: 91).

(14) Classical Armenian (Schmitt 2007: 91) (my glossing)
gtin z-Mariam ew z-Yovsêp‘ ew z-manowk-n
find-AOR.3PL ACC-Mary and ACC-Joseph and ACC-child-DEF
'[the shepherds] found Mary, Joseph and the child'[13]

The presence of accusative cases in the Kartvelian language family is worth discussing, as Kartvelian languages are conventionally described as lacking an accusative case (Tuite 1998: 18; Testelets 2020: 502-503), although the direct and indirect objects are marked with identical suffixes in the accusative alignment of series I in Kartvelian (Tuite 1998: 18). This is true for all Kartvelian languages except Laz, as they all have split-ergative alignment (Fähnrich 1994: 66; Hewitt 1995: 218; Tuite 1997: 21; Harris 2009: 45-47; Rostovtsev-Popiel 2020: 541). These suffixes are conventionally described as being simply instances of the dative case, while the accusative function of the Kartvelian dative case was already acknowledged by Marr & Brière (1931), as they referred to the dative case in Georgian as 'datif-accusatif'. As this thesis has a functional approach, I have decided to analyse the Kartvelian datives as expressing both a dative and an accusative function.

[13] Translated from German 'fanden Maria und Joseph und das Kind'.

(15) Georgian (Harris 2009: 40) (my glossing)
Series I:

nino	*a-čven-eb-s*	*surat-eb-s*	*gia-s*
Nino.NOM	NV-shows-SM-3SG	picture-PL-DAT/ACC	Gia-DAT

'Nino is showing pictures to Gia'
Series II:

nino-m	*a-čven-a*	*surat-eb-i*	*gia-s*
Nino-ERG	NV-showed-AOR.3SG	picture-PL-NOM	Gia-DAT

'Nino showed the pictures to Gia'

The Oblique case function(s)

The fifth most common case is not a true case in the functional sense, as it is the Oblique case, which is a notoriously polysemous concept. The traditional Indo-European notion of the oblique case groups all non-nominative cases, as oblique Indo-European cases often share formal properties such as being formed by an oblique stem that is different from the nominative stem (Haspelmath 2009: 508). The same phenomenon is present in all Nakh-Dagestanian languages, where the oblique case suffixes act as morphological building blocks to construct most non-absolutive cases (Ganenkov & Maisak 2020: 101). The Nakh-Dagestanian oblique case affixes are sometimes identical with the ergative case (Ganenkov & Maisak 2020: 102), but there are numerous Nakh-Dagestanian inflectional paradigms with an oblique case which is separate from the ergative case, e.g. Bagvalal absolutive *misa* 'house' becomes ergative *mis-u-r* (house-OBL-ERG), genitive *mis-u-ɬ:* (house-OBL-GEN), inessive *mis-u-ni* (house-OBL-IN), etc. (Daniel 2001a: 144).

(16) Lak (Friedman 2020: 211) (gloss adapted from source)

t:u-l	*dus-na-l*	*t:u-l*	*lu*	*ka-ni-l*	
1SG.OBL-GEN	friend-OBL-GEN	1SG.OBL-GEN	book	hand-OBL-GEN	

la<w>s-unni
take<I>-TR.PFV.3SG
'My friend took my book with his hand'

(17) Standard Dargwa (Mutalov 2018: 59) (gloss adapted from source)

durħ-na-ni	*q'ac'*	*b-erk-un*
boy-OBL.PL-ERG	bread	NH.SG-eat.PFV-AOR

'The children ate bread'

(18) Hunzib (Van den Berg 1995: 46) (gloss adapted from source)

ož-di-l	*kid-bo-go-s*	*k'az*	*r-αhu-r*
boy-OBL-ERG	girl-OBL-AD-ELA	shawl(V)	V-take-PST

'The boy took the shawl from the girl'

The term ‘oblique’ case has a rather different meaning in two-dimensional case systems, as they are often described as having a two-way distinction between a nominative/absolutive case and an oblique case with miscellaneous functions, e.g. Circassian languages and Ubykh (Arkadiev & Lander 2020: 393-394) and Iranian languages (Stilo 2009: 700-703). The Circassian-Ubykh ‘oblique’ case covers the ergative function, indirect objects, possessors, postpositional objects and locative adjuncts (Arkadiev & Lander 2020: 393-394). In Iranian two-case systems the zero-marked form, labelled the *rectus* case in the Iranistic tradition (Durkin-Meisterernst 2014: 201), is contrasted with the ‘oblique’ case, which covers most non-nominative core case functions. The oblique case expresses definite accusative and genitive functions in Tat (Suleymanov 2020: 97), while in Northern Talysh the oblique conveys definite accusative, ergative and genitive case functions (Schulze 2000: 17).

(19) Adyghe (Arkadiev & Lander 2020: 409) (original glossing, retransliterated)
dwembajə-r *ɬ’ə-m* *qʷaɕʷe-m*
bison-ABS man-OBL boat-OBL
r-a-r-jə-ʁe-ʁe-wətɕʷ-a-ʁ
LOC-3PL.IO-DAT-3SG.ERG-CAUS-CAUS-stand.up-LAT-PST
‘The man ordered them to put [the] bison in the boat (lit. ‘made them make it stand there)’

(20) Northern Talysh (Schulze 2000: 30) (gloss adapted from source)
a. *palang-i* *vind-əš-e* *ba* *čayi* *lona*
leopard-OBL see.PST-3SG.A-AUX.3SG to 3SG.POSS cave
odam *da-šə-da*
man.ABS IN-go.PST-IPFV
‘The leopard saw that a man was entering his cave’
b. *mə* *pi-a* *čayi* *püst-i* *pegat-om*
1SG.OBL want.PST-PFV 3SG.POSS skin-OBL take.off-OPT.1SG.A
‘I wanted to take off his skin’

The Affective or Experiencer case function

The Affective case indicates the *experiencer* of certain sensory verbs (Haspelmath 2009: 514), e.g. ‘see’, ‘hear’, ‘understand’, ‘know’, ‘forget’, ‘think’, ‘find’ and ‘be able to’ (Daniel 2001b: 215), typically corresponding to experiencer datives or genitives in many other languages (Butt 2009: 31-33), e.g. the ‘dative-subject’ in Kartvelian languages (Tuite 1998: 26). Since the affective case is obligatory selected for sensory verbs in these languages, it should arguably be analysed as a core case. The affective case occurs in all Andic languages, i.e. Andi (Salimov 2010: 96), Tindi (Magomedova 2012: 86), Bagvalal (Daniel 2001a: 144), Chamalal (Bokarev 1949a: 46), Akhvakh (Magomedbekova 1967: 55), Ghodoberi (Kibrik 1996: 16; Saidova 2004; 81) and in the Lezgic language Tsakhur (Schulze 1997: 33; Kibrik & Testelets

1999: 54). Avar differs from the Andic languages as it uses either the dative or the superessive case to indicate the experiencer (Forker 2020b: 262-263). It is worth mentioning that the affective case suffix in Andi agrees with the noun class of the absolutive noun (Salimov 2010: 105). The multifunctional case suffix *–(a)x* in Chechen also has an affective function (Aliroev 1999: 58-59).

(21) Andi (Salimov 2010: 105) (my glossing)
imu-bo *qχ:'inkom* *haʁв-o*
father-AFFT.III bull.calf(III).ABS see-WPST
'Father saw the bull calf'[14]

(22) Ghodoberi (Kibrik 1996: 37) (original glossing)
ʕali-ra *haʔ-at-a-da* *mat'-u-tɬi* *hinu*
Ali-AFFT see-PRS-CVB-COP mirror-OBL-INTER inside
ži=w=da
self=M=EMPH
'Ali sees (lit. is seeing) himself in the mirror'

(23) Tsakhur (Schulze 1997: 55) (original glossing)
ʧodʒ-us *aliwʃes-da* *balkan* *za-k'le* *he:ge*
brother-DAT buy.INF-PTCP.III horse(III).ABS 1SG-AFFT show.IMP
'Show me the horse that you will buy for your brother'

[14] Glossed and transliterated from *иму-бо къинком гьакъго*, and translated from Russian '[О]тец бычка видел'.

5.1.1.2. Non-core case functions

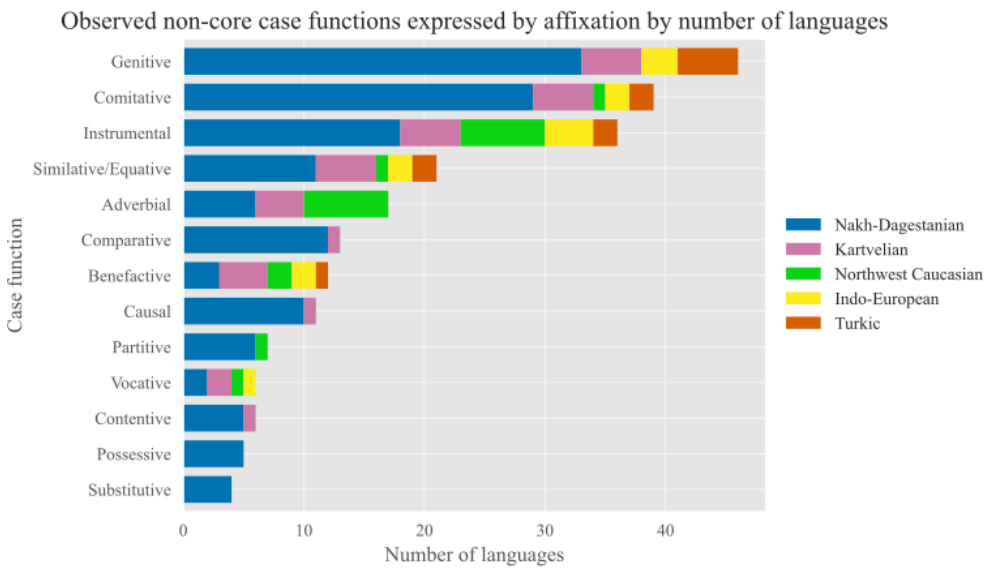


Figure 5.3: Most common observed non-core case functions (excl. local cases) expressed by affixation in the data in descending order according to number of languages.

The Genitive case function

The Genitive case function is the most common non-core case function, and is expressed by affixation in 53 languages of the Caucasus (incl. languages with the multifunctional oblique case). The basic function of the genitive is to mark the possessor in possessive constructions (Haspelmath 2009: 513). Genitive case affixes are found in all Kartvelian languages (Harris 1985: 72), all Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 102) and all Turkic languages in the Caucasus (Ragagnin 2022: 247; Berta & Csató 2022: 325; Karakoç 2022: 358).

Northwest Caucasian languages do not have a specific genitive case as they instead mark the possessed in possessive constructions by means of personal possessive prefixes. The Circassian languages and Ubykh do however mark the possessor as well in these constructions in the so-called oblique case mentioned above, as in e.g. Kabardian (27) and Ubykh (28), which thereby functions as a genitive suffix in possessive constructions. A genitive case affix is also absent in Juhuri as it is either marked by the possessed using an ezafe construction, a periphrastic possessive construction or no marking at all, although the Azerbaijani genitive suffix is sometimes used (Authier 2012: : 82-83).

(24) Georgian (Harris 2009: 115) (my glossing)
zurab-is *da-s* *č'or-av-en*
Zurab-GEN sister-DAT gossip-SM-3PL
'They are gossiping about Zurab's sister'

(25) Avar (Alekseev et al. 2012: 169) (my glossing)
daʁistan-aɬ-ul *muʕr-ul,* *nuž-e-je* *r-ecː,* *barkala*
Dagestan-OBL-GEN mountain-PL, 2PL-OBL-DAT PL-praise, thanks
'The mountains of Dagestan, praise to you, thank you'[15]

(26) Kumyk (Abdullaeva et al. 2014: 259) (my glossing)
o *kitap* *men-i* *ini-m-ni* *kitab-ï*
this book 1SG-GEN younger.brother-1SG.POSS-GEN book-3SG.POSS
'This book is my younger brother's book'[16]

(27) Kabardian (Kumakhov & Vamling 2009: 25) (gloss adapted from source)
fəzə-m *jə-wəne-r*
woman-OBL[17] 3SG.POSS-house-ABS
'the woman's house'

(28) Ubykh (Fenwick 2011: 37) (gloss adapted from source)
ɐ-gʷɨmɜ-n *ʁɜ-ʂɜ*
DEF-cow-OBL 3SG.POSS-head
'the cow's head'

The Instrumental case function

The Instrumental case function indicates that an action is carried out by means of an instrument, typically corresponding to constructions of the type 'with X' or 'by using X' (Haspelmath 2009: 514). However, it is important to differentiate the instrumental 'with X' from the comitative 'with X', as they are often conflated in Indo-European languages (Heine 2009: 467), e.g. the English sentence 'I found him with my dog' can both have an instrumental and comitative interpretation. Instrumental case affixes are found in all the five language families of the Caucasus, but the presence of dedicated instrumental case affixes varies across the Dagestanian branches.

[15] Glossed and transliterated from *Дагъистанальул мугIрул, нужее рецц, баркала*, and translated from Russian 'Горы Дагестана, вам хвала, спасибо'.

[16] Glossed and transliterated from *О китап мени инимни китабы*, and translated from Russian 'Это книга моего младшего брата'.

[17] Kumakhov & Vamling gloss this as OERG or 'oblique ergative'.

Instrumental case affixes are found in all Kartvelian languages (Harris 1985: 72) and in all Northwest Caucasian languages (Arkadiev & Lander 2020: 395). It is almost completely absent in Avar-Andic languages, as e.g. Avar indicates instruments with the ergative case (Forker 2020b: 250), which means that the ergative and instrumental functions are only differentiated by word order in Avar, although the suffixed postposition *-gun* sometimes has an instrumental function (Alekseev et al. 2012: 301). The explicit singular ergative suffix *-cːa* in Avar (Forker 2020b: 249) is most likely cognate with the Nakh instrumental case suffixes *-ca* (Komen, Molochieva & Nichols 2020: 325).

In Iron Ossetic the ablative case suffix *-æj* also has an instrumental function (Abaev 1964: 19; Bagaev 1965: 156), which should be analysed as case syncretism due to identical ablative and instrumental case mergers in Old Persian and Sogdian, cf. the Old Iranian instrumental case suffix **-ayā* (Thordarson 2009: 157).

(29) Ingush (Nichols 2011: 422) (gloss adapted from source)
Muusaa-z guon-a-ca hwaastam chy-tiex-ar
Musa-ERG hammer-OBL-INS nail IN-hit-WPST
'Musa pounded the nail in with a hammer'

(30) Abaza (O'Herin 2002: 195) (reglossed and retransliterated)
a-tɕkʷ'ən a-ʃʷ ʒʷəhʷa-la d-ʕ-a-sə-j-d
DEF-boy DEF-door hammer-INS 3SG.H-DIR-3SG.NH-hit-PRS-DYN
'The boy pounds on the door with a hammer'

(31) Kabardian (Kumakhov & Vamling 2009: 74) (gloss adapted from source)
se mə-r se-m-č'e Ø-s-o-ɕ'-Ø
1SG it/that-ABS knife-DEF-INS 3SG.ABS-1SG.ERG-DYN-do-PRS
'I did it with the knife'

(32) Old Georgian (Fähnrich 2012: 340) (my glossing)
mo-a-rtw-m-id-es mat aklem-eb-ita da
PV-PrV-bring-SM-IPFV-3PL 3PL.DAT camel-PL-INS and
džor-eb-ita da vir-eb-ita da azavr-eb-ita sazrdel-sa
mule-PL-INS and donkey-PL-INS and ox-PL-INS provision-DAT
'They brought them provisions by camels, mules, donkeys and oxen'[18]

(33) Eastern Armenian (Dum-Tragut 2009: 90) (gloss adapted from source)
ašakart-ə gr-um ē matit-ov
pupil-DEF write-PTCP.PRS COP.3SG pencil-INS
'The pupil writes with a pencil'

[18] Translated from German 'Sie brachten ihnen Lebensmittel mit Kamelen und Maultieren und Eseln und Ochsen'.

The Comitative case function

The Comitative case function indicates that an action is carried out together with someone or something else, typically corresponding to constructions of the type 'with X' or 'together with X' (Haspelmath 2009: 514), which is important to differentiate from the instrumental 'with X' discussed above. The comitative case is widespread in all branches of the Nakh-Dagestanian family.

(34) Archi (Chumakina, Bond & Corbett 2016: 33) (original glossing)
anχː-um *a-r-ši* *ikir*
fight(IV)-PL.ABS [III/IV.PL]do-IPFV-CVB [IV.SG]be.ITER
došːob-če-łːu
sister(II).PL-PL.OBL-COM
'… we used to have fights with our sisters'

(35) Bats (Holisky & Gagua 1994: 191) (gloss adapted from source)
obi *labc'-ir* *b-aqː-ai-čo* *ɣaze-čo* *k'nat-i-ci*^{n}
3PL play-IPFV CM-big-PL-OBL good-OBL boy-PL-COM
'They played with the big (i.e. older), good boys'

The situation in the Kartvelian languages warrants a closer examination, as all languages in the family have suffixed postpositions which could be analysed as instances of comitative case, as e.g. the suffixed postposition *–urt* 'with' in Georgian[19] (Bolkvadze & Kiziria 2023: 410) and Old Georgian (Shanidze 1980: 71; Fähnrich 2012: 769). Both Lomtadze and Kajaia describe the suffixed postpositions *–ic'k'ela* and *-ič'k'uma/-ič'k'əma* as comitative case in Megrelian (Lomtadze 1987: 186; Kajaia 2001: 30).

(36) Megrelian
a. (Harris 1991b: 374) (my glossing)
džima-c'k'uma
brother-COM
'with brother'
b. (Lomtadze 1987: 186) (my glossing)
k'oči-c'k'ela
man-COM
'with a/the man'

In Iron Ossetic, the suffix *-imæ* has conventionally been described as a comitative case (Abaev 1964: 20; Bagaev 1965: 160; Thordarson 2009: 165), although Erschler claims that it should rather be analysed as a postposition, as it does not combine

[19] The suffixed postposition *–urt* is however not very productive in modern Georgian.

with the personal pronouns but the possessive proclitics instead (Erschler 2020: 649). Some languages employ spatial cases to convey comitative functions, e.g. the 'animate location' (or rather apudessive) suffix *–de* in Hinuq (Forker 2013: 98). As the comitative and instrumental functions tend to conflate, it is often difficult to properly describe comitative and instrumental cases (Stolz, Stroh & Urdze 2009: 601-602), which further indicates the importance of a usage-based functional approach.

Many Nakh-Dagestanian languages do not differentiate between the comitative and instrumental cases, as e.g. most Dargic languages use the same suffixes for both functions (Magometov 1963: 110; Sumbatova & Mutalov 2003: 25; Temirbulatova 2004: 87-89; Sumbatova 2020: 153; Forker 2020a: 65-66), while Standard Dargwa makes a distinction between comitative *–čil* and instrumental *–čibli* (Musaev 2002: 47).

(37) Sanzhi Dargwa (Forker 2020a: 65-66) (gloss adapted from source)

a. *xural* *d-ax-ul* *hej-ka=či-d-a*
by.foot 1/2.PL-go.IPFV-IPFV.CVB this-DE=SUPER-1/2PL-DIR
di-la *juldašː-a-cː ella…*
1SG.GEN friend-OBL.PL-COM
'We (were) going by foot there with my friends…'

b. *saʕ-q'-aʕn* *zamana=q'ar* *nušːa*
PROX-go-PTCP time=MOD 1PL
mašin-ni-cː ella *saʕ-q'-un=da*
car-OBL-COM PROX-go-IPFV.CVB=1SG/PL
'When we go back, we go by car'

Similar affixes expressing both comitative and instrumental functions are also found in, e.g. Chechen (Aliroev 1999: 57), Ingush (Nichols 2011: 127, 423), Kryts (Saadiev 1994: 418), Ubykh (Fenwick 2011: 43), North Azerbaijani (Širaliev 1971: 146; Ragagnin 2022: 257) and Juhuri (Authier 2012: 52, 60).

The Similative/Equative case function(s)

The Similative or Equative case function indicates that an event 'is performed in a manner that is typical for [noun] X' (Kibrik 1977: 157), thus typically corresponding to constructions of the type 'like X' (Haspelmath 2009: 515; Daniel & Ganenkov 2009: 673). The similative case function is expressed by suffixed postpositions in all Karto-Zan languages (Holisky 1991: 419; Fähnrich 1994: 176; Hewitt 1995: 69; Rostovtsev-Popiel 2020: 539) and possibly the suffix *-šal* in Svan (Gudjedjiani & Palmaitis 1986: 93; Schmidt 1991: 484).

(38) Archi (Kibrik 1977: 157) (my glossing)

to-w	*ɬːann-a-qʿdi*	*qe-r-ši*	*w-i*
that-I.SG	woman-OBL-EQU	dance-IPFV-CVB	I.SG-COP

'He is dancing like a woman'[20]

The similative case suffixes *–cor*/*-čör* have been described in Tliadal Bezhta (Kibrik & Testelets 2004: 243), while Comrie, Xalilov and Xalilova describe *-coj* under 'comparative conjunctions' (Comrie, Xalilov & Xalilova 2015: 437-438).

(39) Tliadal Bezhta (Kibrik & Testelets 2004: 243) (my glossing)

a.

öže	*abo-cor*	*Ø-oːdaː-c*
boy.ABS	father-SIMIL	I-work-PRS

'The boy works like his father'

b.

wahado	*samoljot*	*miːnä-čör*	*b-okʼi-cːa*
this.III	aeroplane(III)	bird-SIMIL	III-fly-PRS

'This aeroplane flies like a bird'

(40) Bezhta (Comrie, Xalilov & Xalilova 2015: 438) (my glossing)

bitło-ʔ	*tʼiga-coj*	*Ø-aqχ-da,*
house-IN	he.goat-SIMIL	I.SG-stand-COND
gisa	*can-coj*	*Ø-aqχ-ca*
outside	she.goat-SIMIL	I.SG-stand-PRS

'At home he acts like a buck, but in the streets like a she-goat'[21]

The Adverbial case function

The Adverbial case function predicates nominals or indicates 'subject complements' (Boeder 2005: 50) and can be described 'as turning a nominal into a secondary predicate' (Arkadiev & Lander 2020: 396). The adverbial case often conveys multiple functions, but it generally corresponds to constructions of the type 'as X' (Boeder 2005: 50). It is important to differentiate the adverbial case from the similative and equative cases, although these cases can all be translated with the construction 'as X' in English, there is an obvious difference between working 'as a teacher' in the adverbial sense (i.e. being a teacher) and working similatively or equatively 'like a teacher'.

The term adverbial is used for case forms in the Kartvelian languages (Harris 1991b: 326; Schmidt 1991: 495-496; Fähnrich 1994: 56; Hewitt 1995: 34), Northwest Caucasian languages (Rogava & Keraševa 1966: 62; Smeets 1984: 401; Lomtatidze & Klychev 1989: 98; Paris 1989: 169; Chirikba 2003a: 23; Kumaxov

[20] Translated from Russian 'Он по-женски танцует'.

[21] Glossed and transliterated from *БилIоъ тIигацой ахъда, гиса цанцой ахъца,* and translated from Russian 'Дома он держится козлом, на улице - козой', while it appears to be a conditional construction in Bezhta.

2006: 94; Kumakhov & Vamling 2009: 24; Fenwick 2011: 42), Chechen (Nichols 1994: 29), Bats (Holisky & Gagua 1994: 165), Lezgian (Haspelmath 1993: 92) and Standard Dargwa (Musaev 2002: 42; Sumbatova 2020: 153). The Avar suffix *-łun* appears to have a function similar to the adverbial case (Alekseev et al. 2012: 295).

(41) Georgian (Aronson 1990: 70) (my glossing)
davit-s *k'arg* *megobr-ad* *Ø-tvl-i-d-it*
Davit-DAT good friend-ADV 2.S-count-SM-IPFV-1/2PL
'You all used to consider David a good friend'

(42)

a. Kabardian (Kumakhov & Vamling 2009: 24)
(gloss adapted from source)
asłen *mezχʷəme-w* *me-ɬaʒe*
Aslan forester-ADV 3SG.S-work.PRS
'Aslan works as [a] forester'

b. Abzakh Adyghe (Paris 1989: 169) (reglossed)
sja-te *ʁʷəč'e-w* *m-e-łaže*
1SG.POSS-father blacksmith-ADV 3SG.S-DYN.PRS-work
'My father works as a blacksmith'[22]

(43) Bats (Hauk 2020: 66) (gloss adapted from source)
mam-is *ħaš-eɣ* *v-ex-iⁿ-sʷ*
aunt-ERG guest-ADV M-invite-AOR-1SG.ABS
'[My] aunt invited me (M) as a guest'

(44) Avar (Alekseev et al. 2012: 171) (my glossing)
dow *škol-alda* *učitel-łun* *ħalt'-ule-w* *w-ugo*
DEM-M school-SUPER teacher-ADV work-PTCP.IPFV-M M-COP
'He works as a teacher at the school'[23]

The Comparative case function

The Comparative case indicates that something is compared to something else, typically corresponding to constructions of the type 'than X' (Haspelmath 2009: 515). The comparative case has been described in a number of Nakh-Dagestanian languages, e.g. all Nakh languages (Nichols 1994: 24; Holisky & Gagua 1994: 206; Nichols & Vagapov 2004: 677, 682; Nichols 2011: 127), in the Lezgic languages Tabasaran (Babaliyeva 2013: 115), Rutul (Maxmudova 2001: 32), Budukh (Talibov

[22] Translated from French 'mon père travaille comme forgeron'.

[23] Glossed and transliterated from *Дов школалда учительлъун хIалтIулев вуго*, and translated from Russian 'Он в школе учителем работает'.

2007: 95) and Archi (Chumakina, Bond & Corbett 2016: 26), in the Tsezic languages Khwarshi (Khalilova 2009: 257), Bezhta (Kibrik & Testelets 2004: 244) and Hunzib (Van den Berg 1995: 50), Lak (Žirkov 1955: 36; Schulze 2007: 4) and Khinalug (Khvtisiashvili 2013: 95).

(45) Chechen (Nichols 1994: 57) (gloss adapted from source)

Aħmad	*dieğanna*	*šie*^{n}	*veš-iel*	*lyeqa*	*v-u*
Ahmed.NOM	in.body	REFL-GEN	brother-COMPR	tall	V-COP

'Ahmed is taller than his brother'

(46) Khinalug (Khvtisiashvili 2013: 95) (reglossed)

šire	*mıda-d*	*ksanžmä*
1PL.EXCL.POSS	mountain-PL	better
sure	*mıda-d-ıq'illi*	
2PL.POSS	mountain-PL-COMPR	

'Our mountains are better than your mountains'

Babaliyeva describes the suffixed postposition *-t'an* as a comparative suffix (Babaliyeva 2013: 115). She does not describe it as a comparative case, although she clearly demonstrates its comparative function (47).

(47) Tabasaran (Babaliyeva 2013: 136) (glossed adapted from source)

jarχla-ʔ	*a-ji*	*čʷe-t'an,*		
distant-IN	be.IN-PTCP	brother-COMPR,		
baga-h	*x-a-ji*	*ʁunši*	*užu*	*vu*
near-AD	AD-be-PTCP	neighbour	good	COP

'A nearby neighbour is better than a distant brother'[24]

There is potentially a comparative case in Megrelian, as it has the two comparative suffixes *–geša* (Rostovtsev-Popiel 2020: 539) and *-oro* (Lomtadze 1987: 186), where the first is a suffixed postposition.

The Benefactive case function

The Benefactive case function indicates that something is the beneficiary of an action, typically corresponding to constructions of the type 'for X' (Heine 2009: 463). The dative case also tends to convey a benefactive function, which means that the label benefactive case should primarily be used in languages that have both a dative case and a separate benefactive case. The benefactive case is found throughout the Caucasus, but is not strongly associated with any of the five language families.

[24] Translated from French 'Un voisin qui se trouve près est préférable à un frère qui se trouve loin'.

Affixes or affixed adpositions expressing a benefactive case function are found in Megrelian (Lomtadze 1987: 178; Harris 1991b: 326; Rostovtsev-Popiel 2020: 541), Abkhaz (Chirikba 2003a: 35), Ubykh (Fenwick 2011: 60), Bats (Holisky & Gagua 1994: 170), South Azerbaijani (Dehghani 2000: 101), Udi (Schulze 1982: 123; Ganenkov 2008: 18), Lak (Žirkov 1955: 36; Friedman 1992: 7) and Talysh (Schulze 2000: 18). The Juhuri benefactive preposition *eri* 'for' also occurs as the proclitic *ey*= (Authier 2012: 55).

(48) South Azerbaijani (Dehghani 2000: 150) (glossed adapted from source)
män *kitab-ı* *ušag-ıčın* *al-dı-m*
1SG book-ACC child-BEN buy-PST-1SG
'I bought the book for the child'

Harris labels the benefactive case in Megrelian as the 'designative case' as she describes it as indicating 'for whom something is intended' (Harris 1991b: 374).

(49) Megrelian (Harris 1991b: 374) (reglossed)
ek *iʔu* *ir* *k'oč-išo* *saxiolo-k*
this.ERG be.PST.3SG every man-BEN joy-ERG
'This was for every man something to enjoy'

The benefactive suffix *-šeni* is used with pronouns in Laz (Holisky 1991: 419; Anderson 1963: 45), and the suffixed postposition *–tvis*/*-twis* expresses a benefactive function in Georgian and Old Georgian (Fähnrich 1994: 175; Hewitt 1995: 70). The Udi examples below show that the benefactive case does not always encode the beneficiary of an event, as e.g. Udi also uses the benefactive case for constructions of the types 'to go out for X' and 'to send for X' (Ganenkov 2008: 39), much like the use of the preposition *for* in English.

(50) Udi (Ganenkov 2008: 37, 39) (original glossing)
a. *bač':ajna-n* *ič-ejnak:* *mes=e* *biq:-e=j*
swallow-ERG self-BEN nest=3SG make-PFV=PST
'The swallow made a nest for itself'[25]
b. *lap* *buruв-oj* *döš-ö=jan* *tac-i* *boʿq:-ejnak:*
exactly mountain-GEN slope-DAT=1PL go.out-AOR boar-BEN
'Right by the foot of the mountain we went out for a wild boar'[26]

[25] Translated from Russian 'Ласточка свила себе гнездо'.

[26] Translated from Russian 'Прямо у подножия горы мы пошли за кабаном'.

The Causal case function

The Causal case indicates the cause of something, typically corresponding to constructions of the type 'because of X' (Haspelmath 2009: 515). The causal case is found in a small number of Nakh-Dagestanian languages, e.g. Standard Dargwa (Musaev 2002: 138), Lak (Friedman 2020: 211), Khwarshi (Khalilova 2009: 73), Bezhta (Kibrik & Testelets 2004: 235), Hunzib (Van den Berg 1995: 50), Archi (Chumakina, Bond & Corbett 2016: 26) and Bats (Holisky & Gagua 1994: 170). Creissels describes a 'purposive' case in Northern Akhvakh which appears to rather have a causal function (Creissels 2010: 136).

The subdirectional case suffix *-kdi* in Lezgian also conveys a causal case function (Haspelmath 1993: 98). Although the causal case is functionally similar to the causal converb, cf. section 6.10.2, the causal case suffixes are generally distinct from the causal converb suffixes, while there are a few exceptions such as the Khwarshi causal case suffix *-tłeru* (Khalilova 2009: 93) and the causal converb suffix *-atłeru* (Comrie, Forker & Khalilova 2012: 160).

(51) Archi (Kibrik 1977: 156; Kibrik 1994a: 313) (my glossing)
wirχʷmul-li-šːi *zon* *qʼˤasːe<w>tːi*
work-OBL-CAUSL 1SG.ABS tired-<I.SG>become.PFV
'Because of work I am tired'

(52) Hunzib (Van den Berg 1995: 50) (gloss adapted from source)
ɑbu-žba *mə* *diʔi* *Ø-ãcʼ-á* *Ø-ɑq-ís*
father-CAUSL 2SG 1SG.DAT I-see-INF I-can-AOR.NEG
'Because of father I could not see you'

Lomtadze mentions what he calls the 'destinative case' *–(š)eni* in Megrelian (Lomtadze 1987: 186), which appears to have a causal function, and this is reiterated by Rostovtsev-Popiel (Rostovtsev-Popiel 2020: 542), who labels it 'Ablative 2'.

(53) Megrelian (Lomtadze 1987: 186) (my glossing)
kʼoč-išeni[27]
man-BEN/CAUSL
'for the man/because of the man'

The Partitive case function

The Partitive case indicates that either a subject or an object is partial, thus typically indicating an indefinite quantity or part of a total, a group or a mass noun (Luraghi & Kittilä 2014: 17-18). The partitive case is primarily associated with the Uralic language family, where one of its function in Finnish is defined as indicating an

[27] Segmented by Lomtadze as *kʼoč-iš-e-ni*.

object which ‘expresses an indefinite, nonlimited quantity (divisible words and plural words)’ (Karlsson 2018: 195) and that the ‘noun complement is in the partitive when it expresses an indefinite quantity of a substance, group or species’ (Karlsson 2018: 198). The partitive case function is also obligatorily used in Finnic languages and Basque to express negative constructions of the type ‘no X’ or ‘not any X’ (Miestamo 2014; Ariztimuño 2014: 326). The partitive case is often defined as indicating ‘partly affected patients’ (Blake 2001: 151) or ‘having to do with partial affectedness of an object argument’ (Haspelmath 2009: 514), but this is hardly the only nor the primary function of the partitive case (Luraghi & Huumo 2014; Karlsson 2018: 188).

(54) Archi (Kibrik 1977: 157) (my glossing)
zari *isː-ib* *čʼabu* *to-w-mu-n*
1SG.ERG 1SG.GEN-PL sheep-PL.ABS DEM-I.SG-OBL-GEN
čʼa-be-qʿiš *kʼolma* *a-w*
sheep-PL.OBL-PART separate do-I.SG
‘I separated my sheep from his sheep’[28]

Partitive cases are not always included in contemporary Nakh-Dagestanian descriptions, but e.g. Kibrik (1977: 59) and Chumkina (2020: 289) mention a partitive case in Archi, and I have found additional examples of nominal affixes with partitive case functions in some other Lezgic languages, e.g. Kryts (Authier 2009: 217), Budukh (Alekseev 1994b: 266; Talibov 2007: 95), and Khinalug (Khvtisiashvili 2013: 86). Other potential partitive case affixes are the suffixed postposition *-nkʲɜ* in Ubykh, as it conveyed a function which is similar to a partitive case (Fenwick 2011: 46), the Chechen case suffix *–(a)x* (Aliroev 1999: 58-59), and its cognate the Ingush suffix *–gh* (Nichols 2011: 345).

(55) Budukh (Talibov 2007: 94, 95) (my glossing)
a. *qʼundži* *kitab-dž-ikir* *sab* *qːe<vi>n*
two.OBL book(III)-OBL-PART one take<III>.IMP
‘Choose one out of two books’[29]
b. *kʼul* *qːaja-dž-ikir* *qːirv-iri*
house stone-OBL-PART build-GNOM
‘A house is built out of stone’[30]

[28] Translated from Russian ‘Я своих овец из его овец выделил’.

[29] Glossed and transliterated from *Къунджи китабджикир саб къевин*, and translated from Russian ‘Из двух книг выбери одну’.

[30] Glossed and transliterated from *КIул кьайаджикир къирвири*, and translated from Russian ‘Дом строят из камня’.

(56) Ubykh (Fenwick 2011: 74) (original glossing)
wɜ-ʁʲɜ-nkʲɜ
that-meat-from.among
Ø-fːɐtsɜ-jt'
3SG.ABS-something.one.customarily.eats-STAT.PST
'[some] of that meat was something he would customarily eat'

The Vocative case function

The function of the Vocative case is to explicitly address an addressee (Daniel & Spencer 2009: 626) and is found in a handful of languages in the Caucasus, i.e. Georgian (Hewitt 1995: 34), Old Georgian (Fähnrich 1994: 56; Tuite 2008a: 149), Ubykh (Fenwick 2011: 35), Hinuq (Forker 2013: 433) and Khwarshi (Khalilova 2009: 72). In Northern Talysh, the vocative case is only explicitly marked by affixation for feminine singular referents (Schulze 2000: 17).

(57) Old Georgian (Tuite 2008a: 151) (my glossing)
c'inac'armet'q'wel-o *davit,* *gw-i-txar*
prophet-VOC Davit, 1PL.O-PrV-tell
'Prophet (voc.) David, tell us…'

(58) Hinuq (Forker 2013: 434) (gloss adapted from source)
nox *di-ho* *ked-iyu!*
come 1SG.OBL-AD[31] girl-VOC
'Marry me, girl!'

The Contentive/Thematic case function

The Contentive or Thematic case indicates the subject or content of a conversation or thoughts, typically corresponding to constructions of the type 'about X' (Sumbatova & Mutalov 2003: 26; Daniel & Ganenkov 2009: 673). The contentive case is found in Standard Dargwa (Isaev 2004: 317), Kubachi (Magometov 1963: 110), Itsari Dargwa (Sumbatova & Mutalov 2003: 20) and Bats (Holisky & Gagua 1994: 165). The suffixed postposition *–šeni* in Megrelian (Harris 1991b: 374) appears to have a function similar to the contentive case in some contexts, while also expressing a causal function that Lomtadze labels as a 'Destinative case' (Lomtadze 1987: 186) and Rostovtsev-Popiel labels it as 'Ablative 2' (Rostovtsev-Popiel 2020: 542). Forker describes the 'Abstract suffix' *-łi* in Hinuq (Forker 2013: 436), which appears to convey a contentive case function.

[31] Forker glosses this as ILOC ''inanimate' location'.

(59) Itsari Dargwa (Sumbatova & Mutalov 2003: 26)
(gloss adapted from source)
maʕhaʕmma-j.čilla *du-l* *ʕaʕħ-ci* *χabar*
Magomed-CONTENT 1SG-ERG good-ATTR news
b=aʕq'-ib-da
NH=hear.PFV-PRET-1SG
'I received good news about Magomed'

(60) Hinuq (Forker 2013: 437) (reglossed)
q'orol *aqila-qo=n* *hayłu* *ked-i* *zonzo*
widow woman.OBL-CONT[32]=ADD that.OBL girl-ERG REFL.SG.GEN
xoddo-łi-žo *cadaq* *žo* *es-o*
husband-CONTENT[33]-GEN all thing tell-PRS
'The girl tells the widow everything about her husband'

The Possessive case function

The Possessive case indicates the possessor in 'to have' constructions or the recipient in dative constructions (Comrie, Xalilov & Xalilova 2015: 268), typically corresponding to constructions of the type 'X (possessor) has Y' (Haspelmath 2009: 511). The possessive case is technically an orientational local case in most Nakh-Dagestanian languages, as it can typically be followed by directional local case affixes (Comrie, Xalilov & Xalilova 2015: 268; Daniel 2001a: 144). Possessive predication is surprisingly complex in certain Dagestanian languages, as they often morphologically differentiate between permanent and temporary possession (Daniel & Ganenkov 2009: 684).

Possessive case affixes have been described in a small number of Nakh-Dagestanian languages, e.g. Aghul (Magometov 1970: 87), Budukh (Alekseev 1994b: 266; Talibov 2007: 90), Bezhta (Kibrik & Testelets 2004: 236; Comrie, Xalilov & Xalilova 2015: 268), Lak (Murkelinskij 1971: 85; Friedman 2020: 211) and Khinalug (Kibrik 1994b: 375; Khvtisiashvili 2013: 85-86). See the local case orientation POSS in section 5.1.2.1 for examples.

The Substitutive case function

The Substitutive case indicates substitution of some sort, typically corresponding to constructions of the type 'instead of X' (Daniel & Ganenkov 2009: 673). The substitutive case is found in a handful of Nakh-Dagestanian languages in Dagestan, i.e. Bagvalal (Daniel 2001a: 144), Archi (Chumakina, Bond & Corbett 2016: 26), Hunzib (Van den Berg 1995: 51) and Mehweb (Chechuro 2019: 57). The function

[32] Forker glosses it as AT 'location 'at''.

[33] Forker glosses it as ABST 'abstract suffix'.

of the substitutive case is identical to the substitutive converb function, see section 6.10.2.

(61) Bagvalal (Daniel 2001b: 216) (gloss adapted from source)
tɬ'or-ab-a-baːsː *išːi-ɬaː* *angiri* *rawnina*
mountain-PL-OBL-SUBST 1PL.EXCL-LOC.POSS here plain
b=uk'a-ʁalir
NH=be-IRR
'If we only had a plain here instead of mountains!'

(62) Mehweb (Chechuro 2019: 63) (original glossing)
nu *adaj-čemadal* *tukaj-ħe* *w-aʕq'-un-na*
I father-SUBST shop.OBL-IN(LAT) M-go:PFV-AOR-EGO[34]
'I went to the shop instead of father'

Remaining case functions

The Abessive, Caritive or Privative case indicates the absence of something, typically corresponding to constructions of the type 'without X' (Haspelmath 2009: 514-515). The abessive case is rare in the Caucasus and the few potential instances of abessive affixes are the privative suffix *–da* in Abkhaz (Chirikba 2003a: 23) and the plausible cognate *-dʔa* in Abaza (Genko 1955: 118; O'Herin 2020: 460). An abessive/caritive case has also been described in Bats (Dešeriev 1953: 73).

(63) Abaza (Genko 1955: 189; Tabulova 1976: 61) (my glossing)
tʃʷəmla-dʔa *ləɣʷra* *Ø-qapal*
ladder-ABESS attic 3SG.NH.ABS-jump
'(It) jumps into the attic without a ladder'[35]

The Involuntary Agent case function has also been proposed, which indicates that the agent involuntarily or accidentally performed an action, typically corresponding to constructions of the type 'X involuntarily did Y' (Haspelmath 1993: 291; Daniel & Ganenkov 2009: 673). It is highly unusual to mark involuntary actions on the agent, but it constitutes a separate case in Bagvalal (Daniel 2001a: 140), while in Lezgian it is expressed by means of the adelative case (Haspelmath 1993: 91) and in Lak by the ablative case (Friedman 2020: 211). It is therefore worth discussing whether it qualifies as a separate case, but it arguably constitutes a separate grammatical function. Involuntary agent constructions should also be compared

[34] Original gloss for 'egophoric'.

[35] Glossed and transliterated from *Чвымладъа лыгъвра хъапал*, and translated from Russian 'Без лестницы прыгает на чердак'.

with verb forms marking intentionality, as they appear to convey a similar functional category, see section 6.5.

(64) Lezgian (Haspelmath 1993: 91) (gloss adapted from source)
dide-di-w-aj *nek* *alax-na*
mother-OBL-AD-ELA milk boil.over-AOR
'The milk boiled over, caused involuntarily by the mother.'

The Addressive case function indicates the addressee of an utterance, typically marking the 'personal object of a verb of speaking' (Friedman 2020: 211). The addressive case is described in Lak, where it is identical to the possessive case (Friedman 2020: 211). Although the addressive function is arguably a separate function, the addressive case in Lak is possibly closer to the affective case function, as the addressee is the experiencer of an utterance. Note that the verb in example (65) appears to not agree with either the gender of the human female (II) nor the human male (I), but *duš* 'girl' actually belongs to the 'non-human' animate gender (III) as only 'mature' women belong to class II (Friedman 2020: 208, 241), see section 5.5.

(65) Lak (Friedman 2020: 211) (gloss adapted from source)
duš-ni-l *but:a-x* *kunu*
daughter-OBL-ERG[36] father-OBL-ADDR say.PTCP.PST
b-u-r
III-COP.PRS-3SG
'The daughter said to her father'

The Durative case 'express[es] atelic meaning with duration adverbials' (Khalilova 2009: 72), typically corresponding to constructions of the type 'during X' or 'for X (amount of time)'. Khwarshi is possibly the only language in the Caucasus described as having a durative case, and it should probably rather be analysed as a derivational suffix or possibly an adverbial case.

(66) Khwarshi (Khalilova 2009: 72) (original glossing)
hada *buco-d* *de* *ɣudul* *n-ež-i*
one.OBL month.OBL-DUR 1SG.ERG garden(IV) IV-sow-WPST
'I sowed a garden for a month'

[36] The ergative and genitive case functions are indicated by the same suffix in Lak (Friedman 2020: 211).

5.1.2. Local case functions

Local case or spatial case is a subcategory of case that encodes spatial relations of nouns or NPs, as spatial relations express how two nominal referents, i.e. the *Figure* or *Trajector* and the *Ground* or *Landmark*, relate to each other in space (Creissels 2009a: 609). Local cases indicate these spatial relations only on the noun or NP of the Ground/Landmark, as the Figure is typically a subject or a direct object. Local cases are notoriously abundant in Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 102), whereas the other language families of the Caucasus typically lack local case systems. Local case systems tend to be structured according to the variables 'relative orientation or spatial configuration, location, destination, source and path' (Blake 2001: 152, Creissels 2009a: 614).

The relative orientation or spatial configuration encodes where the *location*, *destination*, *source* or *path* of the Figure is oriented in relation to the Ground/Landmark, thereby creating a two-dimensional local case matrix (Creissels 2009a: 614). I therefore propose, following Comrie & Polinsky (1998), a two-dimensional distinction between *local case orientations*, i.e. relative orientation, and *local case directions*, i.e. location, destination, source and path, which is the conventional Nakh-Dagestanian and implicit Uralic approach for describing local case systems (Blake 2001: 152-153; Creissels 2009a: 616-617; Daniel & Ganenkov 2009: 674).

Table 5.2: The five most common local case orientations (columns) and directions (rows) with examples of their combined functions with typically associated nouns. Certain combinations are very rare and the allative and addirectional are more or less functionally identical.

	IN **'in'**	**SUPER** **'on'**	**SUB** **'under'**	**AD** **'at, by'**	**INTER** **'in (mass)'**
Essive	Inessive 'in the house'	Superessive 'on the table'	Subessive 'under the bed'	Adessive 'by the tree'	Interessive 'in the water'
Lative	Illative 'into the house'	Superlative 'onto the table'	Sublative 'in under the bed'	Allative 'to (by) the tree'	Interlative 'into the water'
Elative	Inelative 'out of the house'	Superelative 'off the table'	Subelative 'from under the bed'	Adelative 'from (by) the tree'	Interelative 'out of the water'
Directional	Indirectional 'towards (into) the house'	Superdirectional 'towards (onto) the table'	Subdirectional 'towards under the bed'	Addirectional 'towards the tree'	Interdirectional 'towards (into) the water'
Translative	Intranslative 'in through the house'	Supertranslative 'across the table'	Subtranslative 'through under the bed'	Adtranslative 'through by the tree'	Intertranslative 'through the water'

Following Kibrik (1977: 51), the Russian-language literature tends to use the terms 'localisation' for orientation and 'orientation' for direction (Daniel & Ganenkov 2009: 674; Chechuro 2019: 40), which becomes problematic with regards to Comrie & Polinsky's and Blake's terminology. Forker (2013: 78) also uses the term

direction to indicate location, destination, source or path, but 'localisation' for the relative orientation. Local case orientations typically correspond to adpositions with meanings such as 'in', 'on', 'at', 'behind' and 'under' (Creissels 2009a: 614), while specific local case directions depend on the associated local case orientation.

The *destination*, *source* and *path* generally encode the local case directions 'to', 'from' and 'through' (Blake 2001: 151). However, the *location* is completely dependent on the local case orientation, as it does not entail any form of movement, unlike destination, source and path. The location can therefore be analysed as a zero-direction and is often zero-marked, cf. the Avar local case system (Blake 2001: 152; Creissels 2009a: 617). Although the specific spatial relation of a local case is dependent on its local case orientation, there are seemingly no attested local case systems that only encode local case orientations (Creissels 2009a: 614).

Table 5.3: All observed local case orientations expressed through affixation in the data.

Local case orientation	Meaning
IN	'in', 'inside' (open or closed space)
SUPER	'on', 'above'
SUB	'under'
AD	'at', 'by'
CONT	'on (in contact with)', 'on (vertical surface)'
INTER	'in' (mass or liquid)
APUD	'next to'
POST	'behind', 'after'
ANTE	'in front of', 'before'
CUM	'among'
POSS	'in (the possession of)'

Local case systems that only encode local case directions are however abundant across the globe, as *unidimensional* local case systems express spatial relations by means of a tripartite distinction between location (*locative*/*essive*), destination (*allative*/*lative*) and source (*ablative*/*elative*) (Creissels 2009a: 614-615). I suggest that these tripartite systems should be analysed as 'orientationless' local case systems, and these are arguably found in most Turkic languages (Creissels 2009a: 614; Johanson 2022a: 48), as well as the reconstructed Proto-Indo-European case system (Ringe 2017: 25). The locative case usually indicates general locations such as 'in', 'on' or 'at' without further specifications, and can therefore be analysed as a non-specific essive case. Ablative and allative cases are in practice elative and lative local cases without a specified orientation, which means that they can be analysed as bare directional local cases. Affixed adpositions are also frequent around the world, and are mainly found in the Caucasus in the form of suffixed postpositions.

The presence of what is conventionally analysed as suffixed postpositions in Kartvelian languages makes it worth discussing the nature of these spatial indicators. They are found in all Kartvelian languages (Harris 1991b; Holisky 1991; Schmidt 1991; Fähnrich 1994; Hewitt 1995), and their form and function do not differ in any significant way from local cases in other language families (Vogt 1971: 67-74). I have therefore chosen to analyse these as instances of local case, as previously been done by e.g. Vogt (1971), Shanidze (1980), Lomtadze (1987) and Creissels (2009a: 619).

Table 5.4: Kartvelian suffixed postpositions indicating spatial orientation and direction.

Meaning	Georgian (Hewitt 1995)	Old Georgian (Fähnrich 1994)	Megrelian (Harris 1991)	Laz (Holisky 1991)	Svan (Schmidt 1991)
'in'	*-ši*	-	-	-	*-isga, -isk'a, -isa*
'on'	*-ze*	-	-	-	*-ži*
'to'	*-ši, -ze*	*-isa*	*-(i)ša*	*-ša*	*-te*
'from'	*-(i)dan*	*-gan*	*-(i)še*	*-šen*	*-χæn, -χen*
'up to'	*-mde*	*-mde*	*-šax*	*-šaki(s)*	*-nun*
'towards'	*-k'en*	-	-	*-k'ele*	-

It is quite plausible that at least some of these suggested local cases are cognates, making it theoretically possible to reconstruct a Proto-Kartvelian local case system. It is worth considering that Svan, which constitutes its own branch of the family, has the most complex local case system of all Kartvelian languages (Schmidt 1991). The opposite scenario would suggest that the Svan local case system is an innovation, which then might be due to external factors such as language contact. The neighbouring Northwest Caucasian languages Abkhaz and Ubykh do have limited spatial marking on nouns (Chirikba 2003a; Fenwick 2011), but the Svan system is clearly more complex than the Abkhaz and Ubykh systems (Schmidt 1991: 498). There are some surprising phonological similarities between the Abkhaz local cases and the Kartvelian local cases, which would suggest that it is more plausible that Abkhaz borrowed its system from Kartvelian than vice versa, especially since the closely related language Abaza lacks these local cases altogether (Lomtatidze & Klychev 1989).

5.1.2.1. Orientational local case functions

Perhaps the most salient feature of local case systems is that they differentiate at least two spatial orientations, cf. Finnish (Blake 2001: 152-153). Local case orientations are conventionally labelled by using the closest equivalent Latin preposition (Blake 2001: 153), and within contemporary Nakh-Dagestanian linguistics the Latin prepositions are usually given in the upper case, e.g. Sumbatova (2020: 153-154), Friedman (2020: 212), Forker (2020: 250), Chumakina (2020: 289). The Nakh-Dagestanian languages are known for their rich local case systems (Comrie & Polinsky 1998), but the other four language families mainly lack nominal affixes that encode spatial orientation, although both Vogt (1971) and Creissels (2009a: 619) argue that modern Georgian has a local case system. There are also local case-like affixation patterns in Svan (Schmidt 1991: 498) and Talysh (Schulze 2000: 18).

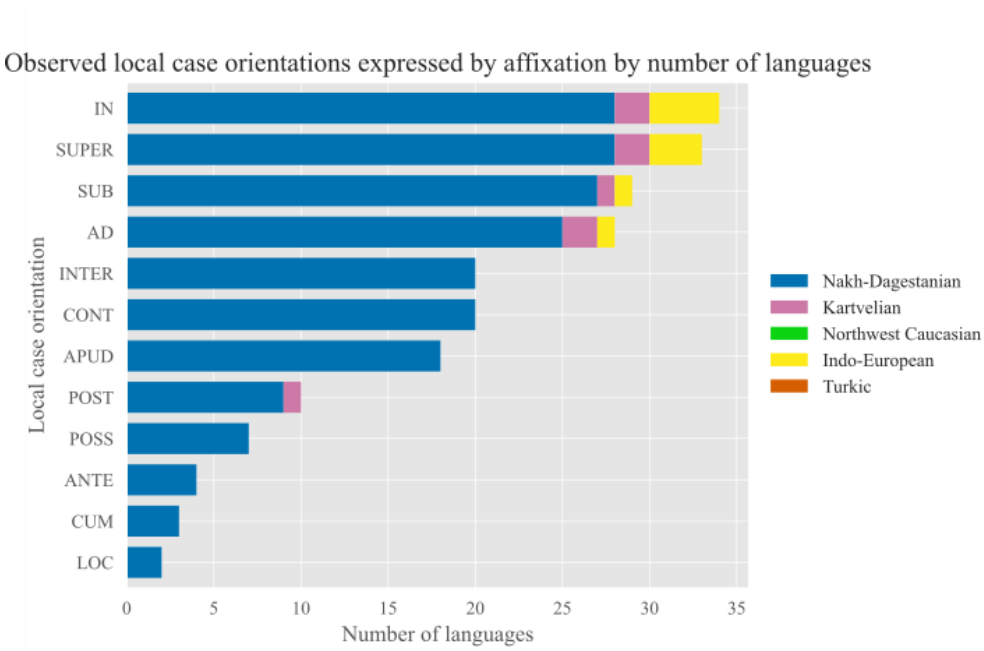


Figure 5.4: All observed local case orientations expressed by affixation in the data.

The local case orientation IN

The most common spatial orientation to be expressed by affixation in the Caucasus is the orientation IN, which indicates that the location is ‘in’ or ‘inside’ something (Blake 2001: 153). The orientation IN is combined with directional local case

affixes to form such combined local cases as the *inessive*, *illative* and the *inelative* (Haspelmath 2009: 516). It is important to differentiate between the orientation IN 'in a hollow space' from the orientation INTER 'in a mass' (Creissels 2009a: 616-617). Some Nakh-Dagestanian languages even seem to distinguish between the orientations 'in a closed hollow space' and 'in an open hollow space', e.g. Andi (Salimov 2010: 112-113), which also appears to holds true for preverb orientations, see section 6.11.1.

The orientation IN is found in almost all Nakh-Dagestanian languages but is absent in Chechen (Nichols 1994), Ingush (Nichols 2011) and Khinalug (Khvtisiashvili 2013), as they do not have local case systems. Specific affixes or affixed adpositions indicating the spatial orientation IN are also found in Georgian (Hewitt 1995: 74), Svan (Schmidt 1991: 498; Gudjedjiani & Palmaitis 1986: 42), Ossetic (Abaev 1964: 20; Bagaev 1965: 143-145) and Northern Talysh (Schulze 2000: 18).

(67) Tabasaran (Babaliyeva 2013: 34) (gloss adapted from source)
xul-a-ʔ *deʔ-na-yi-dar* *mühtal* *ʁa-š-i*
room-OBL-IN sit-PFV-PTCP-PL surprised AOR-be-AOR
'Those who had been in the room were surprised'[37]

(68) Georgian (Tuite 1998: 123) (reglossed)
irem-i *monadire-eb-s* *t'q'e-ši* *da-Ø-e-mal-a*
deer-NOM hunter-PL-DAT forest-IN PV-3.O-PASS-hide-AOR.3SG
'The deer hid from the hunters in the forest'

(69) Svan (Schmidt 1991: 535) (reglossed)
sasaːš-isa *χili* *masard* *χ-e-šan* *megmar-s*
Sasash-IN fruit enough 3.O-OV-grow trees-DAT
'In Sasash enough fruit grows on the trees'

In Avar the orientation IN is only indicated by suffixing the gender/noun class markers (Forker 2020b: 250), cf. examples (70), which could be analysed as IN being zero-marked, as the gender/noun class markers indicate the essive local cases in the Dargic languages (Sumbatova 2020: 154).

[37] Translated from French 'Ceux qui étaient dans la pièce s'étonnèrent'.

(70) Avar (Alekseev et al. 2012: 249, 312) (my glossing)

a. *was* *roqχːʼ-ow* *w-ugo*
boy house.OBL-IN.M M-be.PRS
'The boy is at home (lit. in the house)'[38]

b. *hab* *qχːʼo-jał* *radal-al-da-sa* *naqχːe*
this.NH.SG day-OBL morning-OBL-IN-ELA after
alipat *roqχːʼ-oj* *ji-kʼin-čʼo*
Alipat(F) house.OBL-IN.F.SG F.SG-be-AOR.NEG
'Since the morning of that day Alipat has not been at home (lit. in the house)'[39]

The locative case suffix *-um* in Eastern Armenian is worth mentioning as it only encodes the orientation IN (Dum-Tragut 2009: 101), and it is therefore different from typical locatives that combine the orientations IN and SUPER, cf. the Turkic languages (Johanson 2022a: 48), while the preposition *i* in Classical Armenian sometimes merged with the noun as the prefixed preposition *y-* 'in' (Schmitt 2007: 170).

(71) Eastern Armenian (Dum-Tragut 2009: 178, 307) (reglossed)

a. *ani-n* *zbaʁv-um* *ē* *senyak-um*
Ani-DEF study-PTCP.PRS COP.3SG room-LOC
'Ani studies in her room'

b. *čaš-ə* *seʁan-i* *vra-n* *a*
food-DEF table-DAT on-DEF COP.3SG
'The food is on the table'

The local case orientation SUPER

The orientation SUPER indicates that the location is 'on' or 'on top of' a landmark (De Hoop & Zwarts 2009: 179; Haspelmath 2009: 516). The orientation SUPER seems to be associated with the orientation IN, as almost all languages with an affix for IN also have an affix for SUPER in the Caucasus, with some of the only exceptions being Bats (Dešeriev 1953: 64-65; Holisky & Gagua 1994: 167-168), Udi (Schulze 1982: 119; Ganenkov 2008: 15) and Budukh (Alekseev 1994b: 266). Consequently, nominal affixes indicating the orientation SUPER, i.e. in the combined local cases superessive, superlative and superelative, are found in almost all Dagestanian languages and arguably in Georgian (Hewitt 1995: 74), Svan

[38] Glossed and transliterated from *Вас рокьов вуго*, and translated from Russian 'Мальчик дома (в доме) есть'.

[39] Glossed and transliterated from *Гьаб къоялъ радалалдаса нахъе Алипат рокъой йикIинчIо*, and translated from Russian 'В тот день с утра Алипат дома не было', but the combination of the elative case and *naqχːe* is also translated as *с этих (тех) пор* 'from now on', 'since then' (Alekseev et al. 2012: 243).

(Gudjedjiani & Palmaitis 1986: 38; Schmidt 1991: 498), Ossetic (Abaev 1964: 20; Bagaev 1965: 158-159) and Northern Talysh (Schulze 2000: 18). The orientation SUPER is generally expressed in Eastern Armenian by using the suffix *–i*, which is identical to the dative case (Dum-Tragut 2009: 86), cf. example (71).

(72) Hinuq (Forker 2013: 88) (original glossing)
hayłi *xʷin-tłʼo* *goł* *hes* *aže*
there mountain-SUPER be one tree
'On the mountain there is one tree'

(73) Georgian (Bolkvadze & Kiziria 2023: 498) (my glossing)
mt-eb-ze *tovl-i* *i-d-o*
mountain-PL-SUPER snow-NOM PASS-put-AOR.3SG
magram *bar-ši* *vard-eb-i* *qʼva-od-nen*
but valley-IN rose-PL-NOM bloom-IPFV-3PL
'There was snow in (lit. on) the mountains, but in the valley roses were blooming'

(74) Ossetic (Bagaev 1965: 159) (my glossing)
χœχ-t-il *mit* *ra-warid-is*
mountain-PL-SUPER snow PV-fall.PST-3SG
'Snow has fallen on the mountains'[40]

(75) Svan (Tuite 1997: 36) (gloss adapted from source)
tʼabg-œr-ži *diœr-s* *i* *leɣw-s* *œ-d-isg-œl-i-x*
table-PL-SUPER bread-DAT and meat-DAT NV-lie-SM-VPL-SM-PL
'They put meat and bread on the tables'

The local case orientation SUB

The orientation SUB indicates that something is located under or below a landmark, typically corresponding to 'under X' (Blake 2001: 153; De Hoop & Zwarts 2009: 179). The orientation SUB is expressed by affixation in all Dagestanian languages except Khinalug (Khvtisiashvili 2013), Udi (Schulze 1982) and Mehweb (Chechuro 2019). Outside the Nakh-Dagestanian language family, nominal SUB affixes are only found in Svan (Gudjedjiani & Palmaitis 1986: 39; Schmidt 1991: 498) and Northern Talysh (Schulze 2000: 18).

[40] Glossed and transliterated from *Хœхтыл мит рауарыдис*, and translated from Russian 'На горах выпал снег'.

(76) Kryts (Authier 2009: 173) (reglossed and retransliterated)
tikan *za* *ʁil-i-k* *ča-škar-e*
thorn 1SG.GEN foot-OBL-SUB DE-stick.MP-PRS
'The thorn is stuck under my foot'[41]

(77) Kubachi (Magometov 1963: 336) (my glossing)
ūx=hūč:-ib *bic'-li-ta-l,*
be.afraid-PTCP.PST wolf-OBL-ANTE-ELA,
šūmag-gu *wij-w-iga-d*
rock-SUB SUB-I.SG-hide-AOR.1SG
'Being afraid of the wolf, I hid under a rock'[42]

(78) Upper Svan (Gudjedjiani & Palmaitis 1986: 131-132) (reglossed)
päṭ'w *rok* *mā̄zum* *li* *jede* *mā̄zum* *rok*
millet QUOT which.size COP.3SG or which.size QUOT
ɣər-i *lekwēr-ču?*
go-SM mill-SUB
'Of what size is the millet, that is, how much of it goes in under the millstone?'

The local case orientation AD

The function of the orientation AD is somewhat vaguer than IN and SUPER, both in its definition and in its application, as it indicates that something is located 'at X' (Blake 2001: 153; De Hoop & Zwarts 2009: 179; Haspelmath 2009: 516). The orientations AD, CONT and APUD are functionally related and often difficult to differentiate, where AD is the least specific, cf. the orientations CONT and APUD for further discussion. Affixes indicating the orientation AD are found in almost all Nakh-Dagestanian languages, which possibly includes the allative suffixes in Ingush (Nichols 2011: 127, 424) and the adessive suffix *-χ* in Khinalug (Kibrik 1994b: 375), but possibly only three non-Nakh-Dagestanian languages, i.e. Georgian *–tan* (Hewitt 1995: 75; Creissels 2009a: 619), Svan *-məq'* (Schmidt 1991: 498) and Northern Talysh *–ton* (Schulze 2000: 18).

(79) Lezgian (Haspelmath 1993: 90) (gloss adapted from source)
nadir *rak'-ar-i-w* *aqʷaz-na*
Nadir door-PL-OBL-AD stop-AOR
'Nadir stopped at the door'

[41] Translated from French 'L'épine s'enfonce dans mon pied'.

[42] Translated from Russian 'Испугавшись волка, я спрятался под скалу'.

(80) Sanzhi Dargwa (Forker 2020a: 199) (original glossing)
aba *čːiħri-r=de* *cin-na* *ucːi-li-šːu-r*
mother Chakhri-F=PST REFL.SG-GEN brother-OBL-AD-F
'My mother was in Chakhri, at her brother's place'

The local case orientation INTER

The orientation INTER indicates that something is inside a solid substance, a mass or a liquid (Sumbatova 2020: 154; Forker 2020b: 250), and it must be differentiated from IN, which indicates that a figure is inside a closed or open hollow space (Daniel & Ganenkov 2009: 675). The orientation INTER is also typically used for mass nouns such as 'water', 'flour' and 'sand' (Daniel & Ganenkov 2009: 675; Chumakina, Bond & Corbett 2016: 29).

(81) Mehweb (Chechuro 2019: 66) (gloss adapted from source)
kʼas *ħarkʼʷi-ze-b* *le-b*
fish river-INTER-ESS.NH be-NH
'The fish is in the river'

The orientation INTER must however be differentiated from the orientation CUM, which indicates an orientation inside groups of non-mass nouns. Nominal affixes expressing the orientation INTER are found in most but far from all Nakh-Dagestanian languages, and the Avar-Andic branch stands out as it is the only Nakh-Dagestanian branch in which all languages have INTER affixes.

(82) Avar (Alekseev et al. 2012: 240-241) (my glossing)
du-r *ču* *ɬːe-ɬː* *b-egi-ze* *žub-ana*
2SG-GEN horse water.OBL-INTER NH-lie-INF mix-AOR
he-l-da-łun *b-ičːʼ-ana* *di-da*
DEM-OBL-SUPER-ADV NH-understand-AOR 1SG-SUPER
dur *ču* *ganšː-i-da* *χaχ-un* *b-ukʼ-in*
2SG.GEN horse buffalo-OBL-SUPER suckle-PTCP.PFV NH-be-MSD
'Your horse was lying in the water, and from this I understood that your horse was raised by (water) buffalos (lit. suckled on a buffalo)'[43]

Nichols mentions the 'adverbial' suffix *–l(a)*, which quite clearly indicates the orientation INTER (Nichols 2011: 394), and it is likely cognate with the INTER suffix *–lŏ* in Bats (Dešeriev 1953: 65; Holisky & Gagua 1994: 167). The situation in Chechen is somewhat more complicated, as the suffix *–(a)x* seemingly expresses the orientation INTER among a myriad of other functions that appear to include

[43] Glossed and transliterated from *Дур чу лъелъ бегизе жубана, гьелдалъун бичIчIана дида дур чу ганщида хахун букIин*, and translated from Russian 'Твоя лошадь стала в воду ложиться, из этого я понял, что она вскормлена буйволом'.

affective and partitive functions (Aliroev 1999: 58-59). In Hinuq, the local case suffix *-ɬ* appears to express the orientation INTER (Forker 2013: 81). The orientation INTER is also frequently used for the nouns such as 'village', 'city' and 'forest' in some Nakh-Dagestanian languages, and e.g. in Chamalal 'in the city' can either be expressed with the interessive *qχala-tɬ'a* or with the contessive as *qχala-č'* (Bokarev 1949a: 51).

(83) Chamalal (Bokarev 1949a: 51) (my glossing)
dĩː *wu-k'a* *qχala-tɬ'ã*
1SG.ABS I.SG-be.PST city-INTER
'I was in the city'[44]

The local case orientation CONT

The orientation CONT indicates that something is in contact with or attached to the associated landmark, typically, but not exclusively, indicating orientation on a vertical surface (Haspelmath 2009: 516; Daniel & Ganenkov 2009: 675-676). The orientation CONT is often used to indicate that something is 'hanging on the wall', but it also frequently occurs with human referents, e.g. contexts like 'beard on the cheeks' (Daniel & Ganenkov 2009: 676) or 'ring on a finger' (Chechuro 2019: 58). In Mehweb, the orientation CONT is expressed by the same local case suffixes as for the orientations SUPER and INTER (Chechuro 2019: 58).

(84) Aghul (Magometov 1970: 172) (my glossing)
misa-k *cal-i-k* *t'ut'-ar* *ke*
here-CONT wall-OBL-CONT fly-PL CONT.be
'There are flies here on the wall'[45]

(85) Bagvalal (Kazenin 2001: 586) (gloss adapted from source)
pat'imat-i-r *mač'-alu-č'* *ʁʷarza*
Patimat-OBL-ERG child-OBL.PL-CONT clothes
b=al-irā-χ-da, *kunt'ēna* *χabal-la* *ekʷ'a*
NH-put.on-IPFV-CVB-EMPH[46] husband-COM talk-SUPER be.PRS
'Patimat is dressing (lit. putting clothes on) the children,
and talking to (lit. on a conversation with) her husband'[47]

[44] Glossed and transliterated from *дӣ вукIа хъалалIа̃*, and translated from Russian '[Я] был в городе'.

[45] Translated from Russian '[З]десь на стене есть мухи'.

[46] The original gloss is DA 'логическая и эмфатическая частица' (Kibrik et al. 2001: 881).

[47] Translated from Russian 'Патимат одевает детей и разговаривает [= на разговоре] с мужем'

(86) Bezhta (Comrie, Xalilov & Xalilova 2015: 267) (my glossing)

isi	*muqχ'o-da-ɬ*	*p'alt'o*	*b-oχol-lo*
sister.ERG	nail-OBL-CONT	coat	III.SG-hang-AOR

'The sister hung her coat on a nail'[48]

In Rutul, the semantic overlap between the orientations CONT and INTER described below becomes apparent, as the suffix *-kʲ* covers typical contessive functions as in (87) but also the typical interessive function 'in a substance or liquid' as in (87), as the interessive by definition implies contact with the landmark. This demonstrates that the distinction between the orientations CONT and INTER is not always straightforward, while example (87a) indicates that the primary function of the local case suffix *-kʲ* in Rutul is contessive, as suggested by Maxmudova (2001: 81).

(87) Rutul (Maxmudova 2001: 81) (gloss adapted from source)

a.

masali-kʲ	*šikil*	*k-i*
wall-CONT	picture.ABS	CONT-be

'A portrait hangs on the wall'[49]

b.

xidi-kʲ	*baluʁ*	*k-i*
water-CONT	fish.ABS	CONT-be

'There are fish in the water'[50]

The local case orientation APUD

The orientation APUD indicates that something is located next to or close to the landmark, typically corresponding to 'by X, 'close to X', 'next to X' or 'near X' (Haspelmath 2009: 516; Forker 2020b: 250). The orientation APUD is often difficult to distinguish from the orientation AD due their similar semantics, and unfortunately there even appears to be conflicting definitions of the APUD/AD distinction in, e.g. the Andic and Tsezic languages (Comrie, Xalilov & Xalilova 2015: 264). The distinction is often related to whether the figure is in contact with the landmark or not, as the orientation APUD indicates a lack of contact whereas AD can indicate either in, e.g. Bezhta (Comrie, Xalilov & Xalilova 2015: 264-265). Forker (2013: 98-102) describes 'animate' and 'inanimate location' in the closely related Tsezic language Hinuq, but the animacy-distinction between these orientations is possibly semantically motivated, as the examples of the likely cognate orientations APUD and AD in Bezhta also exhibit a similar tendency to

[48] Glossed and transliterated from *Иси мукъодалъ пIалтIо бохолло*, and translated from Russian 'Сестра повесила пальто на гвоздь'.

[49] Transliterated from *Масалы-к' шикил ки*, and translated from Russian 'На стене висит портрет'.

[50] Transliterated from *Хьиди-к' балугъ ки*, and translated from Russian 'В воде водится рыба'.

correlate APUD with animate nouns and AD with inanimate nouns (Comrie, Xalilov & Xalilova 2015: 263-265).

(88) Bezhta (Comrie, Xalilov & Xalilova 2015: 264-265) (my glossing)
a. *kibba-doj ijo gej*
girl-APUD mother COP
'The mother is next to the girl'[51]
b. *ijo-doj-s gedo b-öčö-jö*
mother-APUD-ELA cat III.SG-leave-AOR
'The cat ran away from the mother'[52]
c. *bitło-ʁa-s qχ'owa b-ö<wä>čö-jo*
house-AD-ELA child.PL I/II.PL.leave<I/II.PL>-AOR
'The children ran away from the house'[53]

(89) Andi (Salimov 2010: 110) (my glossing)
ješi ilu-χa halt'u-mado
daughter mother-APUD work-PRS
'The daughter is working by her mother'[54]

The local case suffix –h/-x in Tabasaran is described as an 'adlocative' by Babaliyeva (2013: 35), whereas its meaning and the examples provided by her and earlier by Alekseev and Šixalieva (2003: 39-40) rather suggests that it conveys the orientation APUD, as there is no apparent contact involved.

(90) Tabasaran (gloss adapted from source)
a. (Alekseev & Šixalieva 2003: 39; Babaliyeva 2013: 35)
urnar-i-x gamuš da<b>qχ-na
gate.PL-APUD buffalo <NH>lie-PFV
'There lies a buffalo by the gate (in front of the gate)'
b. (Alekseev & Šixalieva 2003: 44; Babaliyeva 2013: 48)
jic äraba-ji-x-di кä<b>ʁ-üra
bull cart-OBL-APUD-COM <NH>go-PRS
'The bull walks next to the cart'

[51] Glossed and transliterated from *Киббадой ийо гей*, and translated from Russian 'Около девочки находится мама'.

[52] Glossed and transliterated from *Ийодойс гедо боьчоьйоь*, and translated from Russian 'От мамы ушла (убежала) кошка'.

[53] Glossed and transliterated from *БилIогъас къова боьваьчоьйо*, and translated from Russian 'Из дома дети убежали'. Due to vowel harmony, the expected verb form would be *боьваьчоьйоь*/ *бöwäčöjö*, which is found in a similar example on the bottom of page 435.

[54] Glossed and transliterated from *йеши илуха гьалтIумадо*, and translated from Russian '[Д]очь у матери работает'.

The orientations APUD, AD and CONT appear to be functionally related, as they seem to form a continuum from 'no contact' (APUD) – 'contact unspecified' (AD) – 'in contact' (CONT), which is supported by the presence of all three orientations in all Tsezic languages, in a few Andic languages and in Mehweb.

The local case orientation POST

The orientation POST indicates that something is located behind something else (De Hoop & Zwarts 2009: 179). Nominal affixes expressing the orientation POST are primarily found in some Lezgic languages but also a handful of other Dagestanian languages, e.g. Lak (Friedman 2020: 212) and Sanzhi Dargwa (Forker 2020a: 69). There is also the suffixed postposition *-ʁo*, which indicates the orientation POST in Svan (Schmidt 1991: 498).

(91) Lezgian (Haspelmath 1993: 93) (gloss adapted from source)

ʁuš	*demir-a-n*	*dalu-di-q^{h}*	*čünüχ*	*xa-na*
girl	Demir-OBL-GEN	back-OBL-POST	hide	ANTIC-AOR

'The girl hid behind Demir's back'

(92) Rutul (Maxmudova 2001: 80) (gloss adapted from source)

tɨla	*χalɨ-qla:*	*xuʔ*	*qi-b-q'ɨ-ri*
dog.ABS	house-POST.ELA	forward	POST-III.SG-come.out-PST

'The dog came out from behind the house'[55]

The local case orientation POSS

The orientation POSS indicates that the noun is in the possession of someone and in direct contact with the possessor (Comrie, Xalilov & Xalilova 2015: 268), either in their hands on their body (Daniel 2001a: 141). The POSS orientation has been described in Bagvalal (Daniel 2001a: 141), Bezhta (Comrie, Xalilov & Xalilova 2015: 268), Hunzib (Isakov & Xalilov 2012: 134) and Khinalug (Khvtisiashvili 2013: 85-86). The function is somewhat wider in Bagvalal, as it also includes e.g. family members and possessions in the possessor's home (Daniel 2001a: 141), thus not in direct contact with the possessor.

The possessive local case is often impossible to distinguish from the possessive case, but the presence of posslative, posselative and posstranslative indicates that these are local cases, as the POSS suffix interacts with the directional local case suffixes (Comrie, Xalilov & Xalilova 2015: 268), cf. example (93). The orientation POST has shifted to indicate the POSS functions in many Lezgic languages, e.g. Aghul (Magometov 1970: 87) and Budukh (Alekseev 1994b: 266; Talibov 2007: 90).

[55] Transliterated from *Тыла халы-хълаа хъуъ хъ-ибкьы-ри*, and translated from Russian 'Собака вышла из-за дома'.

(93) Bezhta (Comrie, Xalilov & Xalilova 2015: 268) (my glossing)

a. *aqχ-a-qχa* *kō-ʔ* *bäbä* *gej*
wife-OBL-POSS hand-IN bread COP
'The wife has bread in her hands'[56]

b. *aqχ-a-qχa-s* *kō-ʔo-s* *c'it'* *j-ẽccak-ijo*
wife-OBL-POSS-ELA hand-IN-ELA knife II-escape-AOR
'The knife fell out of the wife's hand'[57]

(94) Khinalug (Khvtisiashvili 2013: 86) (reglossed)
yä *šä* *dädä-š* *k'ičeb* *läk'u-šä-mä*
1SG 1SG.GEN mother-POSS book give-PST-IND
'I gave my mother a book'

(95) Aghul (Magometov 1970: 87) (my glossing)
za-q *q-aja* *ildeš*
1SG-POSS/POST POST-be friend
'I have a friend'[58]

The local case orientation ANTE

The orientation ANTE indicates a location in front of a landmark (Sumbatova 2020: 153). The spatial case orientation ANTE is mainly found in the Dargic languages, where most languages have an identical preverb with the same function, which is also true for Aghul (cf. section 6.11.1). The nominal suffix *-ʈʃ'ʒ* in Ubykh also indicated the orientation ANTE (Fenwick 2011: 60), cf. the Ubykh preverb *ʈʂ'ʒ-* (Fenwick 2011: 113), and the suffixed postposition *-(i)ša-x* in Megrelian appears to have a somewhat similar function (Rostovtsev-Popiel 2020: 539).

(96) Itsari Dargwa (Sumbatova & Mutalov 2003: 30)
(gloss adapted from source)
tukan-ni-sa=d *qːarpuz-i*
shop-OBL-ANTE=NH.PL watermelon-PL
d=irc-a-ca=d
NH.PL-sell.IPFV-PROG-PRS.NH.PL
'They sell watermelons in front of the shop'

[56] Glossed and transliterated from *Ахъахъа коъ баьбаь гей*, and translated from Russian 'У жены на руках хлеб имеется'.

[57] Glossed and transliterated from *Ахъахъас коъос цIumI йеʰццакийо*, and translated from Russian 'У жены из рук выпал нож'.

[58] Translated from Russian '[У] меня есть товарищ'.

(97) Sanzhi Dargwa (Forker 2020a: 151) (gloss adapted from source)
caj-na marka-la hitːi če-r-uq-un ca-r qar
one-time rain-GEN after SUPER.SUS-F-go.PFV-PRET COP-F up
qal-sa qʷaˁrš b-arqʼij
house-ANTE sweep NH-do.PFV-INF
'Once after the rain (she) went up to sweep in front of the house'

The local case orientation CUM

The orientation CUM indicates that the figure is found among or together with a group that constitutes the landmark (Comrie, Xalilov & Xalilova 2015: 266). The term 'collocative' has also been used in Kryts (Authier 2009: 31). In Avar the suffix *-ł* primarily indicates the spatial orientation INTER (Forker 2020b: 250; Alekseev et al. 2012: 164), but it also encodes a cumessive function as in e.g. *łaraʕa-zu-ł* 'among the Kumyks' (Alekseev et al. 2012: 164).

(98) Bezhta (Comrie, Xalilov & Xalilova 2015: 266) (my glossing)
biše-jōl waj-ā-ʁoj qχʼacʼo j-õqχʼo-jo
calf(III)-PL cow-PL-CUM together III.PL-come-AOR
'The calves came together with (lit. among) the cows'[59]

(99) Kryts (Authier 2009: 222) (gloss adapted from source, retransliterated)
nahčavan zin mal-dži-vas-ar la-lsal-džiz
evening 1SG cattle-OBL-CUM-ELA PV-return-SIM.CVB
'This evening, when I will return from herding the sheep… (lit. from among the cattle)'[60]

The local case orientation LOC

The orientation LOC has been used by e.g. Forker (2020a) to indicate an orientation that combines the orientations IN, SUPER and possibly AD (Forker 2020a: 66). The local case suffixes *–a/-e* in Budukh express both the orientations IN and SUPER (Alekseev 1994b: 266; Talibov 2007: 86-87), and a similar phenomenon is also found in Ghodoberi, as the suffix *–(j)alda*, which has been borrowed from Avar, expresses both the orientations IN and SUPER (Kibrik 1996: 82). It is therefore worth discussing whether LOC is an unspecified orientation in its own right or if it is just a combination of multiple orientations, which is common with other combined orientations such as IN and INTER in e.g. the Xaidaq suffix *-cːi* (Temirbulatova 2004: 97) and the Bats suffixes *–lŏ/-lʷ* and *-loħ* (Dešeriev 1953: 64-65; Holisky & Gagua 1994: 167; Hauk 2020: 42).

[59] Glossed and transliterated from *Бишейӧл вайā-гъой къацIо йоʰкъойо*, and translated from Russian 'Телята вместе с коровами пришли'.

[60] Translated from French 'Ce soir, quand je rentrerai de garder les moutons…'.

5.1.2.2. Directional local case functions

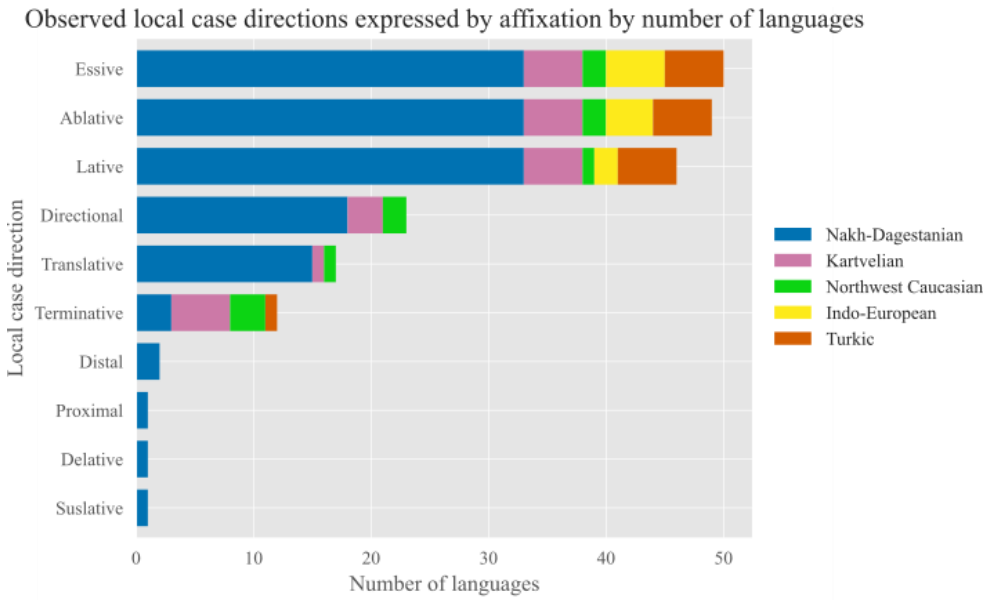


Figure 5.5: All observed local case directions (expressed through affixation in the data.

Directional local cases constitute the second dimension of local case systems, as they typically encode *location*, *destination*, *source* and *path* (Blake 2001: 152; Creissels 2009a: 614). The three most common directional local cases, *essive*, *lative* and *elative*, often occur as the independent cases *locative*, *allative* and *ablative* in languages that lack true local case systems (Creissels 2009a: 614-615). The Nakh-Dagestanian local case systems vary with regards to which direction is zero-marked, as the lack of an explicit directional local case affix either implies the essive case as in Avar (Alekseev et al. 2012: 164-165), the lative case as in Dargwa (Musaev 2002: 48; Isaev 2004: 317), or both in some languages such as Tindi (Magomedova 2012: 86).

Table 5.5: All observed local case directions expressed through affixation in the data.

Local case direction	***Meaning***
Essive/Locative	Ø-movement (Location)
Elative/Ablative	'from' (Source)
Lative/Allative	'to' (Destination)
Directional	'towards' (Destination)
Translative	'through, along' (Path)
Terminative	'up to' (Destination)
Proximal & Distal	'hither' & 'thither' (Destination/Source)
Delative & Suslative	'down' & 'up' (Path)

The Locative case function and the Essive local case direction

The Locative case function indicates the location of something, typically corresponding to 'in X', 'on X' or 'at X' (Blake 2001: 151). The locative case could be analysed as a non-specific local case, as it often does not distinguish between 'in', 'on', a distinction which is typically found in local case systems (Creissels 2009a: 615-617). The locative case is found in all Turkic languages of the Caucasus (Dehghani 2000: 101; Ragagnin 2022: 247; Berta & Csató 2022: 325; Karakoç 2022: 358) and in Classical and Eastern Armenian, Juhuri and Talysh (Meillet 1936: 64-65; Schulze 2000: 17-18; Dum-Tragut 2009: 71; Authier 2012: 50). In Classical Armenian the locative case was, with a few exceptions, identical to the dative case in the singular but to the accusative case in the plural (Meillet 1936: 64-65), which clearly indicates a case syncretism.

(100) Kumyk (Abdullaeva et al. 2014: 201) (my glossing)
zaripat *škola-da* *joq* *edi*
Zaripat school-LOC NEG.EXIST COP.PST
'Zaripat was not at school'[61]

(101) Juhuri (Authier 2012: 51) (gloss adapted from source)
pineçi *veno* *des=e* *e=şimşil* *xüşde* *pesde* *ser=e*
cobbler put.on.AOR hand-DAT LOC=sword REFL then head=DAT
e=hovo *tik* *gür-de* *e=asmu* *denişi-re*
LOC=air upright take-CVB LOC=sky look-CVB
'The cobbler put his hand on his sword, then raised his head in the air and looked into the sky'[62]

The Kartvelian dative also has a locative function, which means that the dative case can convey a locative meaning without any postpositions. The locative function of the Kartvelian dative case can be found in all Kartvelian languages except standard Georgian (Harris 1991b: 372; Holisky 1991: 452; Fähnrich 1994: 56; Tuite 1997: 17; Tuite 2008a; Öztürk & Pöchtrager 2011: 27), and the dative case has even been explicitly described as a dative/locative in Svan (Gudjedjiani & Palmaitis 1986: 38). A similar phenomenon has been observed in Classical Greek, where particularly in Homeric Greek the dative case alone could encoded locative relations (Luraghi 2009: 149).

[61] Glossed and transliterated from *Зарипат школада ёкъ эди*, and translated from Russian 'Зарипат не было в школе'. As *joq* expresses negative existentials, the literal translation is 'There was no Zaripat at school'.

[62] Translated from French 'Le cordonnier mit la main à son épée, puis levant la tête et regardant en l'air vers le ciel'.

(102) Old Georgian (Tuite 2008a: 150) (my glossing)
xiq'o *igi* *ierusalem-s*
be-AOR-3SG 3SG Jerusalem-DAT/LOC
'He was in Jerusalem'

(103) Megrelian (Harris 1991b: 372) (my glossing)
baɣan-ep-i *ʔude-s* *skid-un-a*
child-PL-NOM house-DAT/LOC stay-SM-3PL
'The children are staying in/at the house'

(104) Svan (Gudjedjiani & Palmaitis 1986: 29) (gloss adapted from source)
hant'o *ar-da* *mest'ja-s*
Hanto be-IPFV.3SG Mestia-DAT/LOC
'Hanto lived in Mestia'

The status of the 'locative' datives is further complicated by the widespread use of spatial preverbs in Kartvelian languages (cf. section 6.11), as it is not always possible to disambiguate which element of the clause that carries the spatial reference. Unsurprisingly, 'locative' datives and spatial preverbs often co-occur, and similar patterns are also found in the Northwest Caucasian languages, where the oblique case indicates the locus of the event (Arkadiev & Lander 2020: 413).

(105) Megrelian (Reseck 2015: 63) (gloss adapted from source)
a. *baɣana-k* *c'q'ar-s* *ino-sxap'-u*
child-ERG water-DAT IN-jump-AOR.3SG
'The child jumped into the water'
b. *uškur-i* *k'aračxa-s* *ino-dzə*
apple-NOM basket-DAT IN-lie.PRS.3SG
'The apple lies in the basket'

The presence of locative affixes in the Northwest Caucasian languages is somewhat more complicated, as Arkadiev & Lander (2020: 395) only describe Ubykh as having had a locative case, while suffixed postpositions with a locative function have also been described in Abkhaz (Aristava 1968: 28-29; Chirikba 2003a: 23).

(106) Ubykh (Fenwick 2011: 47) (original glossing)
sɨ-tʷ *ʁʒ́-tɕʷjʒ-ʁʒ*
1SG.POSS-father.ERG 3SG.POSS-room-LOC
jʒ-Ø-txɨ́-n
NULL.ABS-3SG.ERG-write-PRS
'My father is writing in his room'

(107) Abkhaz (Aristava 1968: 142) (my glossing)
aw-χa *zegʲə* *anə-tɕʷa,*
DEM-night everyone TEMP.CVB-sleep
aʃʃəħʷa *d-ɥagəla-n*
quietly 3SG.H.ABS-get.up-PST.IND
a-mats'urta-ɬʂ'ə *l-ɬʂə-l-ʃ-əjt'*
DEF-kitchen-LOC 3SG.F.OBL-REFL-3SG.F.ERG-kill-AOR
'That night, when everyone was sleeping, she quietly got up and hanged herself in the kitchen'[63]

The Essive encodes the *location* of spatial relations (Creissels 2009a: 617). Essive affixes are found in almost all languages of the Caucasus in some form, either implicitly as part of a local case system, as a locative case or as affixed adpositions (Creissels 2009a: 619). The essive case is often the zero-marked form in Nakh-Dagestanian local case systems (Daniel & Ganenkov 2009: 674), and is therefore intertwined with the orientational affixes (Creissels 2009a: 617). There are some languages in which the essive is explicitly marked by a separate affix, e.g. most Dargic languages (Magometov 1963: 68; Sumbatova & Mutalov 2003: 22; Isaev 2004: 315-317; Chechuro 2019: 58; Forker 2020a: 43). The essive suffixes in Dargic languages and Avar are interesting, as they usually only consist of the noun class marker, which consequently has to agree in noun class with the absolutive noun or pronoun it relates to (Chechuro 2019: 58; Forker 2020b: 250).

(108) Lezgian (Haspelmath 1993: 229) (gloss adapted from source)
zi *gada-di-n* *rik'-e* *wa-z* *qarši*
1SG.GEN boy-OBL-GEN heart-IN.ESS you-DAT opposite
mühübbat *xu-raj*
love be-OPT
'May there be love in my son's heart toward you'

(109) Mehweb (Ganenkov 2019: 196) (gloss adapted from source)
madina-s *ʔali* *urč-e-w* *le-w*
Madina-DAT Ali.ABS heart-IN-ESS.M be-M
'Madina remembers Ali'

It is rather difficult to assess the presence of locatives in the Nakh languages, as the descriptions of Ingush diverge substantially, but the suffixes *–ie, -a, -ga* constitute some sort of locative case in Ingush (Jakovlev 2001: 196-198), although they are analysed as adverbs by Nichols (Nichols 2011: 430). The same suffixes in

[63] Glossed and transliterated from *Ауха зегьы аныцәа, ашьшьыҳәа дфагылан амаӷурҭақәы л-е-ылшьит,* and translated from Russian 'Ночью (в ту ночь), когда все заснули, она тихо вышла и повесилась (убила себя) на кухне'.

Chechen have an allative meaning, but if the suffix *–ħ* is added they indicate the location of something (Jakovlev 1960: 33; Nichols 1994: 24), thereby functioning as a locative case. Locative cases in the narrow sense are quite rare in Dagestan since most Dagestanian languages have more or less intricate local case systems (Ganenkov & Maisak 2020: 102-104).

Khinalug possibly has the best example of a locative case in the Dagestanian family (Kibrik 1994b: 375; Khvtisiashvili 2013: 81), while Udi appears to have a locative case which has syncretised with the dative case (Schulze 1982: 113; Harris 2002: 25; Ganenkov 2008: 20-21). The syncretism of the dative and locative cases in Udi becomes apparent as the Nic dialect differentiates the dative *–a/-ä* from locative *–e* for certain words (Ganenkov 2008: 21), suggesting that there is a dative-locative syncretism in the other dialects of Udi. Ghodoberi has borrowed the suffix *–(j)alda* from Avar and it has the function of a locative case in certain contexts (Kibrik 1996: 82), even though Ghodoberi has a fully-fledged local case system (Saidova 2004: 83).

(110) Nic Udi (Ganenkov 2008; 139, 146) (gloss adapted from source)

a. *ʁe* *beš* *joldaš-χo-n* *samdži* *demiš=tːun*
today 1PL.POSS friend-PL-ERG first time=3PL
akː-sa *beš* *ajiz-a*
see-PRS 1PL.POSS village-DAT
'Today our friends are seeing our village for the first time'[64]

b. *beš* *ajiz-e* *gele* *odžaχ-χo=no*
1PL.POSS village-LOC many sacred.place-PL=COP.3SG
'In our village there are many sacred places'[65]

The Ablative case function and the Elative local case direction

The Ablative case function indicates that something originates from a source, typically corresponding to constructions of the type 'from X' (Haspelmath 2009: 515). The ablative case can be part of a larger local case system, but it is more common to only have a three-way distinction between locative, ablative and allative, which is arguably true for Turkic (Creissels 2009a: 614). The ablative case is found in all Turkic languages (Seegmiller 1996: 13; Schönig 1998: 251; Csató & Karakoç 1998: 336; Dehghani 2000: 101; Abdullaeva et al. 2014: 183) and all Caucasian Indo-European languages except Juhuri and Tat (Abaev 1964: 19-20; Schulze 2000: 17; Van Damme 2004: 14-16; Dum-Tragut 2009: 71; Authier 2012; Suleymanov 2020).

[64] Translated from Russian 'Сегодня наши друзья впервые видят наше село'.

[65] Translated from Russian 'В нашем селе есть много священных мест («оджахов»)'. The word *odžaχ* is of Turkic origin, cf. Azerbaijani *ocaq* 'stove, hearth'.

(111) Eastern Armenian (Dum-Tragut 2009: 308) (reglossed)
ax *mayrik-s* *indz*
INTERJ mother-1SG.POSS 1SG.DAT
vṙnd-el *ē* *tn-ic^h*
expel-PTCP.PFV COP.3SG house-ABL
'Oh, my mother has expelled me from the house'

(112) Nogai (Karakoç 2005: 185) (gloss adapted from source)
ötpek-tiŋ *bir-ew-i* *peš-ten* *šïɣ-ïp* *turï*
bread-GEN one-COLL[66]-3G.POSS oven-ABL take.out-CVB AUX
'One of the breads is taken out of the oven'

An ablative case has been described in some Kartvelian languages, e.g. Megrelian and Laz (Harris 1991b: 326; Holisky 1991: 408; Öztürk & Pöchtrager 2011: 28; Rostovtsev-Popiel 2020: 541), while the presence of ablatives in Georgian, Old Georgian and Svan is a matter of debate (Schmidt 1991: 498; Hewitt 1995: 70, 76). The instrumental case suffix *–it* also had an ablative function in Old Georgian (Fähnrich 1994: 173; Tuite 2008a: 151), which later merged with the suffix postposition *-gan* to form the suffixed ablative postposition *–idan* in modern Georgian (Fähnrich 2012: 766).

The Svan suffix *–χän*/*-χen* 'from' (Schmidt 1991: 498; Tuite 1997: 46-48) should likely also be analysed as an ablative case. Since most Nakh-Dagestanian languages have local case systems, they tend to have an elative suffix that conveys the same function as the ablative case within a local case system (Creissels 2009a: 615-617), suggesting that the distinction between the elative and ablative local case directions is largely terminological and not functional.

(113) Laz (Lacroix 2009: 414) (gloss adapted from source)
ma *andğa* *noğa-şen* *köy-şe* *eşa-p-t-i*
1SG today city-ABL village-ALL PV-1SG-go.up-AOR
'I came up from the city to the village today'[67]

The Elative local case directions indicate the source of an action or an object, typically corresponding to construction 'from X' (Blake 2001: 153). Although the terms *ablative* and *elative* are often used synonymously in Caucasian linguistics, the direction 'from' is almost exclusively labelled *elative* in Nakh-Dagestanian local case systems (Creissels 2009a: 617; Daniel & Ganenkov 2009: 674). However, it is relevant to theoretically differentiate between *ablatives* indicating 'from' and

[66] Karakoç's glossing.

[67] Translated from French 'Aujourd'hui, je suis venu au village depuis la ville'.

elatives indicating 'out of', cf. the local case systems in Finnish (Blake 2001: 153) and Hungarian (Creissels 2009a: 616).

(114) Standard Dargwa (Mutalov 2018: 61) (gloss adapted from source)
durħ-ni *wac'a-li-zi-b-ad*
boy-PL forest-OBL-IN-HPL-ELA
čar-b-uq-i *sa<b>i*
RE[68]-HPL-come.PFV-CVB COP<HPL>
'The children have come back from the forest'

(115) Karata (Magomedbekova 1971: 72) (my glossing, retransliterated)
cː'erti-tɬːi-gal *b-oqχː-e* *k'uncː'e*
mud-INTER-ELA NH.SG-pull.out-PST puppy
'They pulled the puppy out of the mud'[69]

The Allative case function and the Lative local case direction

The Allative case function indicates that something is moving towards a goal, typically corresponding to constructions of the type 'to X' (Haspelmath 2009: 515). Similar to the ablative case, the allative case can either be part of a larger local case system, often as a Lative affix, or as one of the constituents of a three-way local distinction together with a locative and an ablative (Creissels 2009a: 614). Allative affixes have been described in Old Georgian (Fähnrich 1994: 56-57, 178; Tuite 2008a: 151), Megrelian (Harris 1991b: 326; Rostovtsev-Popiel 2020: 541) and Laz (Holisky 1991: 408; Lacroix 2018: 835), while the suffix *–te* in Svan clearly has an allative function (Schmidt 1991: 498; Tuite 1997: 38).

The presence of allatives in the Kartvelian family is somewhat complicated by the apparent loss of a distinct allative in modern Georgian (Bolkvadze & Kiziria 2023: 28), as the Old Georgian allative suffixes *-d* and *–isa* have been replaced by *–ši, -ze* and *-tan* in various lative functions (Creissels 2009a: 619). The Pazar dialect of Laz has also lost a distinct allative affix, as the ablative and allative suffixes have merged to *–şe* (Öztürk & Pöchtrager 2011: 28), whereas some dialects have retained the distinction between allative *–ša* and ablative *–šen* (Holisky 1991: 408), which is similar to the Megrelian distinction between allative *–ša* and ablative *–še* (Harris 1991b: 326).

(116) Svan (Tuite 1997: 38) (gloss adapted from source)
tæš *sg-oːt-šq'æd* *lemesg-teː-sga*
cheese.NOM PV-PV.O3SG.OV-fall.AOR fire.DAT-ALL-IN
'His cheese fell into the fire'

[68] Mutalov glosses it simply as 'back-'.

[69] Translated from Russian '[И]з грязи вытащили щенка'.

Allative affixes have also been described in Chechen (Jakovlev 1960: 24; Nichols 1994: 24; Aliroev 1999: 59), Udi (Schulze 1982: 119; Harris 2002: 24), Abkhaz (Chirikba 2003a: 23) and Ossetic (Abaev 1964: 20; Bagaev 1965: 142; Erschler 2020: 647).

(117) Udi (Harris 2002: 135) (reglossed)
ta-al-le-c-i *kalabal-t'-uč'*
DIST-ADD-3SG-go-AOR servant-OBL-ALL
'And she went to the servant'

(118) Iron Ossetic (Erschler 2020: 670) (original glossing)
abon *dukani-mæ* *a-sæw-zæn* *ʃoʃlan*
today shop-ALL PV-go-FUT.3SG Soslan
'Today, Soslan will go to the shop'

The allative function coincides with the dative case in all Turkic languages of the Caucasus (Širaliev 1971: 45; Seegmiller 1996: 13; Dehghani 2000: 146; Abdullaeva et al. 2014: 193; Ragagnin 2022: 251; Karakoç 2022: 358), and Širaliev even describes the dative case suffix *–(y)a/-(y)ə* in North Azerbaijani as the 'dative-directional case' (Širaliev 1971: 43). Allative/lative-dative syncretism, where the allative/lative case function is formally identical to the dative case, is a widespread phenomenon (Creissels 2009a: 621), and is generally present in Turkic languages (Johanson 2022a: 48) and in, e.g. Lezgian (Haspelmath 1993: 87). Since this thesis has a functional approach, I have decided to analyse the dative case affixes in Turkic as having both a dative and an allative function, enabling comparisons with other allative affixes.

(119) North Azerbaijani (Širaliev 1971: 46) (my glossing)
a. *bakı-ya* *oxu-mağ-a* *gəl-ib*
Baku-DAT/ALL read-INF-DAT come-PFV.3SG
'He has come to Baku to study'[70]
b. *meşə-yə* *odun-a* *ged-ib*
forest-DAT/ALL firewood-DAT go-PFV.3SG
'He has gone to the forest for firewood'[71]

[70] Translated from Russian '[О]н приехал в Баку учиться'.

[71] Translated from Russian '[О]н пошел в лес за дровами'.

(120) Kumyk (Abdullaeva et al. 2014: 174) (my glossing)
erten-in-de *men* *maʁačqala-ʁa*
morning-3SG.POSS-LOC 1SG Maxačkala-DAT/ALL
jol-ʁa *tüš-dü-m*
road-DAT/ALL go.down-PST-1SG
'I set off for the road to Maxačkala in the (lit. its) morning'[72]

(121) Nogai (Karakoç 2005: 32) (gloss adapted from source)
bazar-ɣa *bar-ar* *e-ken-men* *aqš-am* *yoq*
bazaar-DAT/ALL go-FUT COP-EVID-1SG money-1SG.POSS NEG
'I would go to the market, but I have no money'

(122) Lezgian (Haspelmath 1993: 402) (gloss adapted from source)
har *gat-u-z* *am* *wiči-n* *č'eχi* *buba-di-n* *pataw*
every summer-DAT he.ABS self-GEN big father-GEN to
χür-ü-z *χk^{w}e-da*
village-DAT/ALL return-FUT
'Every summer he goes back to the village to his grandfather'

The Lative or Allative local case direction indicates the destination of an action or an object, typically corresponding to the preposition 'to X' (Blake 2001: 151). The terms *lative* and *allative* are often used synonymously, but the term allative in *sensu stricto* refers to the combined local case of the orientation AD and the lative direction (Blake 2001: 153).

(123) Avar (Forker 2020b: 257) (gloss adapted from source)
dun *ħež-al-d-e* *ine* *b-ugo*
1SG.ABS Hajj-OBL-SUPER-LAT go.INF NH.SG-COP
'I will go on the Hajj'

(124) Khwarshi (Khalilova 2009: 329) (original glossing, retransliterated)
bazar-tł'a-l-in *Ø-õk'-un,* *hadam-il*
market-SUP-LAT-ADD I-go-PFV.CVB people-LAT
ise.iso *zihe-bo* *l-eɣ^{w}-a* *Ø-eq-un* *žu*
REFL.GEN cow-PL.ABS NH.PL-sell-INF I-begin-UWPST that.ABS
'When he came to the market, he began to sell his own cows to the people'

[72] Glossed and transliterated from *Эртенинде мен Магъачкъалагъа ёлгъа тюшдюм*, and translated from Russian 'Утром я отправился в Махачкалу'.

The Directional local case direction

The Directional, Directive or Versative local case direction indicates the path towards a landmark, typically corresponding to the preposition 'towards' (Friedman 2020: 212; Haspelmath 2009: 515). The term *directive* is not optimal however, as it is sometimes regarded as synonymous with the allative/lative (Haspelmath 2009: 515; Creissels 2009a: 614), and directive is widely used with a completely different meaning within speech act theory.

(125) Standard Dargwa (Musaev 2002: 53) (my glossing)
il-di *anqχ-li-zi-baʿħ* *ha-j-b-iʔ-ub*
DEM-ERG garden-OBL-IN-DIR ANTE-LAT-HPL-be-AOR
'They headed towards the garden'[73]

(126) Hunzib (Van den Berg 1995: 48) (reglossed)
hɑn-ɑ-ɬ-do *Ø-ẽtɬ'e-r*
forest-OBL-CONT-DIR I-go-AOR
'[He] went in the direction of the forest'

The Translative local case direction

The Translative or Perlative local case direction indicates the path through or across a landmark, typically corresponding to the prepositions 'through' and 'along' (Haspelmath 2009: 516; Daniel & Ganenkov 2009: 674). Translative local case affixes are exclusively found in the Nakh-Dagestanian languages (Haspelmath 2009: 516) with two possible exceptions in the Caucasus, i.e. the Svan suffixed postpositions *–ka* (Gudjedjiani & Palmaitis 1986: 93) and the Abaza suffixes *–ta*/*-šta* (Lomtatidze & Klychev 1989: 98) with similar translative functions.

(127) Ghodoberi (Kibrik 1996: 264) (gloss adapted from source)
ho=w *han-tɬí-ritɬ'ì* *wu=n-atɬ-áwq'atɬì,*
3SG=M village-INTER-TRANS M=go-PRS-TEMP.CVB,
χʷáni-č'u *hiɬʲa-lá* *hank'-ú,* *t'orda* *hingur-abé*
horse-CONT above-ADD sit.down.PFV-CVB all window-PL
hinc:-abé *hiš:-ú* *ru=k'-á-da,*
door-PL close.PFV-CVB NH.PL=be.PFV-CVB-COP
adámi *is:íra* *ba=k'-uč'-á-da*
man outside HPL=be.PFV-NEG-CVB-COP
'As he was riding his horse through the village, all the windows (and) doors were closed, there was no[t] anybody outward [sic]'

[73] Glossed and transliterated from *илди анхълизибяхI гьайбиуб*, and translated from Russian '[О]ни направились в сторону сада'.

(128) Bats (Dešeriev 1953: 66) (my glossing)
uq *naχ-guiħ-ren-daħ* *v-aχ-r-as* *sk'ol-e*
DIST.OBL people-APUD[74]-ABL-TRANS I-go-AOR-1SG school-ALL
'I went to school through (lit. from beside) those people'[75]

The Terminative local case direction

The Terminative or Limitative local case direction indicates 'motion until reaching the domain' (Daniel & Ganenkov 2009: 674) or reaching 'its endpoint' (Haspelmath 2009: 515), typically corresponding to the preposition 'up to' (Creisseils 2009: 610) or to constructions of the type 'to X but not further' (Chumakina 2020: 290). The terminative local case direction is rather unusual in the Nakh-Dagestanian languages as it appears to only be found in Archi (Chumakina, Bond & Corbett 2016: 30; Chumakina 2020: 290), Khwarshi (Khalilova 2009: 74) and Bats (Holisky & Gagua 1994: 170).

Terminative affixes or clitics are found in all language families of the Caucasus except Indo-European, as a terminative case has been described in North Azerbaijani (Ragagnin 2020: 247-248), while suffixed postpositions with a terminative function are found in all Kartvelian languages, e.g. Georgian and Old Georgian *–mde* (Fähnrich 1994: 172; Hewitt 1995: 76), Megrelian *–šax* (Harris 1991b: 374; Reseck 2015: 171-172), Laz *–šaki(s)* (Anderson 1963: 89; Holisky 1991: 419) and Svan *–nun* (Schmidt 1991: 498), and in some Northwest Caucasian languages, i.e. Abkhaz *–ndza* (Chirikba 2003a: 35), Abaza *-dza* (Lomtatidze & Klychev 1989: 99; O'Herin 2020: 460) and Ubykh *-ʒwndzʒ* (Fenwick 2011: 60).

(129) Khwarshi (Khalilova 2009: 90) (original glossing)
muħamad *qod-o-ɣo-q'a* *Ø-õk'-i*
Magomed(I) wall-OBL-APUD-TERM I-go-WPST
'Magomed almost reached the wall (lit. reached until the wall)'

(130) Georgian (Bolkvadze & Kiziria 2023: 406) (my glossing)
is *dil-idan* *saɣamo-mde* *bibliotek'a-ši-a*
3SG morning-ABL evening-TERM library-IN-COP.3SG
'He is in the library from morning till evening'

[74] Dešeriev analyses *–guiħ-re(ⁿ)* as a case form meaning *со стороны* 'from the side' or 'from the direction of', while Holisky & Gagua (1994) and Hauk (2020) analyse *guiħ* as a postposition meaning 'toward', which becomes problematic if Dešeriev's examples are considered (Dešeriev 1953: 65-66).

[75] Glossed and transliterated from *укх нах-гуихьрендахь вахрас скIоле*, and translated from Russian '[В] сторону этого народа и через него пошел я в школу'.

The Distal and Proximal local case directions?

The Distal and Proximal directions are marginal functions which encode either movement 'towards here/hither' or 'towards there/thither', and they are functionally identical to the distal and proximal preverb orientations, see section 6.11.1. It is worth discussing whether these are local case directions, but I see no reason to exclude them. The distal direction is only found in two languages in the data, and the opposite proximal direction is found in Xaidaq (Temirbulatova 2004: 91-92) and Tsez (Polinsky 2015). The distal local case suffix in Tsez can be combined with all other local case suffixes (Alekseev & Radžabov 2004: 125). Xaidaq is the only language in the data to have both distal and proximal nominal suffixes, i.e. *–ten* and *–žen*, and these two suffixes also combine with all other local case suffixes (Temirbulatova 2004: 91-92). These directions are also found in a few other Dargic languages not included in the data, e.g. Tanti Dargwa (Sumbatova 2020: 155).

(131) Tsez (Comrie, Forker & Khalilova 2012: 167) (gloss adapted from source)
b-iɦinay-xo xex-bi iškol-āz-ay nex-si
HPL-fight-IPFV.CVB child-PL.ABS school-DIST-IN.ABL come-WPST
'Fighting, the children came from the school (lit. thence)'

(132) Xaidaq (Temirbulatova 2004: 101-102) (my glossing)
a. *mask:aw-c:i-r-žen*
Moscow-INTER-ELA-PROX
'From Moscow towards here/hither'[76]
b. *burχan-gu-r-ten*
roof-SUB-ELA-DIST
'From under the roof towards there/thither'[77]

The Delative and 'Suslative' local case directions?

The Delative (from Latin *de-* 'from, down, off') and what I have chosen to label the Suslative (from Vulgar Latin *sūsum* 'upwards') directions are also only found in Xaidaq (Temirbulatova 2004: 91-92), but these directions are present in some of the other Dargic languages not included in the data, e.g. Tanti Dargwa (Sumbatova 2020: 155). The delative direction indicates movement downwards and the suslative direction indicates movement upwards, and these directions are conveyed by the nominal suffixes *-k'en* and *–χen* in Xaidaq (Temirbulatova 2004: 91-92). It is important to compare these directional suffixes to their preverbal counterparts, i.e. the delative preverb *ka-* and the suslative preverb *ha-* (Temirbulatova 2004: 187),

[76] Glossed and transliterated from *Масккавцциржен*, and translated from Russian 'из Москвы по направлению сюда'.

[77] Glossed and transliterated from *бурхангуртеи*, and translated from Russian '[П]од крышею по направлению туда'.

since the delative and suslative directions are primarily associated with preverbs in the Caucasus, cf. the delative and suslative preverb directions for a discussion regarding terminology in section 6.11.2.

(133) Xaidaq (Temirbulatova 2004: 101) (my glossing)
a. *kːalkːan-sa-r-χen*
tree-AD-ELA-DE
'From the tree downwards'[78]
b. *qal-e-r-k'en*
house-IN-ELA-SUS
'From the house upwards'[79]

5.2. Number-marking functions

The grammatical category of number is found in all 56 languages of the affixal data set, indicating whether the noun is singular or plural. Singular and plural were the only numbers found in these languages, as none of the Indo-European languages of the Caucasus have retained the dual number, as it was lost already in Classical Armenian (Meillet 1936: 93), Parthian and Middle Persian (Durkin-Meisterernst 2014: 230). Number is often conflated with case in inflectional case paradigms, which is typically not the case in the Caucasus, since most languages in the region tend to prefer agglutinative morphology. There is no affix in the data which only encodes that the noun is singular, as the singular is almost universally zero-marked, and all singular affixes primarily encode case or gender/noun class.

(134) Georgian (Hewitt 1995: 138) (gloss adapted from source)

mšobl-eb-s	*Ø-u-q'var-t*	*švil-eb-i*
parents-PL-DAT	3.O-OV-love-PRS.PL	child-PL-NOM

'The parents love the children'

(135) Kabardian (Kumakhov & Vamling 2009: 70) (gloss adapted from source)

fəz-χe-m	*bostej-χe-r*	*Ø-ja-də-r*
woman-PL-ERG	dress-PL-ABS	3.ABS-3PL.ERG-sew-PRS

'The women are sewing dresses'[80]

[78] Glossed and transliterated from *ккалккансархен*, translated from Russian '[О]т дерева по направлению вниз'.

[79] Glossed and transliterated from *хъалеркIен*, and translated from Russian '[И]з дома, по направлению вверх'

[80] Translated by Kumakhov & Vamling as 'The women sewed the dresses' while the verb is glossed as present tense.

(136) Lezgian (Haspelmath 1993: 340) (gloss adapted from source)
itim-ri *čül-ler-a* *iji-zwa-j* *wiri* *kʷ'alaχ-ar*
man-PL.ERG field-PL-IN.ESS do-IPFV-PST all work-PL
dišehli-jri-n *χiw-e* *hat-na*
woman-PL-GEN neck-IN.ESS fall-AOR
'All the work that the men used to do in the fields fell on the women'

(137) Hinuq (Forker 2013: 467) (original glossing, retransliterated)
obu-be *bitłe-be* *r-uː-ho* *b-iči-š*
father-PL house-PL NH.PL-do-IPFV.CVB H.PL-be-PST
'The fathers were building houses'

In contrast, nominal affixes only encoding plurality are found in all five language families of the Caucasus, particularly modern Kartvelian (Boeder 2005: 14), Circassian (Arkadiev & Lander 2020: 393) and Turkic (Berta & Csató 2022: 324), while Nakh-Dagestanian languages predominantly differentiate between the absolutive plural suffixes and the various oblique suffixes to form all other plural cases (Ganenkov & Maisak 2020: 101). In some Nakh-Dagestanian languages the oblique plurals are formed by changing the last vowel of the absolutive plural suffix, e.g. the Lak absolutive suffixes *–nu*, *-lu* and *–du* become the oblique forms *–na-*, *-la-* and *–da-*, while there are also irregular oblique plurals such as the absolutive *–ru* can either become *-irtːa-* or *–irda-* depending on the noun (Friedman 2020: 210).

5.2.1. Numeral functions

Northwest Caucasian language also have numeral nominal affixes to indicate the lower numerals 'one' to 'ten', either as prefixes in Abkhaz, Abaza and Ubykh (Arkadiev & Lander 2020: 402) or as suffixes in Circassian (Rogava & Keraševa 1966: 82-83; Smeets 1984: 233; Kumaxov 2006: 144; Konuk 2022: 152). It is probably more appropriate to analyse these numeral affixes as instances of derivation or numeral incorporation, as Arkadiev & Lander (2020: 402) describe these constructions as compounds, and the singular agreement was more common in Ubykh, as the plural agreement was optional (Fenwick 2011: 91), (cf. example 138a and 138b).

(138) Ubykh (Fenwick 2011: 91) (original glossing)
a. *jɨ-t'qʷ'ʒ́-miz-ʈʂ'ɜ* *Ø-sɨ́-Ø-miɕ[ɜ]-ɜwːt*
this-two-child-good 3SG.ABS-1SG.ERG-CAUS-read-FUT.II
'I will make these two good children study'
b. *jɨ-t'qʷ'ʒ́-miz-ʈʂ'ɜ* *Ø-z-ʁʒ́-miɕɜ-n[ɜ]-ɜwːt*
this-two-child-good 3SG.ABS-1SG.ERG-CAUS.PL-read-PL-FUT.II
'I will make these two good children study'

(139) Kabardian (Colarusso 1992: 49)
(gloss adapted from source, retransliterated)
pɕaaɕa-əj-p'tɬ'ə-r *ma-a-də+a-ha(-r)*
girl-NUM-four-ABS 3PL.ABS-PRS-sew+ITR-PL(-PRS)
'The four girls are sewing'

5.3. Definiteness-marking functions

Definiteness is a grammatical category that primarily encodes the *identifiability* of nominal referents (Lyons 1999: 277-278). In the Caucasus it is only systematically expressed by means of affixation in the Northwest Caucasian languages (Arkadiev & Lander 2020: 388-390), Eastern Armenian (Dum-Tragut 2009: 102; Belayev 2020: 588) and Classical Armenian (Meillet 1936: 88; Schmitt 2007: 120).

(140) Eastern Armenian (Dum-Tragut 2009: 323) (gloss adapted from source)
vardan-ə *stip-um* *ē* *ir* *professor-i-n*
Vardan-DEF force-PTCP.PRS COP.3SG his professor-DAT-DEF
čanač^{h}-el *iren* *orpes* *karewor* *gitnakan*
recognise-INF himself.DAT as important scholar
'Vardan makes his professor recognise him as an important scholar'

(141) Kabardian (Kumakhov & Vamling 2009: 21) (reglossed, retransliterated)
a. *fəzə-r* *q^{w}'aže-m* *Ø-k^{w}'-a-ɕ*
woman-ABS.DEF village-OBL.DEF 3SG.ABS-go-PFV-IND
'The woman went to the village'
b. *fəz-Ø* *q^{w}'aže-Ø* *Ø-k^{w}'-a-ɕ*
woman-ABS.INDEF village-OBL.INDEF 3SG.ABS-go-PFV-IND
'A woman went to a village'

There are even indefinite affixes in Abkhaz and Abaza (Lomtatidze & Klychev 1989: 98; Chirikba 2003a: 23; Arkadiev & Lander 2020: 390), which is typologically rare and it could be a counterargument to Lyons, who questions whether true indefinite-marking exists at all (Lyons 1999: 89). The Abkhaz-Abaza indefinite suffix *-k'* is also obligatory with all numerical prefixes, while O'Herin glosses it as a quantifier (O'Herin 2020: 459).

(142) Abkhaz (Chirikba 2003a: 22) (my glossing, retransliterated)

a. *a-ɥnə́*
DEF-house
'(a/the) house'

b. *ɥnə-k'*
house-INDEF
'a house'

c. *ɥən-kʷá-k'*
house-PL-INDEF
'some houses'

The Ubykh 'indefinite article' *zɜ-* is identical to the numeral 'one' (Fenwick 2011: 45), which means that it should rather be analysed as a numeral prefix, see section 5.2.1. Most languages of the Caucasus lack definiteness completely, and the closest thing to definite-marking affixes in the Nakh-Dagestanian languages is possibly the definite particle *–so* in Khwarshi (Khalilova 2009: 256). North Azerbaijani (Širaliev 1971: 45), South Azerbaijani (Dehghani 2000: 145), Talysh (Schulze 2000: 17), Tat (Suleymanov 2020: 250) and Juhuri (Authier 2012: 47) only indicate definiteness on direct objects, which is a feature they share with Turkish (Göksel & Kerslake 2005:156) and Persian (Yousef 2018: 40).

(143) South Azerbaijani (Dehghani 2000: 145, 167) (gloss adapted from source)

a. *ali* *bir* *alma* *al-dı*
Ali one apple buy-PST.3SG
'Ali bought an apple'

b. *ali* *alma-nı* *al-dı*
Ali apple-ACC buy-PST.3SG
'Ali bought the apple'

(144) Şirvan Tat (Suleymanov 2020: 96, 250) (gloss adapted from source)

a. *män* *korda* *yoft-um*
1SG knife find.PST-1SG
'I found a knife / (some) knives'

b. *in* *korda=ra* *sän* *in=ä*
this knife=OBL PROSP this=OBL
ä *däs=i* *b-ustun-um*
from hand=3SG.POSS MOD-get-1SG
'I am going to snatch this knife out of his hand'

5.4. Possessive functions

Possessive affixes and clitics indicate the possessor of a possessed noun, thus expressing the same grammatical function as possessive pronouns in other languages (Johanson 2022a: 33). It is a grammatical category that is widely found in Northwest Caucasian (Arkadiev & Lander 2020: 389), Turkic (Johanson 2022a: 33-34) and all Caucasian Indo-European languages (Schulze 2000; Dum-Tragut 2009; Authier 2012; Erschler 2020; Suleymanov 2020: 98). It is one of few grammatical categories that appear to be completely absent in Nakh-Dagestanian, and it is generally not found in Kartvelian with the exception of Laz, which distinguishes all persons with possessive suffixes (Lacroix 2009: 75). The enclitic 1st person possessive *–čem* and the suffixed 2nd person possessive *–šen* can also attach to certain kinship terms in Georgian, e.g. *mama-čem-s* 'my father (dative)' and *deda-šen-i* 'your mother (nominative)' (Hewitt 1995: 202, 558).[81]

The Circassian languages and Ubykh stand out in this category, as they have a full three-person distinction in singular and plural, while they also tend to have reciprocal and relative possessive affixes, e.g. Kabardian (Kumakhov & Vamling 2009: 26), Abzakh Adyghe (Paris 1989: 174) and Shapsug Adyghe (Smeets 1984: 383). Fenwick uses the term 'dyadic possession' for a reciprocal possessive prefix in Ubykh (Fenwick 2011: 51).

(145) Adyghe (Rogava & Keraševa 1966: 68) (my glossing)
a. *sə-ne*
1SG.POSS-eye
'my eye'
b. *wə-ne*
2SG.POSS-eye
'your eye'

(146) Kumyk (Abdullaeva et al. 2014: 245) (my glossing)
biz-in muallim-ibiz bir jaxšï adam
1PL.GEN teacher-1PL.POSS one good man
'Our teacher is a really good man'[82]

[81] A similar pattern is also found in Megrelian according Revaz Tchantouria (p.c.).

[82] Glossed and transliterated from *Бизин муаллимибиз бир яхшы одам*, and translated from Russian 'Наш учитель - очень хороший человек'.

(147) Laz (Lacroix 2009: 501, 678) (reglossed)

a. *hemu-k* *cumal-epe-çkimi-s* *oşkui*
DEM-ERG brother-PL-1SG.POSS-DAT apple
n-u-xir-u
PV-IOV[83].O3-steal-AOR.3SG
'He stole the apple from my brothers'

b. *bere-pe-muşi-k* *i-bgar-nan*
child-PL-3SG.POSS-ERG SV[84]-cry-PRS.3PL
'His/her children are crying'

As per the Indo-European languages, Eastern Armenian has possessive affixes or clitics for the 1st and 2nd person singular and plural (Dum-Tragut 2009: 113) whereas Iron Ossetic, Talysh, Tat and Juhuri have a full three-person distinction in both singular and plural (Schulze 2000: 18, 23; Authier 2012; Erschler 2020: 648; Suleymanov 2020: 98). The Iron Ossetic possessive proclitics are peculiar, as they are the only non-Northwest Caucasian possessive affixes or clitics to precede the noun. Possessive affixes or clitics do not appear to have been present in Classical Armenian (Meillet 1936; Van Damme 2004), which is not surprising as Indo-European languages typically employ possessive pronouns instead.

(148) Eastern Armenian (Dum-Tragut 2009: 254) (reglossed)
hima *kardal-u* *em* *hodvac-d*
now read-PTCP.FUT COP.1SG paper-2SG.POSS
'I will read your paper now'

(149) Iron Ossetic (Erschler 2020: 673) (gloss adapted from source)
jɐ=mad-ɐn *nikʷə* *nisə* *ʒaχt-a*
3SG.POSS=mother-DAT never nothing say.PST-PST.3SG
'She never told anything to her mother'

(150) Şirvan Tat (Suleymanov 2020: 98) (gloss adapted from source)

a. *biror=män*
brother=1SG.POSS
'my brother'

b. *xune=šmun*
house=2PL.POSS
'your house'

[83] Glossed as VAL3 'opérateur de valence 3'.

[84] Lacroix labels the prefix *i-* as 'opérateur de valence 2', while the term 'subjective version' is used in previous literature (Lacroix 2009: 456-457).

5.5. Gender or noun class functions

Gender or noun class is a grammatical category that divides the nouns of a language 'into classes of nouns reflected in the behavior of associated words' (Hockett 1958: 231), where the core gender assignment is typically based on semantic criteria (Corbett 1991: 8). Gender is defined, determined and realised by *gender agreement* (Corbett 1991: 105), and although gender is by definition connected to the noun, it is often not realised on the noun itself. It is instead rather realised on e.g. demonstratives, adjectives, articles, numerals and verb agreement (Corbett 1991: 106-110). In the Caucasus, gender or noun class is only found in the Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 100) and in Abkhaz and Abaza (Lomtatidze & Klychev 1989; Chirikba 2003a).

Nakh-Dagestanian genders are mainly realised as verbal and adjectival affixes, but there are also examples of gender agreement on case forms (Ganenkov & Maisak 2020: 100). The noun classes of the Caucasus generally distinguish masculine/feminine, humanness and animacy (Ganenkov & Maisak 2020: 100; O'Herin 2020: 464), but the semantic division in the larger systems is notoriously opaque (Nichols 2011: 142-144; Hauk 2020: 40). There are some vestiges of the Iranian gender system left in the feminine vocative in Northern Talysh (Schulze 2000: 17), but gender was seemingly already lost in Parthian and Middle Persian (Durkin-Meisterernst 2014). Grammatical gender was also already lost in Classical Armenian (Meillet 1936: 92).

Table 5.6: The general criteria for gender or noun class assignment observed in the data for semantic and formal categorisation of Nakh-Dagestanian genders into noun classes I to VI (Friedman 2020: 208; Forker 2020b: 429; Chumakina 2020: 286; Komen, Molochieva & Nichols 2020: 324).

Semantic criteria, Singular	**Semantic criteria, Plural**	**Formal criteria, Singular & Plural**
Human Masculine (I)	Human (I & II)	Generally semantic I & II
Human Feminine (II)		
Non-Human Animate (III)	Non-Human (III & IV)	Opaque III
Inanimate (IV)		Opaque IV
-	-	Opaque V
		Opaque VI

Almost all Nakh-Dagestanian languages have some sort of a gender or noun class system, typically based on semantic criteria (Corbett 1991: 24), but gender assignment in the largest Nakh-Dagestanian gender systems is essentially opaque (Nichols 2011: 144; Ganenkov & Maisak 2020: 100), cf. table 5.6. I therefore suggest that it is relevant to differentiate between gender assignment based on semantic criteria and formal criteria, while recognising that the semantic systems are formally related to the opaque formal systems. Gender or noun class has only been lost in a handful of Lezgic languages, i.e. Lezgian (Haspelmath 1993), Aghul

(Magometov 1970), Udi (Schulze 1982; Ganenkov 2008) and a few dialects of Tabasaran (Ganenkov & Maisak 2020: 100).

Bats has been described as having the largest noun class system in the Nakh-Dagestanian family, as it has potentially eight genders (Ganenkov & Maisak 2020: 100). Only five of these genders are considered to be 'full grammatical genders' while the remaining three genders are *inquorate genders* (cf. Corbett 1991: 170-175), as they only occur with a limited number of nouns (Hauk 2020: 39-40). If these inquorate genders are disregarded, the largest systems distinguish up to six genders, which is the case in e.g. Chechen (Nichols 1994), Ingush (Nichols 2011) and Gagatli Andi (Salimov 2010).

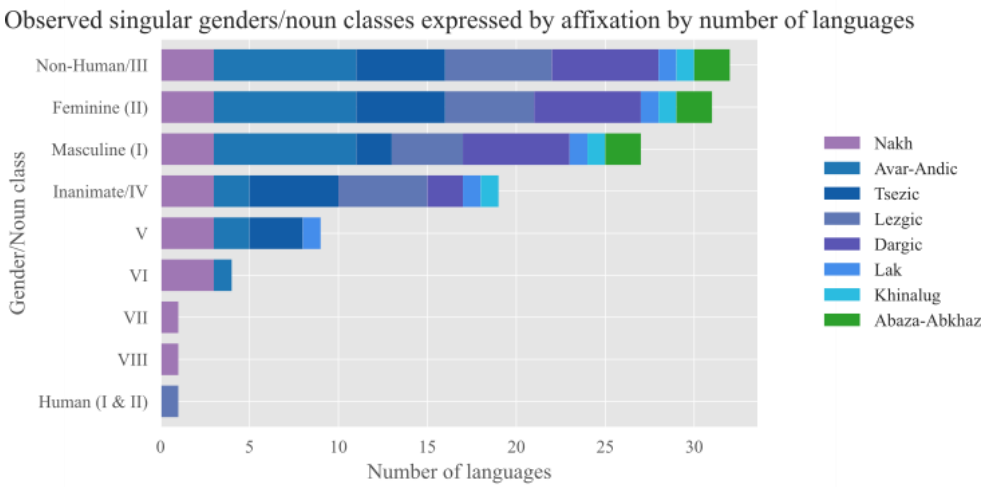


Figure 5.6: All observed singular genders or noun classes expressed through affixation in the data according to number of languages. The gender categories are tentatively based on both semantic and formal categorisation.

In the plural there is a general pattern in both Nakh-Dagestanian and Abkhaz-Abaza of merging the human masculine and human feminine classes into one joint human plural, which has also happened in the singular for Tabasaran (Babaliyeva 2013: 164-165). The gender system in Tabasaran has been reduced to a two-way distinction between human and non-human (Ganenkov & Maisak 2020: 100; Babaliyeva 2013: 164-166), which explains why the non-human gender is the most common in the data, as the difference between human feminine and non-human is just one language, i.e. Tabasaran, which lacks both the human feminine and masculine genders. Human feminine affixes are more common than human masculine affixes in the Caucasus, as certain languages, particularly the Tsezic languages, indicate the singular human masculine with zero-marking (Imnaišvili 1963: 43). The various genders and noun classes of the Caucasus are presented briefly below.

The Human Masculine gender or noun class I

The human masculine gender (M) or noun class I is found in all Nakh-Dagestanian languages with grammatical gender except Tabasaran (Babaliyeva 2013: 164-166), and it typically also includes non-human male *sentient beings* such as male gods and male spirits (Friedman 2020: 208). The human masculine gender is also found in Abkhaz and Abaza (O'Herin 2020: 464). It is even possible for participles to agree with both the agent and the patient in relative constructions in Tindi (Ganenkov & Maisak 2020: 122), cf. example (152).

(151) Avar (Alekseev et al. 2012: 94) (my glossing)
musa *łik'a-w* *χuru-qχːan* *w-ugo*
Musa(M) good-M field-AN M-COP
'Musa is a good fieldworker (lit. field-er)'[85]

(152) Tindi (Magomedova 2012: 176) (my glossing)
[*bac'a* *b-ixʲːu-w*] *hekʷ'a*
wolf.ABS NH.SG-catch-M.SG man.ABS
'A man (M) who caught a wolf (NH)'[86]

(153) Abkhaz (O'Herin 2020: 465) (my glossing)
di-j-ba-t'
3SG.H.ABS-3SG.M.ERG-see-DYN.AOR
'He saw him/her'

Zero-marking is generally used to mark the singular human masculine class in all Tsezic languages (Van den Berg 1995: 79; Khalilova 2009: 42; Forker 2013: 189; Comrie, Xalilov & Xalilova 2015: 214) and Chamalal (Magomedova 2004: 42).

(154) Bezhta (Comrie, Xalilov & Xalilova 2015: 284) (my glossing)
a. *Ø-uqχ'o* *öžö*
I.SG-big son
'older son'
b. *b-uqχ'a-jo* *öž-dä*
I.PL-big-PL son-PL
'older sons'[87]

[85] Glossed and transliterated from *Муса лъикIав хурухъан вуго*, and translated from Russian 'Муса хороший полевод есть'.

[86] Glossed and transliterated from *бацIá бихьʼӱ/в гьекIва*, and translated from Russian 'волка поймавший (I гр. кл.) человек'.

[87] Both examples glossand and transliterated from *Ø-укъо оьжоь* and *б-укъайо (// б-укъа) оьждаь*, and the first example is translated from Russian 'старший сын'.

The Tsezic language Bezhta has a set of infixes that differentiate the singular human masculine from the human feminine and non-human, e.g. human masculine *g-o-wal*, human feminine *g-i-jal* and non-human *g-u-wal* 'come' (Comrie, Xalilov & Xalilova 2015: 216). The human masculine is almost zero-marked in Budukh, as it is only marked with the infix *–r-* in the imperative and terminative, which is the same infix used for the singular human feminine (Alekseev 1994b: 277; Talibov 2007: 170).

The Human Feminine gender or noun class II

The human feminine gender (F) or noun class II is also found in all Nakh-Dagestanian languages with grammatical gender except Tabasaran (Babaliyeva 2013: 164-166), and the human feminine gender is also found in Abkhaz and Abaza (O'Herin 2020: 464).

(155) Hinuq (Forker 2013: 297) (gloss adapted from source)
[*y-eti-n y-ese-yo-me*] [*y-iɣ-no*] *kur-o de!*
II-want-CVB II-be.probable-COND-NEG II-take-CVB throw-IMP 1SG
'If you do not love me (fem.), take me and throw me away!'

(156) Abaza (O'Herin 2002: 140) (gloss adapted from source)
jə-l-rə-tə-t'
3SG.OBL-3SG.F.ABS-3PL.ERG-give-DYN
'They gave it to her'

The human feminine differs from the human masculine in some languages as it does not always cover all female human referents. In Lak, Friedman defines it as covering *mature female sentient beings* (Forker 2020b: 208), as e.g. the word *duš* 'girl, daughter' belongs to the non-human gender or class III (Friedman 2020: 241). Similarly, Authier describes the corresponding feminine gender in Budukh as 'human adult feminine' while all 'non-adult human females' belong to what he labels the animate gender, which is formally identical to the non-human animate (Authier 2010: 145).

(157) Mehweb (Dobrushina 2019b: 133, 155) (original glossing)
a. *w-aš-e-ca hečʼ xunul ʔaʿχ r-aqʼ-as*
M-go:IPFV-IMP-PTCL that.higher woman good F-do:PFV-INF
'Let's help that woman'
b. *urši w-aqʼ-a-la ħu-ni d-aqʼ-a dursi!*
boy M-do:PFV-IRR-APPR you.SG-ERG F1[88]-do:PFV-IMP.TR girl
'[I am afraid that] you give birth to a boy, [better] give birth to a girl!'

[88] *F1-* indicates the unmarried human feminine gender (Daniel 2019: 74-75).

A separate pattern of gender agreement is used for unmarried women and girls in Mehweb (157), in which the singular agreement affixes are formally identical to the plural non-human agreement affixes (Daniel 2019: 74-75).

In some languages the gender affix can attach directly to the noun to differentiate male and female human referents, e.g. Tanti Dargwa *w-eʕʔ* 'male proprietor (I)' and *r-eʕʔ* 'female proprietor (II)', while the same noun can sometimes change gender agreement to indicate male or female humans (Sumbatova 2020: 151).

(158) Dargwa (Sumbatova 2020: 151) (my glossing)

a. *direktur* *w-ač'ib*
director M-come-PRET
'the (male) director came'

b. *direktur* *r-ač'ib*
director F-come-PRET
'the (female) director came'

The Non-Human animate gender or noun class III

The non-human animate gender (NH) or noun class III gender typically includes animals but also inanimate nouns such as e.g. 'sun' and 'moon', body parts, fruits, household objects and certain food terms (Imnaišvili 1963: 44; Murkelinskij 1971: 63; Alekseev & Radžabov 2004: 139; Talibov 2007: 55-57). Animacy should therefore be understood within the cultural contexts of the specific languages, and not animacy in *sensu stricto*. The non-human gender is found in all semantically defined gender systems of the Caucasus, as it is also found in the Tabasaran human/non-human gender system (Babaliyeva 2013: 164-166) and in Abkhaz and Abaza (O'Herin 2020: 464).

In Nakh-Dagestanian languages with only three genders the non-human gender is sometimes referred to as the *neuter* gender (Ganenkov & Maisak 2020: 100; Forker 2020b: 249), which is not optimal since the Indo-European neuter gender is typically not used for animals (Carling 2019: 199-200). The non-human gender can indicate animacy in languages where there are more than three genders or classes, as the typical four-gender system has a distinction between non-human animate (III) and inanimate (IV), e.g. Tsez (Imnaišvili 1963: 44; Alekseev & Radžabov 2004: 139), and Lak (Murkelinskij 1971: 62-63; Friedman 2020: 208).

(159) Lak (Friedman 2020: 234) (reglossed)

buːta-l	*čʷu*	*b-a<w>χː-unu*	*b-u-r*
father.OBL-ERG	horse(III)	III-sell<III>-PTCP.PST	III-COP-3SG

'Father sold the horse (apparently)'

(160) Tindi (Magomedova 2012: 204) (my glossing)

ħaži,	*ogʲa<b>á-w*[89]	*itɬːí-b*	*amaχa*	*χu-ri*
Haji,	there<NH>-M	our.INCL-NH	donkey	field-INTER.LAT

b-etɬː-ṓ
NH-go-UWPST
‘Haji, look, our donkey has got into the field’[90]

(161) Abkhaz (O’Herin 2020: 463, 465) (adapted from source gloss)

a.

a-tʃɨ	*awaʔa*	*j-k’a-ħa-t’*
DEF-horse	there	3NH.SG.ABS-DE-fall-AOR.IND.DYN

‘The horse fell (down) there.’

b. *(dara)*
3M.SG
d-rɨ-la-na.gala-jt’
3H.SG.ABS-3PL.APPL-among-3NH.SG.ERG-bring-AOR.IND.DYN
‘He found himself among them (lit. ‘It brought him among them.’)’

The Inanimate gender or noun class IV

The inanimate gender (INAN) or noun class IV is found in a dozen Nakh-Dagestanian languages, and it typically includes inanimate and abstract nouns covering particularly materials, liquids, temporal nouns, clothes and certain body parts (Imnaišvili 1963: 44; Murkelinskij 1971: 63-64; Talibov 2007: 57). The inanimate gender or noun class IV is only found in the Nakh-Dagestanian languages in the Caucasus, e.g. Gagatli Andi (Salimov 2010: 47, 179-181), in the Lezgic languages Rutul (Alekseev 1994a: 229), Aghul (Kibrik 1994a: 308), Kryts (Authier 2009: 141-142), Budukh (Talibov 2007: 57) and Archi (Kibrik 1994a: 308; Chumakina 2020: 286). Noun class IV has a zero-prefix in languages such as Archi (Chumakina, Bond & Corbett 2016: 23-25) and Budukh (Talibov 2007: 57), and in Tsakhur both noun class I and IV have a zero-prefix in the singular, which means that these two classes are only differentiated when the class markers are infixes or suffixes (Schulze 1997: 46).

[89] Magomedova describes this as a particle which draws the attention of a male interlocutor to an animal or an object (Magomedova 2003: 272).

[90] Glossed and transliterated from *ХIажи, ог'абáв илIúб амаха хури белIṓ*, Translated from Russian ‘Гаджи, вон наш осел забрался на пашню’.

(162) Tsakhur (Schulze 1997: 33, 62, 66) (gloss adapted from source)

a. *za-k'le* *Ø-ats'a* *def*
1SG-AFFT IV-know.PRS NEG
'I do not know (it)'

b. *džamalij,* *Ø-ats'a-χe-ne* *t'ak'u* *bazara-qa*
Džamali I-know-COND-Q Taku(I).ABS bazaar-ALL
uˁqˁ-aˁs *hˁaˁzər-qa-je*
go-INF ready-AUX-Q
'Džamali, do you know whether Taku is ready to go to the bazaar?'

c. *jed-ik'le* *urus* *miz* *w-ats'a* *wo-b*
mother-AFFT Russian language(III).ABS III.know.PRS AUX.PRS-III
'Mother knows Russian'

The inanimate gender or noun class IV is also found in the Dargic languages Xaidaq (Temirbulatova 2004: 60, 63) and Mehweb (Magometov 1982: 76), in the Tsezic languages Tsez (Imnaišvili 1963: 44; Alekseev & Radžabov 2004: 139), Khwarshi (Khalilova 2009: 42) and Hinuq (Forker 2013: 189), in Lak (Murkelinskij 1971: 64; Friedman 2020: 208) and in Khinalug (Kibrik 1994b: 387-388; Khvtisiashvili 2013: 51).

(163) Khwarshi (Khalilova 2009: 307) (original glossing)
q'ebed-qo *l-ogu* *tir* *l-i-ya* *l-eqʷ-i*
smith-CONT IV-good sabre(IV) IV-do-INF IV-can.WPST
'The smith could make a good sabre'

(164) Lak (Friedman 2020: 239) (adapted from source gloss)
[*butːa-l* *ars-na-n* *d-ull-usːa*] *č'ila*
father.OBL-ERG son-OBL-DAT IV-give.PTCP.PST knife(IV).ABS
'The knife that father gave to his son'

The remaining genders or noun classes

The remaining genders or noun classes can generally not be analysed semantically or functionally (Nichols 2011: 144; Ganenkov & Maisak 2020: 100). There is a tendency to treat young or small animals differently from adult animals in some languages, as e.g. noun class V in Khwarshi covers both inanimate nouns and specifically the young of animals (Khalilova 2009: 42), while small animals and the young of animals belong to noun class IV in Archi (Chumakina 2020: 286). The same phenomenon of assigning the young of animals as inanimate, has also been observed in Indo-European gender assignment (Carling 2019: 199-200). Possibly the only system with more than five classes outside of the Nakh branch in the Caucasus is found in some dialects of Andi, such as Rikwani Andi, where class VI primarily covers insects (and arachnids), e.g. 'butterfly' and 'scorpion' (Salimov 2010: 7).

(165) Andi (Salimov 2010: 113) (my glossing)

χur-u-kːu	*kartuš-ol*	*b-oɢʁ-o*
field(V)-IN-ELA	potato(IV)-PL.ABS	IV.PL-bring-WPST
išba	*imu-di*	
home	father(I).OBL-ERG	

'Father brought home potatoes from the field'[91]

(166) Khwarshi (Khalilova 2009: 138) (adapted from source gloss)

ħalt'i-tł'o-z	*j-ux-šezuq'un,*	*iłelo*	*dandil*	*k'it'-is*
work-SUPER-ABL	II-come-DUR.CVB	that.GEN	towards	cat-GEN
hũho	*j-ux-še*	*j-eč-i*		
kitten(V)	V-come-IPFV.CVB	V-be-WPST		

'When she was coming back from the work, she met a kitten on her way'

The descriptive tradition of the Nakh languages stand out from the other Nakh-Dagestanian languages in that their noun classes are conventionally not described with Roman numerals but with the form of the noun class prefixes (Jakovlev 1960: 198-199; Komen, Molochieva & Nichols 2020: 324). However, e.g. Aliroev (1999: 49) attempts to classify the noun classes in Chechen as I-VI. I will therefore give some examples of Nakh noun classes below, while also acknowledging that the noun classes in the Nakh languages cannot be analysed according to the framework of this thesis. Consider examples (167) from Bats, as they indicate how the nouns 'dog' and 'basket' are categorised together while 'cat' and 'donkey' belong to another class, which is difficult if not impossible to explain by factors such as animacy or size.

(167) Bats (Hauk 2020: 44, 49, 50) (gloss adapted from source)

a. *equs* *daħ* *b-itː* *e* *pħu*
this.one.ERG PV B/D-wash this dog(B/D)
'She [the mother] is washing this dog'

b. *t'at'en* *d-a* *is* *k'uit'* *e* *daħ* *c'em-o-d*
wet D/D-be that.one cat(D/D) and PV clean-PRS-D/D
'That cat is wet, and (she) is cleaning (it)'

c. *oqus* *k'alt-i* *ħal* *qoxk'-d-i-en*
yon.one.ERG basket(B/D)-PL up hang.many-B/D-TR-AOR
vir=mak
donkey(D/D)=SUPER
'They (SG) hung the baskets on the donkey'

[91] Glossed and transliterated from *хурукку картушол бокъго ишба имуди* and translated from Russian '[И]з поля картошки принес домой отец'.

5.6. Copular or predicative functions

Copular or predicative affixes or clitics express the functions of copular verbs, where copulas can be characterised as carrying *verbal inflection*, appearing 'in contexts where the predicate is nonverbal, and used to link the predicate and the subject' (Arche, Fábregas & Marín 2019: 3). Copular affixes typically merge a subject and a predicate or turn an adjective into a predicate rather than linking them. Copular affixes are not very common in the Caucasus, but they do occur, particularly in the Turkic languages (Ragagnin 2022: 249; Berta & Csató 2022: 328). Copular affixes form an interesting category as they constitute an intersection between nominal and verbal morphology, since these affixes attach to nominals while they convey grammatical functions typically associated with verbs, i.e. person and tense (Johanson 2022a: 35-36).

(168) Kumyk (Abdullaeva et al. 2014: 205) (my glossing)
men *doxtur-man,* *sen* *busa* *muallim-sen*
1SG doctor-COP.1SG 2SG but teacher-COP.2SG
'I am a doctor, but you are a teacher'[92]

(169) Juhuri (Authier 2012: 135) (gloss adapted from source)
me *dusd* *en* *biror=tü=düm*
1SG friend GEN brother=2SG.POSS=COP.1SG
'I'm a friend of your brother's'

(170) Standard Dargwa (Van den Berg 2001: 37) (gloss adapted from source)
nu *buk'un-ra*
1SG.ABS shepherd.ABS-COP.1SG
'I am a shepherd'

Most of these languages encode person in their copular affixes, which is the typical situation in both Turkic and Indo-European copular affixation in the region (Authier 2012: 133; Suleymanov 2020: 140; Ragagnin 2022: 249; Berta & Csató 2022: 328). In the Nakh-Dagestanian languages, copular person-marking appears to only be found in Standard Dargwa (Van den Berg 2001: 37), Itsari Dargwa (Sumbatova & Mutalov 2003: 80), Sanzhi Dargwa (Forker 2020a: 162) and Kryts (Authier 2009: 111-112).

The copular clitics in the Kartvelian languages typically only encode the 3rd person singular copula in the present tense, e.g. Georgian *-a* (Hewitt 1995: 561; Bolkvadze & Kiziria 2023: 454) and Megrelian *-re*/*-e* (Harris 1991b: 376), which is also true for Udi (Schulze 1982: 34, 43).

[92] Glossed and transliterated from *Мен дохтурман, сен буса муаллимсен*, and translated from Russian 'Я доктор, а ты учитель'.

(171) Georgian (Bolkvadze & Kiziria 2023: 454) (my glossing)
čemi *da* *ekim-i=a*
1SG.POSS sister doctor-NOM=COP.3SG
'My sister is a doctor'

The number of languages in the Caucasus that encode tense in their copular affixes is quite limited, and they only distinguish between present and past tense. Copular affixes encoding both present and past tense are found in North Azerbaijani (Ragagnin 2022: 253), South Azerbaijani (Dehghani 2000: 121-122), Khinalug (Kibrik 1994b: 388, 396-398), Kryts (Authier 2009: 111-112) and Sanzhi Dargwa (Forker 2020a: 164). The copular affixes in Khinalug and Kryts also agree in gender or noun class (Kibrik 1994b: 398; Authier 2011: 111).

(172) North Azerbaijani (Širaliev 1971: 128) (my glossing)
müəllim-di-m
teacher-COP.PST-1SG
'I was a teacher'

(173) Sanzhi Dargwa (Forker 2020a: 164) (gloss adapted from source)
du *ustːa=de*
1SG master=COP.PST
'I was a master'

5.7. Conjunctive functions

Conjunctive or Additive clause-linking affixes generally correspond to conjunctions in most other languages and they can both be nominal or verbal depending on the word class they attach to. The validity of this category could be questioned however, as the constituents of the group are often described as particles and clitics rather than affixes. The typical conjunctive affix is the coordinate conjunction 'and', cf. the Latin enclitic conjunction *–que*, but in the Nakh-Dagestanian languages the additive enclitic also encodes meanings such as 'also', 'too', 'as well' and 'even' (Forker 2020a: 171). The conjunctive affixes are either obligatory on all nouns of the NP, as in Bezhta (Comrie, Xalilov & Xalilova 2015: 437), or optional in order to e.g. express emphasis as in Lezgian (Haspelmath 1993: 327).

Coordinate conjunctive affixes or enclitics are found in almost all Nakh-Dagestanian, except e.g. Khinalug (Dešeriev 1959: 151-152; Khvtisiashvili 2013: 164), and in most Northwest Caucasian languages. There is also the conjunctive clitic *–c(a)* 'also, too' in Georgian (Hewitt 1995: 88). The Turkic languages of the Caucasus mainly have separate conjunctions, e.g. the borrowed Persian and Arabic conjunctions *va*/*wa* 'and', thereby lacking a conjunctive affix or enclitic

(Abdullaeva et al. 2014: 380-381; Ragagnin 2022: 255; Berta & Csató 2022: 333), whereas South Azerbaijani has also retained the conjunctive suffix *–InAn*, which is identical to the instrumental case suffix (Dehghani 2000: 211).

(174) Ingush (Nichols 2011: 252) (original glossing)
taxan vuuchara wa so k'alxar-vaaqqaragh,
today V.perish.PTCP.NZ.ABL 2SG.ERG 1SG save-V.LV.VN.LAT
aaz shortta axcha=ji soughatazh=ji luddy
1SG.ERG plenty money=& gift.PL=& give.FUT.D
hwuona='a, hwa dottaghazhta='a
2SG.DAT=& 2SG.GEN friend.PL.DAT=&
'Since you saved my life today, I'll give you and your friends plenty of money and gifts'

(175) Adyghe (Rogava & Keraševa 1966: 336) (my glossing)
njepe nurjet-re asłančerije-re ze-fe-gubzə-ʁ-eχ
today Nurijet-ADD Aslančerij-ADD REC-BEN-get.angry-PST.PFV-3PL
'Nurijet and Aslančerij had quarrelled today'[93]

(176) South Azerbaijani (Dehghani 2000: 181) (original glossing)
gız-ınan oglan öp-üš-dü-lär
girl-CONJ boy kiss-REC-PST-3PL
'The girl and the boy kissed each other'

[93] Glossed and transliterated from *Непэ Нурет-рэ Аслъанчэрие-рэ зэфэгубжьыгъзх*, and translated from Russian 'Сегодня и Нуриет и Асланчерий поссорились'.

5.8. Summary of the nominal affixation systems in the Caucasus

The grammatical functions expressed through affixation in the Caucasus can be summarised into various affixation systems. The variation described above is quite remarkable, even within the same families, so I will therefore summarise the observed affixation systems for each grammatical category, before making a typological comparison between the nominal affixation systems of all 56 languages. All systems are categorised on an alphabetical scale, where A is the smallest observed system and zero indicates that the category is either absent or not expressed by affixation. Most of these categories include ranges of optional functions where all functions do not have to be present in the system, as it would become overly detailed to attempt a stricter categorisation that would prevent any further generalisations. This approach could be criticised as being somewhat arbitrary, but the purpose is to tentatively explore and visualise general tendencies in the data.

The core case systems can be summarised into eight categories, where the smallest systems only differentiate between zero-marking and a marked oblique case. The largest observed systems include four core cases, as these systems have a dedicated affective case affix.

Table 5.7: Tentative summary of all core case systems expressed by affixation in the Caucasus

Core case affixation systems by size, where each system type contains any of the listed functions.	
A	1: OBL
B	1: DAT(/GEN)
C	2: ABS, OBL
D	2: ERG, DAT
E	2-3: NOM, ACC, DAT
F	2-3: NOM/ABS, ERG, DAT/ACC
G	3: ERG, OBL, DAT
H	4: ERG, OBL, DAT, AFFT

The situation is much more varied when it comes to the non-core case functions, although the number of tentative categories is lower as there is a hierarchical tendency for certain functions to be completely absent from the smaller systems. The smallest non-core case system only includes a genitive case (which in some languages is identical to the dative case), while almost all larger systems contain the genitive, instrumental, comitative, and/or equative/similative cases.

Table 5.8: Tentative summary of all non-core case systems expressed by affixation or affixed adpositions in the Caucasus

Non-core case affixation systems (incl. affixed adpositions) by size, containing any of the listed functions.	
A	1: GEN
B	2: INS/COM, ADV
C	2-4: GEN, INS, COM, EQU/SIMIL, BEN, VOC
D	3-5: GEN, INS, COM, EQU/SIMIL, COMPR, PART
E	3-6: GEN, INS, COM, EQU/SIMIL, CAUSL, SUBST
F	5-7: GEN, INS, COM, EQU/SIMIL, ADV, BEN, PART, VOC
G	7-9: GEN, INS, COM, EQU/SIMIL, ADV, COMPR, BEN, CAUSL, PART, VOC, CONTENT, SUBST

The orientational local case systems exhibit less variation than the non-core case, as the less frequent functions are only found in the larger systems, while the smaller systems are generally more similar. The largest orientational local case systems are found in Aghul and Chamalal as they distinguish nine and eight orientations respectively.

Table 5.9: Tentative summary of all local case orientation systems expressed by affixation or affixed adpositions in the Caucasus

Local case orientation systems (incl. affixed adpositions) by size, containing any of the listed functions.	
A	2-3: IN, SUPER, AD
B	3-4: IN, SUPER, SUB, AD, INTER
C	5-6: IN, SUPER, SUB, AD, INTER, CONT, APUD, POST, ANTE
D	6-9: IN, SUPER, SUB, AD, INTER, CONT, APUD, POST, ANTE, POSS, CUM

The directional local case systems are somewhat more complicated, as the last category contains the unusual functions distal, proximal, delative and suslative, which are more or less only found in the Dargic languages. All systems would neatly fit within categories A to E if these four directions are disregarded. Category B is perhaps better described as true postpositions and not local case directions.

Table 5.10: Tentative summary of all local case direction systems expressed by affixation or affixed adpositions in the Caucasus

Local case direction systems (incl. affixed adpositions) by size, containing any of the listed functions.	
A	2: Essive(/Lative), Ablative
B	2: Translative, Terminative
C	3: Essive, Ablative, Lative
D	4-5: Essive, Ablative, Lative, Directional, Translative, Terminative
E	6: Essive, Ablative, Lative, Directional, Translative, Terminative
F	5-8: Essive, Ablative, Lative, Directional, Distal, Proximal, Delative, Suslative

The gender systems have been categorised into seven categories, which is primarily motivated by the actual number of genders or noun classes in these languages. Some of these categories could be merged though, e.g. D-E and F-G.

Table 5.11: Tentative summary of all gender or noun class systems in the Caucasus

Gender/noun class systems in the Caucasus by size.	
A	2: Human, Non-human
B	3: Human masculine (I), Human feminine (II), Non-human (III)
C	4: Human masculine (I), Human feminine (II), Non-human animate (III), Inanimate (IV)
D	4: Human masculine (I), Human feminine (II), Opaque III-IV
E	5: Human masculine (I), Human feminine (II), Opaque III-V
F	6: Human masculine (I), Human feminine (II), Opaque III-VI
G	6-8: Human masculine (I), Human feminine (II), Opaque III-VIII

The tentative summary of all nominal affixation systems in the Caucasus (table 5.12) suggests that certain grammatical categories are more stable than others, but also that closely related languages differ to a surprising degree. This could either indicate that there are issues with the underlying data, that the categorisations are problematic or that there actually are considerable differences even between sister languages. It is also remarkable to see how related languages generally belong to the same categories while differing in just one or two categories. This suggests that morphology changes gradually, possibly one function at a time, which in a longer perspective increases the morphological distance as languages diverge.

There are also some areal tendencies, as many unrelated or distantly related languages belong to the same categories. The Kartvelian languages generally belong to non-core case category F, which is also true for Abkhaz and Ubykh but not for Abaza, whereas Abaza belongs to category B together with the Circassian languages. This could indicate that Abkhaz and Ubykh have shifted towards a Kartvelian type system, while Circassian and Abaza have not. There are also some apparent similarities between some Nakh-Dagestanian languages that are likely best explained by geographical vicinity rather than their position in the Nakh-Dagestanian family tree.

The importance of dividing orientational and directional local cases become apparent with this summary, as many languages have rich directional systems while lacking any orientational affixes as previously observed. The Kartvelian languages have a rather stable set of suffixed directional postpositions, while both Georgian and Svan encode orientation as well. The Georgian orientational system shares similarities with both Iron Ossetic, Udi and potentially Eastern Armenian, while the Khinalug system should likely belong to a separate category as it has been described as only encoding the orientation AD.

Table 5.12: Tentative summary of the nominal affixation systems in the Caucasus.

Core	*Non-core*	*LC OR*	*LC DIR*	*NC*	*DEF*	*POSS*	*NUM*	*Languages*
F	F	A	D	0	No	0/2	0	Georgian
F	F	0	D	0	No	0	0	Old Georgian
F	G	0	D	0	No	0/2	0	Megrelian
D	F	0	D	0	No	6	0	Laz
F	F	C	D	0	No	0	0	Svan
G	D	0	C/D	F	No	0	0	Chechen, Ingush
G	G	B	E	G	No	0	0	Bats
G	C	D	D	B	No	0	0	Avar
H	C	C	C	F	No	0	0	Andi
H	C	D	C/D	B	No	0	0	Tindi, Ghodoberi
H	E	C/D	D	B	No	0	0	Bagvalal, Akhvakh
H	C	D	D	E	No	0	0	Chamalal
G	C	D	C	B	No	0	0	Karata
G	C	D	F	C	No	0	0	Tsez
G	G	D	E	C	Yes	0	0	Khwarshi
G	C	D	D	E	No	0	0	Hinuq
G	E/G	D	D	E	No	0	0	Bezhta, Hunzib
G	C/E	C	D	0	No	0	0	Lezgian
G	C	D	D	0	No	0	0	Aghul
G	D	D	D	A	No	0	0	Tabasaran
G	C	A	C	0	No	0	0	Udi
G	D	C	C	C	No	0	0	Rutul
H	C	B	C	D	No	0	0	Tsakhur
G	D	C/D	C/D	C	No	0	0	Kryts, Budukh
G	G	C	D	C	No	0	0	Archi
G	D	A	C	C	No	0	0	Khinalug
G	G	C	D	B	No	0	0	Standard Dargwa
G	E	D	D	C	No	0	0	Mehweb
G	C	D	D	B	No	0	0	Kubachi
G	C	D	F	C	No	0	0	Xaidaq
G	C	C	D	B	No	0	0	Itsari & Sanzhi Dargwa
G	G	D	D	E	No	0	0	Lak
C	B	0	0	0	Yes	6	10	Circassian
A	F	0	D	0	Yes	6	10	Ubykh
0	F	0	D	B	Yes	6	10	Abkhaz
0	B	0	B	B	Yes	6	10	Abaza
B	C	0/A	A	0	Yes	4	0	Eastern Armenian
B/E	C	0	A	0	Yes	0	0	Classical Armenian
E	C	A	C	0	Yes	6	0	Iron Ossetic
A	C	B	A	0	Yes	6	0	Talysh
A	0/C	0	0/A	0	Yes	6	0	Tat, Juhuri
E	C	0	C/D	0	Yes	6	0	S & N Azerbaijani
E	A/C	0	C	0	No	6	0	Kipchak

6. Verbal affixation in the Caucasus

6.1. Tense functions

Tense can be defined as the 'grammaticalisation of location in time' (Comrie 1985) and verbal tense is one of the few categories that is expressed by affixation in all 56 languages of the affixal data set. Tense is conventionally analysed either according to *absolute* or *relative tense*, where the absolute tense 'takes the present moment as deictic' (Comrie 1985: 36) while for relative tense 'the reference point for location of a situation is some point in time given by the context' (Comrie 1985: 56). I have operationalised tense in accordance with Comrie's definition of absolute tense with the traditional three-way distinction between *past*, *present* and *future tense*, while redefining the deictic as the time of the speech act instead of the present moment, where past tense indicates events prior or anterior to the time of the speech act, i.e. the present, and the future tense indicates events posterior to the speech act (Comrie 1985: 36-47). The reference point should be redefined as the time of the speech act, since while it is possible to make absolute time reference to a speech act in spoken language, it is impossible to make absolute time reference to the present moment (Comrie 1985: 36).

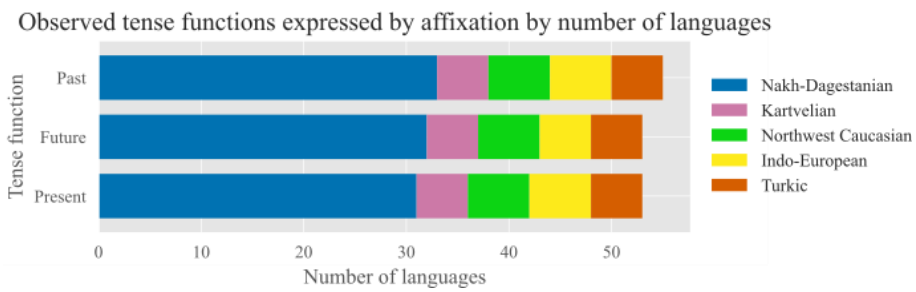


Figure 6.1: All observed tense functions expressed by means of affixation in the data according to number of languages.

This approach therefore does not analyse *perfect* and *pluperfect* as tense categories as they constitute combinations of tense and aspect (Hewson 2012: 508; Ritz 2012: 881). Both Comrie (1976) and Dahl (1985) argue that the perfect should be analysed as a cross-linguistically valid category, based on the perfect functions

of *perfect of result, experiential perfect, perfect of persistent situation* and *perfect of recent past* (Comrie 1976: 56-61; Dahl 1985: 129-133), but Dahl does not list it as one of his tense categories (Dahl 1985: 103-128). A strict absolute temporal interpretation should analyse all functions of perfect given by Comrie referring to past events, as the events are not occurring at the time of the speech act even though the results are felt in the present (Dahl 1985: 129-133).

The rationale for differentiating tense and aspect in perfect and pluperfect can by exemplified by the perfect and pluperfect paradigms in Italian and English, cf. table 6.1. These paradigms also present a compelling case for the analysis of preterite/aorist as an aspect rather than a tense.

Table 6.1: The 3rd person singular forms of perfect, pluperfect and future perfect of the verb 'to speak' in Italian and English.

Tense & Aspect	**Italian**	**English**
Present perfective (Perfect)	*ha* *parla-to* AUX.PRS.3SG speak-PTCP.PFV 'has spoken'	*has* *spoken* AUX.PRS.3SG speak.PTCP.PST
Past perfective (Pluperfect)	*av-ev-a* *parla-to* AUX-PST.IPFV-3SG speak-PTCP.PFV 'had spoken' (past imperfective perfective)	*had* *spoken* AUX.PST speak.PTCP.PST
	ebbe *parla-to* AUX.PST.PRET.3SG speak-PTCP.PFV 'had spoken' (past preterite perfective)	
Future perfective (Future perfect)	*av-rà* *parla-to* AUX-FUT.3SG speak-PTCP.PFV 'will have spoken'	*will* *have* *spoken* AUX.FUT AUX.INF speak.PTCP.PST

Since English does not have a synthetic distinction between the imperfective and preterite aspects, the pluperfect is simply formed with the past tense of the auxiliary. In Romance languages such as Italian and French there is a distinction between pluperfects formed with the imperfect or the preterite (L'Huillier 1999; Maiden & Robustelli 2013), although the preterite pluperfect is only used in formal literary French since the preterite (or *passé simple*) has largely been replaced by the perfect in spoken French (L'Huillier 1999). This demonstrates how the historical three-way aspectual distinction between preterite, perfective and imperfective has been reduced to a perfective/imperfective distinction in contemporary spoken French.

A binary tense contrast between either 'non-past tense' and past tense or 'non-future tense' and future tense is found in many languages of the world (Comrie 1985: 48; Hewson 2012), but 'non-past tense' and 'non-future tense' are categorical concepts rather than functions as they only group two of three tense functions. Comrie categorises Finnish and German as having a binary past/non-past tense system (Comrie 1985: 49), which is problematic as both German and Finnish have periphrastic constructions with unambiguous future reference (Donaldson 2006: 111; Karlsson 2018: 313). Tense forms with multifunctional tense reference are

unproblematic from a functional perspective, as the time referencing function can be carried by other constituents, e.g. temporal adverbs, much in the same way that Turkic languages generally do not allow plural suffixes if numerals or other quantifiers are present in the NP, e.g. Turkish *iki at* 'two horses' (Johanson 2022a: 45). Hewson categorises Russian as having a binary tense system since the future tense is indicated by the present perfective, e.g. *napišet* 'he will write' (Hewson 2012: 510). An identical construction is found in Georgian, where the co-occurrence of preverbs and the present tense suffixes express the future tense, e.g. Georgian present tense *v-xat'-av* 'I'm painting' and future tense *da-v-xat'-av* 'I will paint' (Vamling 1989: 20). I would analyse these constructions as instances of future tense, although the Russian prefix *na-* and the Georgian *da-* also convey perfectivity in other constructions.

The analysis of future tense in Georgian and Megrelian is complicated by the fact that they indicate future reference by attaching preverbs to the present tense forms (Tuite 1998: 16; Rostovtsev-Popiel 2020: 551), which means that the preverbs express a future function, but it would be misleading to label them as future tense prefixes.

The three-way tense distinction between past, present and future is morphologically motivated as all five language families generally distinguish between these tenses, either by means of affixation or periphrasis. All languages in the data have at least one past tense affix and all but two languages have a future tense affix, as it is absent in Bezhta (Comrie, Khalilov & Khalilova 2015) and Tat expresses future tense by means of the prospective particle (Suleymanov 2020: 164)

(177) Budukh (Talibov 2007: 211) (my glossing)

vɨn	*ala*	*daʁ-dž-e*	*ča<ma>ʁar,*
2SG	DEM	mountain-OBL-LOC.LAT	go<PROH>.MSD,
ije-r	*vɨn*	*jolkol-a*	
there-ELA	2SG	roll.MSD-FUT	

'Don't go up that mountain, you will fall down (lit. roll down) from there'[94]

Present tense affixes are found in 53 languages, which means that the remaining languages either express present tense by periphrastic constructions or that the present tense is zero-marked. There is substantial overlap between the tense functions and various aspectual, modal, evidential and person-marking functions, which is a well-known phenomenon around the world, i.e. TAM. Since many languages of the Caucasus have agglutinative affixation patterns (Ganenkov & Maisak 2020: 112), verbal affixes that seemingly express only tense and not aspect nor mood are quite frequent. A small number of languages in the Caucasus have separate present tense affixes, but it is often difficult to make a clear-cut distinction

[94] Glossed and transliterated from *Вын ала дагъдже чамагъар, ийер вын йолкола*, and translated from Russian 'Ты на ту гору нс ходи, оттуда ты сорвешься'.

between imperfective/progressive affixes and present tense affixes, e.g. the Lezgian suffix *–z(a)wa* which Haspelmath describes as an imperfective suffix, while all examples indicate a present tense function (Haspelmath 1993: 140).

Similarly, the Kartvelian languages have a set of suffixes either labelled *thematic suffixes* or *present(-future) stem formants* (Aronson 1990: 40; Hewitt 1995: 118; Bolkvadze & Kiziria 2023: 196). These thematic suffixes could also be analysed as expressing an imperfective aspect (Aronson 1990: 44), as they can combine with the suffix *–(o)d* to form the past imperfective or imperfect tense (Bolkvadze & Kiziria 2023: 208-209; Öztürk & Pöchtrager 2011: 71-72), but they indicate present tense when they do not combine with any non-spatial preverbs or other TAM affixes (Bolkvadze & Kiziria 2023: 166). Subsequently, affixes that simultaneously express present tense and imperfective aspect reiterate the claim by Bybee, Perkins and Pagliuca (1994) that an 'imperfective restricted to the present is simply a present' (Bybee et al. 1994: 126), which makes it counterintuitive not to analyse present imperfective affixes as instances of present tense. The presence of present tense affixes in the Northwest Caucasian languages is complicated by the distinction between dynamic and static verbs, as the prefixes *e-*/*me-* in the Circassian languages also convey present tense, while being analysed as only a dynamic prefix by e.g. Arkadiev & Lander (2020: 418-421).

(178) Shapsug Adyghe (Smeets 1984: 86, 198, 221) (reglossed)
- a. *s-jə-ɬeʁʷə-ʁ*
  1SG.ABS-3SG.ERG-see-PST.PFV
  '(s)he saw me'
- b. *s-j-e-ɬeʁʷə*
  1SG.ABS-3SG.ERG-DYN.PRS-see
  '(s)he sees me'
- c. *s-jə-ɬeʁʷə-ʃt*
  1SG.ABS-3SG.ERG-see-FUT
  '(s)he will see me'

I have instead followed Paris (1989; 188) and Konuk (2022: 223) by analysing the prefixes *e-/me-* and the suffix *–re*/*-rɜ* as dynamic present tense affixes, as they always occur with dynamic verbs in the present tense, and they are absent in the past and future tenses, cf. table 6.2.

Table 6.2: Tense distinction in Shapsug Adyghe (Smeets 1984: 86, 198, 221)

1SG.ABS	3SG.ERG	DYN.PRS	Root 'see'	Tense suffix	Translation
s-	*-jə-*	*-Ø-*	*-ɬeʁʷə-*	*-ʁ* (PFV)	'(s)he saw me'
s-	*-j-*	*-e-*	*-ɬeʁʷə-*	*-Ø* (PRS)	'(s)he sees me'
s-	*-jə-*	*-Ø-*	*-ɬeʁʷə-*	*-ʃt* (FUT)	'(s)he will see me'

(179) Abzakh Adyghe (Konuk 2022: 472) (gloss adapted from source)

wɜ	*ɐdəgɐ-bzɜ-t͡ʃʔɜ*	*wə-zə-gʷəʃəʔɜ-rɜ-m*	
2SG	Adyghe-language-INS	2SG.ABS-TEMP[95]-speak-DYN.PRS-OBL	
ɐ-xɜ-mɜ	*jɐ-gʷɐpɜ*	*mɜ-χʷə*	
DEM.DIST-PL-OBL.PL	3PL.POSS-pleased	DYN.PRS-be	

'When you speak Adyghe, they are pleased'[96]

The General or Generic Present tense?

Apart from the three tenses mentioned above many Dagestanian language have been described as having a *general*, *general present* or *generic present* tense that refers to events 'that lack a concrete temporal reference' and typically have a habitual meaning (Forker 2020b: 254; Chumakina 2020: 299). The general tense is often labelled as a 'simple present' as it has the same function as the simple present in English (Forker 2020b: 254), but some of these forms should probably rather be analysed as habitual aspect, which is formally supported by the Avar simple present *–ula* forming the past habitual *–ula-ʔan* (Forker 2020b: 254-255). There is a functional difference between true habituals and *generic* statements (Carlson 2012: 830-831), e.g. 'snow is usually cold' and 'snow is cold' are not synonymous, while both lack a concrete temporal reference. I therefore analyse the general tense affixes as having either habitual or generic aspect, depending on the apparent function.

The Aorist tense?

The term aorist is used for various tense forms in Kartvelian (Testelets 2020: 508-509), Turkic (Johanson 2022a: 39) and Nakh-Dagestanian (Ganenkov & Maisak 2020: 114), although it is problematic from a typological and functional perspective as the term refers to different phenomena in the different families. In Kartvelian languages the aorist generally refers to past events that are either perfective or imperfective depending on the presence of a preverb (Bolkvadze & Kiziria 2023: 237). In North Azerbaijani, the aorist *signals disposition* with habitual and future reference (Ragagnin 2022: 251), while in Avar and other Nakh-Dagestanian languages it refers to the past perfective (Forker 2020b: 255; Ganenkov & Maisak 2020: 114). I have therefore chosen to analyse aorist as an aspectual category and not a tense, cf. section 6.2 below.

[95] Konuk glosses it as 'soi_même', while acknowledging that it possibly belongs to a different morphosyntactic category that expresses temporality (Konuk 2022: 472). I analyse this as a temporal converb, cf. the identical prefix *zə-* with a temporal converb function in Shapsug Adyghe (Smeets 1984: 254) and the Temirghoy Adyghe prefix *z-* Arkadiev & Lander gloss as 'REL.TEMP' (Arkadiev & Lander 2020: 385, 406).

[96] Translated from French 'Quand tu parles adyghé, ils sont contents'.

6.2. Aspectual functions

Aspect is a grammatical category that could be defined as 'different ways of viewing the internal temporal constituency of a situation' (Comrie 1976: 3), which is not an optimal definition as aspect does not *per se* specify any temporal information, since the temporal reference is conveyed by its interaction with tense (De Swart 2012: 753). Although the grammatical categories of tense and aspect are intrinsically connected, evidence from languages across the world indicates that tense and aspect should be analysed as autonomous grammatical categories (Gvozdanović 2012: 791). Comrie postulates three aspectual oppositions, i.e. *perfective/imperfective*, *habitual/continuous* and *non-progressive/progressive* (Comrie 1976: 25) and Timberlake gives four aspectual operators, i.e. *progressive*, *iterative/habitual*, *perfect* and *perfective* (Timberlake 2007: 287-292). The term *perfect* should be avoided in comparative typology as previously discussed (cf. section 6.1), and Bybee, Perkins and Pagliuca proposes the term *anterior* instead (Bybee et al. 1994: 54).

Bybee, Perkins and Pagliuca define the perfect as indicating that the event 'occurs prior to reference time and is relevant to the situation at reference time' while the perfective is contrasted as implying that the event 'is viewed as bounded temporally' (Bybee et al. 1994: 54). Timberlake similarly defines the perfect as presenting 'a situation as a state and that the state extends back, as a continuous interval, to include the actual event reported by the predicate' (Timberlake 2007: 290), while the perfective 'imposes boundaries on situations at the contextual occasion' (Timberlake 2007: 292). Bybee, Perkins and Pagliuca's distinction between perfect and perfective is not exclusive, which means that both can be true, while Timberlake's definition of the perfect does not strictly refer to the event itself, which Comrie explicitly acknowledges as he states that 'the perfect is rather different from these aspects, since it tells us nothing directly about the situation in itself' (Comrie 1976: 52).

Since the definitions of perfect generally do not relate to an event alone but rather to how an event relates to other past events or the current speech act, it is worth questioning whether perfect is actually an aspectual category in the narrow sense. I have therefore chosen to not include the notion of perfect aspect in this study, as it is exceedingly difficult to operationalise cross-linguistically. The prototypical perfect aspects are regularly expressed by the perfective aspect in e.g. Russian (Wade et al. 2020: 269, 295), which makes it almost impossible to distinguish perfect and perfective aspect in the Russian-language literature.

The functions of the perfect aspect given by Comrie (Comrie 1976: 56-61) can all be analysed as perfective, while the *perfect of persistent situation* is also 'bounded temporally' by an obligatory temporal phrase (cf. the examples in Comrie (1976: 60)), as the event in *Matilda has lived in Sydney for two years (and she still lives there)* is bounded by the temporal phrase *for two years*. The absence of the obligatory temporal phrase will inevitably yield a perfective reading, e.g. *Matilda*

has lived in Sydney, and the *perfect of persistent situation* should therefore rather be analysed as a perfective *stative predicate* (Bybee et al. 1994: 55). Timberlake's merging of iterative and habitual is not valid however, as the habitual aspect refers to events that are 'customarily repeated on different occasions' while the iterative aspect refers to an event that is 'repeated on a particular occasion' (Bybee et al. 1994: 127).

Aspect is found in some form in all 56 languages of the affixal data set, and the Northwest Caucasian languages generally have the most complex aspectual affixation patterns in the Caucasus as many of them distinguish more than ten aspectual categories (Arkadiev & Lander 2020: 422-423). The exact definitions of the various aspectual functions are not universally agreed upon and there is considerable overlap between different definitions, which means that the summary of aspectual affixation in the Caucasus below (figure 6.2.) is tentative at best.

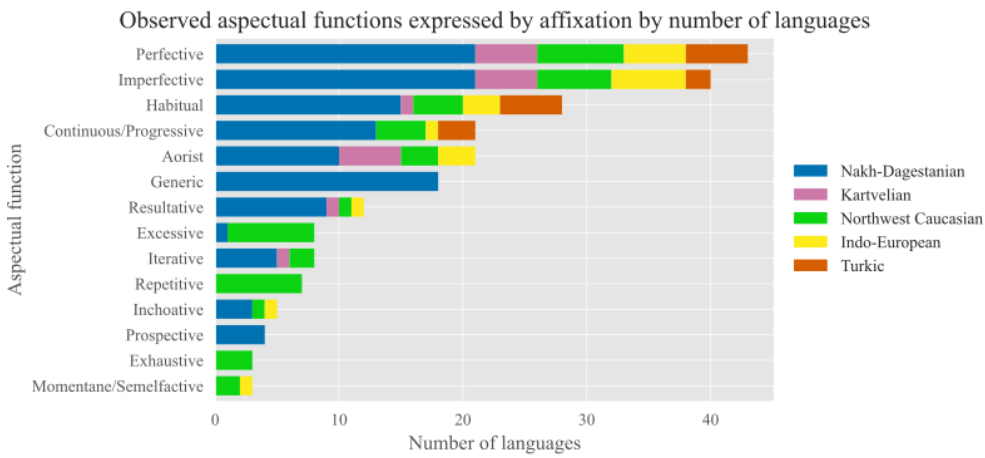


Figure 6.2: The most common observed aspects expressed by affixation in the data according to number of languages.

The Perfective aspectual function

The Perfective aspect indicates that an event 'is viewed as a single whole' (Comrie 1976: 16) or 'as bounded temporally' (Bybee et al. 1994: 54). However, there are differences between how the perfective aspect is defined in Western traditions and in the Russian-language tradition, where the Russian definition generally implies telicity, or limitation by a relative boundary, and totality of an event (Gvozdanović 2012: 784-785). Perfective affixes are found in at least 40 languages across all five languages families of the Caucasus, where Nakh-Dagestanian languages typically differentiate between perfective and imperfective verb stems by means of stem vowel alternation, reduplication and affixation (Ganenkov & Maisak 2020: 111-112).

(180) Sanzhi Dargwa (Forker 2020a: 538) (gloss adapted from source)

d-e<r>čːi-ij=qʼar	*it-i-l*	*čaˤj*
NH.PL-drink<PFV>-INF=MOD	that-OBL-ERG	tea
d-učː-an=de,	*amma*	*itːa-l*
NH.PL-drink.IPFV-PTCP=PST	but	those.OBL-ERG
a-a<l>t-ur	*ca-w*	
NEG-let<IPFV>-PRET	COP-M	

‘As for drinking, he would have drunk the tea, but they did not let him’

The perfective aspect is expressed by means of preverbs in all Kartvelian languages except Laz (Tuite 1998: 13; Öztürk & Pöchtrager 2011: 71) and in Iron Ossetic (Abaev 1964: 76-78; Erschler 2020: 658).

(181) Megrelian (Vamling & Tchantouria 1993: 70) (reglossed)

osur-k	*buneba*	*do-Ø-xantʼ-u*
woman-ERG	landscape.NOM	PV-3.O-paint-AOR.3SG

‘The woman painted a landscape’

(182) Iron Ossetic (Erschler 2020: 658) (original glossing)

ʃoʃlan	*dɐʃ*	*minut-mɐ*	*ʧinəg*	*ba-kaʃti*
Soslan	ten	minute-ALL	book	PV-read.PST.3SG

‘Soslan read a book in ten minutes’

The North Azerbaijani suffix *–mḶš* has been described as expressing perfect-like functions (Ragagnin 2022: 251) and is conventionally analysed as a perfect (Širaliev 1971: 125), while its function when not preceded by *-(y)Ḷr* appears to be perfective (cf. example (183)).

(183) North Azerbaijani (Širaliev 1971: 126) (my glossing)

ay oğul, sən-ə	*elə*	*adam-ın*	*qız-ın-ı*	
oh son, 2SG-DAT	such	man-GEN	girl-3SG.POSS-ACC.DEF	
al-mış-ıq	*ki,*	*bütün*	*ömr-ü*	*boy-u*
take-PFV-1PL	that,	all	life-ACC.DEF	length-3SG.POSS
var-dövlət	*içində*	*üz-əcək-sən*		
wealth	inside	swim-FUT-2SG		

‘Oh son, we have taken the daughter of such a man for you to marry, that you will live your whole life swimming in riches (lit. swim in wealth the whole length of life)’[97]

[97] Translated from the Russian translation ‘О сын мой, мы тебе в жены взяли дочь такого человека, что всю жизнь свою ты будешь плавать как сыр в масле’.

The Imperfective and Continuous/Progressive aspectual functions

The Imperfective aspect indicates that an event is 'in progress at a particular reference point, either in the past or present' (Bybee et al. 1994: 125). Comrie defines the imperfective aspect as 'explicit reference to the internal temporal structure of a situation, viewing a situation from within' (Comrie 1976: 24), which is unsatisfactory from a functional perspective because it is difficult to operationalise cross-linguistically.

(184) Lezgian (Haspelmath 1993: 186, 223) (gloss adapted from source)

a. *alfija-di-z wiči-z wič güzgü-d-a akwa-zwa*
Alfija-OBL-DAT self-DAT self mirror-OBL-IN.ESS see-IPFV
'Alfija sees herself in the mirror.'

b. *muallim-di-n čkadal ada-z wiči-n buba akwa-zwa-j*
Teacher-OBL-GEN instead he-DAT self-GEN father see-IPFV-PST
'Instead of the teacher he saw his father'

(185) North Azerbaijani (Širaliev 1971: 149, 314) (my glossing)

a. *hamı abbas-a və atlas-a bax-ır*
everybody Abbas-DAT and Atlas-DAT look-PRS(3SG)
'Everybody is looking at Abbas and Atlas'[98]

b. *mən sakit-cə dayan-ıb ağac-a bax-ır-dı-m*
1SG silent-ADV stand-CVB tree-DAT look-PRS-PST-1SG
'While standing silently, I was looking at a tree'[99]

(186) Georgian (Hewitt 1995: 93, 553) (reglossed)

a. *ai es k'ac-i-a, rom*
lo this man-NOM-COP.3SG, that
g-e-ubn-eb-od-i
2.O-PASS-tell-SM-IPFV-1SG/2SG[100]
'look/lo, this is the man I was telling you about!'

b. *deda bavšv-s t'q'uil-s*
mother.NOM child-DAT lie-DAT
(Ø-)e-ubn-eb-a
(3.O)-PASS-tell-SM-PRS.3SG[101]
'The mother is telling the child a lie'

[98] Translated from Russian 'Все смотрят на Аббаса и Атлас'.

[99] Translated from Russian 'Я стоял спокойно и смотрел на дерево'.

[100] Glossing based on personal communication with Manana Kock Kobaidze. Hewitt gives the form as *(Ø-)g-e-ubn-eb-od-i* and glosses it as '(it-)you-IOV-tell-TS-IMPERF-INDIC'.

[101] Glossing based on on personal communication with Manana Kock Kobaidze. Hewitt gives the form *(Ø-Ø-)e-ubn-eb-a* which he glosses as '(it-him-)IOV-tell-TS-she(PRES)'.

Bybee at al. (1994) builds upon Comrie's definition and define the imperfective as contrastive to perfective, viewing 'the situation not as a bounded whole' and more precisely 'as in progress at a particular reference point' (Bybee et al. 1994: 125). The imperfective aspect therefore typically stands in opposition to the perfective aspect, and explicit perfective and imperfective affixes are found in more than half of the languages of the Caucasus, while some languages leave either the perfective or the imperfective zero-marked.

The aspects labelled Progressive, Continuous, Continuative and Durative all indicate that an action is 'ongoing at reference time' (Bybee et al. 1994: 126-127). They are defined as different aspects by Bybee et al., while their distinction between the three categories is completely based on a dynamic/stative distinction. These aspects are all functionally related to the imperfective, as Comrie describes them as subcategories of the imperfective, and Comrie further defines progressive as a subcategory of continuous aspect (Comrie 1976: 25), which has led me to merge these categories into a joint continuous/progressive function. The suffix *–mAK-dA* in North Azerbaijani has been described as a 'focal present' (Ragagnin 2022: 251), but Širaliev demonstrates that it has a continuous/progressive function (Širaliev 1971: 137), while in Kumyk it appears to only carry a stative continuous function (Abdullaeva et al. 2014: 203, 340-341). The term continuative is also occasionally used, e.g. for the Lezgian suffix *–ma* (Haspelmath 1993: 127).

(187) North Azerbaijani (Širaliev 1971: 134) (my glossing)
yaz-maqda *i-miş-əm*
write-PROG AUX-EVID-1SG
'I have apparently been writing for a while (for a long period of time)'[102]

(188) Lezgian (Haspelmath 1993: 389) (reglossed)
ahmed aniz fi-zma-j *qʷ'an,* *zun ada-w* *raχa-da-č*
Ahmed DIST go-PROG-PTCP as.long.as 1SG 3SG.AD talk-FUT-NEG
'As long as Ahmed still goes there, I won't talk to him'

The Habitual and Generic aspectual functions

The Habitual aspect indicates that an event is 'customarily repeated on different occasions' (Bybee et al. 1994: 127). Habitual affixes are found in all Turkic languages of the Caucasus (Dehghani 2000: 116; Johanson 2022b: 113; Ragagnin 2022: 251; Berta & Csató 2022: 331; Karakoç 2022: 361), but they often syncretise with the future. Habitual affixes are also present in numerous Nakh-Dagestanian languages, but previous descriptions tend to describe these verb forms with other labels such as 'imperfect', e.g. the past habitual in Dargwa (Van den Berg 2001: 39). Habitual constructions in Khwarshi exemplify the connection between habitual

[102] Translated from Russian '[Я], оказывается/вероятно/кажется, писал (длительно), (но отчетливо об этом не помню)'.

aspect and the general tense in Nakh-Dagestanian, as the general tense typically conveys habitual meanings, although it is used in narratives with a clearly non-habitual function as well (Khalilova 2009: 187-188), which also holds true for Tindi (Magomedova 2012: 171). The *permansive* or *iterative* aspect in Old Georgian has been described as conveying both iterative and habitual functions (Fähnrich 1994: 85; Tuite 2008a: 159-160).

(189) North Azerbaijani (Širaliev 1971: 298) (my glossing)
ağac-a su ver-ər-lər qayda-dır ək-il-əndə
tree-DAT water give-FUT/HAB-3PL rule-COP.3SG plant-PASS-TEMP
'As a rule, trees are usually watered when they are planted'[103]

(190) Karachay-Balkar (Aliev 1973: 314) (my glossing)
biz otun taš-ïwču[104] *arba sïn-dï*
1PL firewood carry-MSD/HAB cart break-PST.3SG
'The cart, with which we usually carry firewood, broke down'[105]

(191) Nogai (Karakoç 2005: 110) (gloss adapted from source)
ol bulay-da oltïr-ataɣan
3SG here-LOC sit-HAB
'He usually sits here'[106]

The exact distinction between the habitual and the generic aspect described below is not trivial, as e.g. the habitual suffix *–(i)da* described in Ghodoberi by Kibrik (1996: 49) seems to rather convey a generic aspect, cf. (192). The presence of a generic aspect (or 'generic present' in the Russian tradition, cf. section 6.1) in the Nakh-Dagestanian languages should therefore either be seen as a broader generic/habitual aspect, or that there truly is a functional difference between these two aspects.

(192) Ghodoberi (Kibrik 1996: 34) (gloss adapted from source)
t'orda ʁand-e r=eč'er ru=k'-ida
all crow-PL NH.PL=black NH.PL=be-HAB
'All crows are black'

[103] Translated from Russian 'Деревья поливают, как правило, когда сажают'.

[104] Aliev (1973) describes this form as a masdar, while it is apparently only used to indicate habituality.

[105] Glossed and transliterated from *Биз отун ташыгъан арба сынды*, and translated from Russian 'Арба, на которой (обычно) мы возили дрова, сломалась'.

[106] Translated from German 'Er pflegt immer hier zu sitzen'.

The Generic or Gnomic aspect expresses events that generally occur or facts that are generally true, which typically involves non-specific referents (Carlson 2012: 830-831). The generic aspect is functionally related to the habitual aspect but they are not identical, as e.g. the generic statement 'all crows are black' is not synonymous with the illicit habitual statement '*all crows are usually black', as the quantifier prohibits such an interpretation. The generic aspect has generally not been described in the Caucasus, but the traditional term 'general tense' has been used for verb forms in Nakh-Dagestanian languages that should rather be analysed as instances of generic aspect. Implicitly encoding a generic aspect is typologically very rare (Carlson 2012: 831), so if the 'general' tense in Nakh-Dagestanian qualifies as generic aspect it could constitute a rare example of generic aspectual affixes. The generic aspect in Avar is described as a 'simple present' by Forker (2020b: 255), which is unproblematic as the simple present in English also expresses the generic present.

(193) Avar (Alekseev et al. 2012: 216) (my glossing)
dow kidago c'alu-le-w w-uk'-una
3SG always read-PTCP.IMPF-M M-be-GNOM
'He is always reading'[107]

The Aorist, Preterite and Semelfactive aspectual functions

The term aorist has a long tradition of being applied to various aspects in the Caucasus, in e.g. Classical Armenian (Lauer 1869; Meillet 1936), Georgian (Marr & Brière 1931; Vogt 1938) and later Lezgian (Haspelmath 1993). The term and the concept are borrowed from Classical Greek grammars where the aorist indicated a 'simple occurrence' or 'single event' in the past (Goodwin 1897: 16; Morwood 2002: 61), corresponding to the *preterite* or *simple past* in English (Goodwin 1897: 16).[108] The aorist in Georgian similarly refers to simple actions in the past (Hewitt 1995: 242), and it can refer to both perfective and imperfective events (Hewitt 1995: 242; Bolkvadze & Kiziria 2023: 237).

(194) Old Georgian (Harris 1985: 48) (reglossed)
xolo čwen ɣmert-man cecxl-i
and 1PL.DAT God-ERG fire-NOM
samsaxurebl-ad mo-gw-c-a
servant-ADV PV-1PL.O-give-AOR.3SG
'And God gave fire to us to use (lit. as a servant)'

[107] Glossed and transliterated from *Дов кидаго цIалулев вукIуна*, and translated from Russian 'Он всегда читающим бывает (читает)'.

[108] Morwood claims that it corresponds to the pluperfect, but exemplifies the aorist by the simple past example 'I did this' (Morwood 2002: 61).

The co-occurrence of the aorist aspect with either imperfective or perfective forms is also found in e.g. Bulgarian (Gvozdanović 2012: 791-792), indicating that the aorist aspect does not necessarily express perfectivity as it should rather be analysed as the opposite of the habitual aspect. The aorist in Eastern Armenian has a function that is similar to the Classical Greek aorist as it indicates single events, successive actions and 'ingressive meanings' in the past, but Dum-Tragut describes it as expressing perfective aspect (Dum-Tragut 2009: 230-232). Aorists have also been described in e.g. Bezhta (Comrie, Khalilov & Khalilova 2015: 363) and the verb forms labelled as preterite in many contemporary Dargic descriptions appears to be functionally identical to the aorist aspect.

(195) Eastern Armenian (Dum-Tragut 2009: 231) (gloss adapted from source)
hayk-ə tun ek-av banali-n grpan-ə dr-ec^{h}
Hayk-DEF house come-AOR.3SG key-DEF pocket-DEF put.AOR.3SG
ew štap k^{h}ayl-er-ov durs gn-ac^{h}
CONJ quick step-PL-INS out go-AOR.3SG
'Hayk came home, put the key into the pocket, and went out with quick steps'

(196) Bezhta (Comrie, Xalilov & Xalilova 2015: 365) (my glossing)
łanas-la wod-iʔ Ø-õqχ'-ojo mähämmä is bitło-ʁa.
third-TRANS day-IN I-come-AOR Muhammad brother house-AD
beta χisa tłoba-tł'a ijo-abo-doj ẽdo-k-ijo,
then next.day noon-SUPER mother-father-APUD IN-go-AOR
nisdajlil di-doj-na gic'a-g<o>c'-ojo
afternoon 1SG-APUD-ADD REDUP-visit<I>-AOR
'Brother Muhammad came to the village for three days. Then the next day at noon he went to his parents and in the afternoon he visited me as well'[109]

The aorist in Lezgian and Avar are likewise described as indicating perfective past but without expressing the functions associated with the aorist in Classical Greek (Haspelmath 1993: 142; Forker 2020b: 255), which suggests that these aorist forms should be analysed as perfectives. The term aorist is also used within Turkic linguistics, but with a completely different function (Johanson 2022a: 39), as the 'aorist' in North Azerbaijani expresses habitual and future reference (Ragagnin 2022: 251). The term aorist is almost exclusively used for aspects in non-Western Indo-European, Turkic and Caucasian languages, which makes it both a problematic term and a difficult concept to operationalise comparatively, as the aorist aspect and

[109] Glossed and transliterated from *Лъанасла водиъ оʰкъойо Маьгьаьммаь ис билӀогъа. Бета хиса лӀобакьа ийо-абодой эʰдокийо, нисдайлил дидойна гицӀа гоцӀойо*, and translated from Russian 'На три дня приехал брат Магомед в село. На следующий день в обеденное время навестил родителей и после обеда зашел ко мне'.

the various preterite aspects, including the Anglo-French terminology *simple past*/*passé simple*, appear to be in complementary distribution.

The Momentane or Semelfactive aspect indicates a punctual event that 'takes place once and only once' (Comrie 1976: 42), and is therefore almost functionally identical to the aorist aspect. The semelfactive prefix *ê-* has been described in Shapsug Adyghe (Smeets 1984: 270) and there are possibly momentane prefixes in Abkhaz (Chirikba 2003a: 38, 54) and Ossetic (Abaev 1964: 45) but are not described as such. The prefix *a-* in Iron Ossetic functions both as a spatial preverb and to indicate 'a rapid, brief and superficial action' (Abaev 1964: 77)

(197) Shapsug Adyghe (Smeets 1984: 272) (gloss adapted from source)
sə-z-j-e-łeʁʷə-m
1SG.ABS-TEMP.CVB-3SG.ERG-SEM-see-REL
'When he saw me'

(198) Iron Ossetic (Abaev 1964: 77) (my glossing)
a-lœww-id
MOM-stand-PST.3SG
'He stood for a little while'[110]

The Resultative aspectual function

The Resultative aspect expresses 'a state implying a previous event' (Nedjalkov & Jaxontov 1988: 6) or that 'a state exists as a result of a past action' (Bybee et al. 1994: 54). The resultative aspect is therefore notionally connected to Comrie's *perfect of result*, defined as a present state 'being the result of some past situation' (Comrie 1976: 133), while briefly defining resultative as 'successful completion of a situation' (Comrie 1976: 20) which is rather a *completive* (Bybee et al. 1994: 54). Nedjalkov and Jaxontov give examples from Eastern Armenian to demonstrate how the resultative differs from the perfect, cf. examples (199).

(199) Eastern Armenian (Nedjalkov & Jaxontov 1988: 16) (reglossed)
a. *Na (*deṙ) ənk-əl ē*
he (*still) fall-PTCP.PFV be.3SG.PRS
'He has fallen (*still)'
b. *Na deṙ ənk-ac ē*
he still fall-PTCP.RES be.3SG.PRS
'He is still fallen'

Nedjalkov and Jaxontov's example above inadvertently reveals that Comrie's definition of *perfect of result* actually appears to reflect the function of the

[110] Abaev gives the translation 'I stood for a little while', but this must be an error as the first person singular form would be *alœwwidtæn*.

resultative instead, which consequently supports the analysis that the present perfect does not alone imply present tense reference. In both Avar and Archi the 'perfect' and 'pluperfect' express a resultative function (Kibrik 1988: 171-172; Forker 2020b: 256; Chumakina 2020: 303), demonstrating that resultatives are not restricted to present tense reference. All four forms of conjugation series IV in Megrelian have been described as 'resultative' by Kajaia (2008: 215), while Rostovtsev-Popiel instead describes it as a 'dedicated evidential series' (Rostovtsev-Popiel 2020: 549).

The Excessive aspectual function

The Excessive aspect indicates that an action is carried out in an excessive manner or to an excessive extent (Fenwick 2011: 126; Chirikba 2003a: 53). In the Caucasus, excessive affixes are almost exclusively found in the Northwest Caucasian language family, e.g. in Abkhaz (Chirikba 2003a: 53; Hewitt 2010: 272), Abaza (Lomtatidze & Klychev 1989: 108), Kabardian (Kumakhov & Vamling 2009: 49), Adyghe (Rogava & Keraševa 1966: 309), Abzakh Adyghe (Paris 1989: 198), Shapsug Adyghe (Smeets 1984: 276) and Ubykh (Fenwick 2011: 126). Kumakhov and Vamling gloss the excessive as 'abundance' (Kumakhov and Vamling 2009: 49), but the function appears to be identical to the excessive aspect in Abkhaz (cf. examples (200) and (201)). The persistive suffix *–mat* in Archi indicates that the duration of an action is longer than anticipated (Chumakina, Bond & Corbett 2016: 40; Chumakina 2020: 303), which is possibly the closest equivalent to an excessive affix outside of the Northwest Caucasian language family, although its function is quite different.

(200) Abkhaz (Chirikba 2003a: 53) (gloss adapted from source)
jə-r-fa-cʷá-Ø-jt'
3SG.NH.ABS-3PL.ERG-eat-EXC-AOR-DYN.FIN
'They ate too much'

(201) Kabardian (Kumakhov & Vamling 2009: 49) (reglossed)
a. fə-ʃxe-ɕ-a-q'əm
2PL.ABS-eat-EXC[111]-PFV-NEG
'You didn't eat too much'
Kabardian (Kumaxov 2006: 93) (my glossing)
b. *məɕe dəʁʷəƶ f'e-belace-ɕ*
bear wolf seem-shaggy-EXC
'A wolf seems too shaggy to a bear'[112]

[111] Kumakhov & Vamling gloss it as ABU 'abundance'.

[112] Translated from Russian 'Медведю волк кажется слишком лохматым'.

The Iterative and Repetitive aspectual functions

The Iterative aspect indicates that an action is 'repeated on a particular occasion' (Bybee et al. 1994: 127). The term iterative is sometimes misleadingly used for the habitual aspect that instead refers to events repeated over multiple occasions, e.g. the 'iterative' suffix *–la* in Akbhaz, which only appears to have a habitual function (Spruit 1986: 55; Chirikba 2003a: 53). The term *repetitive* is sometimes used for the iterative aspect (Bybee et al. 1994: 127), but it is relevant to differentiate the iterative and repetitive aspects (Chirikba 2003a: 53-54), where the repetitive aspect refers to a *single repetition* in the Northwest Caucasian terminology (Arkadiev & Lander 2020: 423), which sets it apart from the iterative aspect. The terminology for these aspectual functions is not optimal, but it is important to differentiate them somehow, cf. the iterative suffix *–re* and the repetitive suffix *–ž* in example (202). Seegmiller (1996: 23) mentions the past iterative suffix *–lwču* with the auxiliary *e-* in Karachay, but the function appears to be purely habitual (Aliev 1973: 314). The suffixes *- æ:l/-i-e:l* in Svan can also express an iterative function (Schmidt 1991: 512; Tuite 1998: 8).

(202) Kabardian (Kumakhov & Vamling 2009: 51) (reglossed)
sə-q'a-kʷ'e-re *sə-kʷ'e-ž-w*
1SG.ABS-PROX-go-ITER.CRD 1SG.ABS-go-REP[113]-PTCP
'I come (repeatedly) and leave'

The Repetitive aspect indicates that an event is repeated, often as a *single repetion*, typically corresponding to the adverb 'again' (Chirikba 2003a: 54; Arkadiev & Lander 2020: 423). The 'reversed action' described by Kumakhov and Vamling (2009: 99-100) is analysed as a repetitive by Arkadiev & Lander (2020: 423), although it also has a retrolative function, cf. section 6.11.2. The repetitive aspect is therefore functionally related to the iterative and habitual functional aspects, but they differ in the iterative aspect implies a repetition on a single occasion while the habitual implies repetition over multiple occasions (Bybee et al. 1994: 127). The iterative suffix *-ɐj(i)* in Ubykh appears to rather have a repetitive function (Fenwick 2011: 124-125), although it could also be analysed as a retrolative affix, cf. section 6.11.2.

(203) Ubykh (Fenwick 2011: 125) (reglossed)
ɐ-wɨ́-s-tʷ-ɐji-n
3SG.ABS-2SG.OBL-1SG.ERG-give-REP-PRS
'I give it back to you'

[113] Glossed as REV 'reversed action' by Kumakhov & Vamling.

The Inchoative/Ingressive aspectual function

The Inchoative or Ingressive aspect indicates 'the beginning of a situation' (Comrie 1976: 19). Inchoative affixes are rare in the Caucasus but can be found in e.g. Eastern Armenian (Dum-Tragut 2009: 173), Abzakh Adyghe (Paris 1989: 186), Bezhta (Kibrik & Testelets 2004: 270) and Hunzib (Van den Berg 1995: 111). The inceptive suffix in Chechen appears to have both an inchoative and terminative function (Nichols 1994: 46). These inchoative affixes should possibly rather be analysed as derivational, as most of the examples I have found attach to either nouns or adjectives, cf. (204) and (205).

(204) Eastern Armenian (Dum-Tragut 2009: 173) (reglossed)
caʁik-ner-ə *čor-an-um* *en*
flower-PL-DEF dry-INCH-PTCP.PRS COP.3PL
'The flowers are drying'

(205) Abzakh Adyghe (Paris 1989: 186) (gloss adapted from source)
Ø-šemə-ʔʲe-Ø-x
3.ABS-cow-INCH-PRS-PL
'These are heifers (lit. they are becoming cows)'[114]

The Prospective/Inceptive aspectual function

The Prospective or Inceptive aspect indicates that an event is 'about to happen' or 'going to happen' (Maisak & Tatevosov 2001a: 276; Alekseev et al. 2012: 212; Forker 2020b: 257). Van den Berg describes a terminative aspect in Hunzib indicating that an event 'is almost happening or almost accomplished' (Van den Berg 1995: 111), which appears to share some but not all of its functions with the prospective aspect described in Bagvalal.

(206) Bagvalal (Maisak & Tatevosov 2001a: 277) (gloss adapted from source)
o-šːu-r *kaʁal* *qʷa-rā-di-b-o* *ekʷ'a*
he-OBL.M-ERG letter write-POT-PROSP-III-PTCP[115] be.PRS
'He intends to write a letter'[116]

(207) Avar (Charachidzé 1981: 114) (my glossing, retransliterated)
hʷe *ʕadin* *χʷ-eze-qin* *w-ug-in*
dog like die-INF-PROSP M-COP-EMPH
'I'm about to die like a dog'

[114] Translated from French '[C]e sont des génisses'.

[115] The original glossing is 'писать-MS.POT-PROSP=N=CONV'.

[116] Translated from the Russian translation 'Он собирается писать письмо'.

The Exhaustive/Completive aspectual function

The Exhaustive or Completive aspect indicates an ‘action done to completion or to fulfilment’ (Fenwick 2011: 125) or ‘to do something thoroughly and to completion’ (Bybee et al. 1994: 54). In the Caucasus, exhaustive affixes are only documented in the Northwest Caucasian languages, e.g. Abzakh Adyghe (Paris 1989: 185), Shapsug Adyghe (Smeets 1984: 276) and Ubykh (Fenwick 2011: 125).

(208) Ubykh (Fenwick 2011: 126) (gloss adapted from source)
кʒ-łɐp’ʒ́ *dʁʒ-Ø-Ø-p’tɕ’ʒ-lʒ́-t’in…*
3SG.POSS-foot SUB-3SG.ABS-3SG.ERG-clean-EXH-TEMP.CVB
‘When he had finished cleaning his feet..’

6.3. Modal functions

Grammatical mood or modality encodes the 'status of the proposition that describes the event' while not referring to 'any characteristic of the event' (Palmer 2001: 1). Mood has previously been defined as the grammaticalisation of 'speakers' (subjective) attitudes and opinions' (Bybee et al. 1994: 176), which still holds true but only covers some of the notions expressed by mood and modality. The variation of moods expressed by affixation in the Caucasus is remarkable as there are almost 30 separate grammatical moods in the data. Mood is also one of the few grammatical categories expressed by means of affixation in all 56 languages of the affixal data set. The exact definition and categorisation of the various moods is an ambitious endeavour, as the description of grammatical mood and modality is a notoriously thorny field of linguistic research (Nuyts 2016: 2). Figure 6.3 below is therefore a tentative distribution of modal functions found in the languages of the Caucasus.

The Indicative or Realis is unsurprisingly the most common mood, but it is mainly an implicit mood that is only observable due to the presence of contrastive non-indicative or non-realis forms in a verbal paradigm. The indicative suffix *–mä* in Khinalug is a rare example of an explicit indicative affix in the Caucasus (Kibrik et al. 1972: 106; Khvtisiashvili 2013: 166). The 'assertive' or 'declarative' suffix *-ҏ* in Kabardian is obligatory for all indicative stative forms and all indicative non-present dynamic forms (Kumakhov & Vamling 2009: 47), while absent in other moods (Arkadiev & Lander 2020: 416), which makes me analyse it as an explicit indicative suffix.

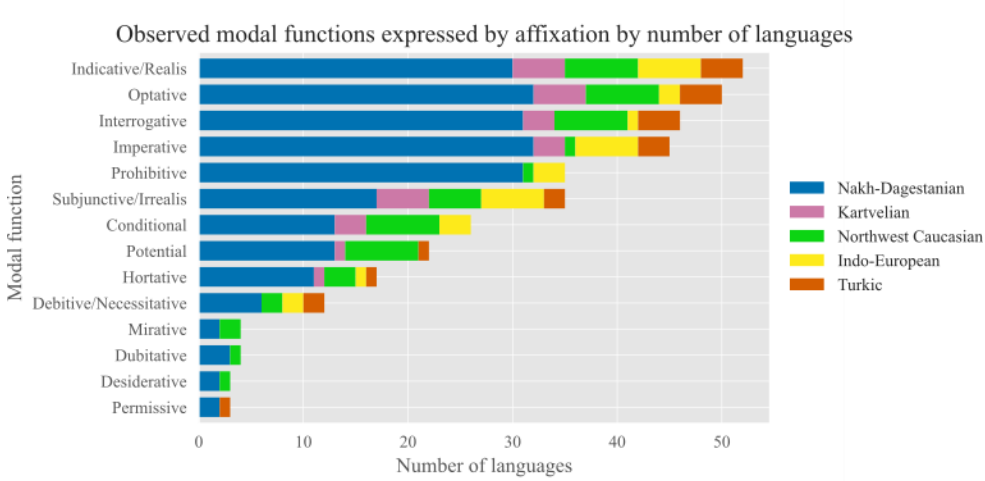


Figure 6.3: The most common observed moods expressed through affixation in the data according to number of languages.

The Optative and Desiderative modal functions

The Optative mood is the most frequent explicit mood expressed by affixation in the data and it indicates that the speaker wishes or hopes that an action or a state is true (Bybee et al. 1994: 179; Nikolaeva 2016: 77). Optatives typically correspond to constructions of the type 'may X happen' (Polinsky 2020: 13), and are therefore frequently used for vows and curses. Optative affixes are found in 48 of the 56 languages in the data and are found in all five language families of the Caucasus. The functionally related Votive suffix *–tir* as been described in Kryts but it is used synonymously with the optative (Authier 2009: 273).

(209) Svan (Tuite 1997: 44) (gloss adapted from source)
ɣerte-m *či-v* *dž-a-mzər-a-x*
God-ERG all-OPT[117] 2PL.O-NV-bless-OPT-PL
'May God bless all of you'

(210) Kubachi (Magometov 1963: 204) (my glossing)
dig *ix^{w}le* *b-ēc'-ab!*
meat quickly III-roast-OPT.3SG
'May the meat roast quickly!'[118]

The optative mood previously described in Iron Ossetic is analysed as a subjunctive by Erschler (2020: 662), which appears to be a more appropriate analysis as none of the optative examples given by Abaev (1964) and Bagaev (1965: 306-308) indicate the optative function described above. In Xaidaq, the optative is zero-marked but there is also a permissive-optative aspect that conveys both a permissive and an optative function (Temirbulatova 2004: 172). The 'permissive I' described in Bezhta is functionally identical to the definition of the optative mood described above (Comrie, Xalilov & Xalilova 2015: 382), while the moods labelled 'optative I' and 'optative II' should rather be analysed as instances of desiderative and hortative mood (Comrie, Xalilov & Xalilova 2015: 380-381). The widespread presence of optative affixes in the Caucasus is highly unusual from a typological perspective and is a well-known Caucasian peculiarity, and it is arguably an areal feature of the Caucasus (Polinsky 2020: 13).

[117] Tuite glosses this as OPT 'optative particle'.

[118] Translated from Russian 'Да изжарится быстро мясо!'.

(211) Karachay-Balkar (Aliev 1973: 97) (my glossing)
oj, *an-dan* *al-ɣa* *öl-ej-im*
INTERJ, 3SG-ABL before-DAT die-OPT-1SG
'Alas, may I die before him!'[119]

(212) Juhuri (Authier 2012: 177) (gloss adapted from source)
bebe=y=tü *su* *gir-o!*
father=EZ=2SG.POSS fire take-OPT.3SG
'May your father burn (lit. catch fire)!'

The Desiderative mood indicates that a state or an event is desired or wished for (Timberlake 2007: 329; Palmer 2001: 131), and is similar but not identical to the optative mood. The primary difference between the desiderative and the optative is that the desiderative can be 'used for the wish of a participant in the state of affairs related to in the utterance' while the optative by definition expresses the wish of the speaker (Dobrushina et al. 2005: 299). This distinction can be made in English as well, e.g. the desiderative *he wishes to eat something* simply expresses the wish of the subject while the optative *may he eat something* explicitly expresses the wish of the speaker. Bezhta and Hunzib have been described as having both desiderative and optative affixes (Van den Berg 1995: 76, 112; Comrie, Xalilov, Xalilova 2015: 380), which also might hold true for Shapsug Adyghe (Smeets 1984: 270, 278).

(213) Hunzib (Van den Berg 1995: 112) (gloss adapted from source)
iyu-u *xank'al* *r-uw-aq'e-n* *zuq'u-r*
mother-DAT khinkal(V) V-do-DESID-PTCP.PFV be-PST
'Mother liked to make khinkal'

The desiderative in Abkhaz and Abaza described by O'Herin (2020: 475) should rather be analysed as an optative as it expresses the wish of the speaker (Aristava 1968: 121; Tabulova 1976: 159; Chirikba 2003a: 46), cf. (214).

(214) Abkhaz (Aristava 1968: 122) (my glossing)
a-t'rakt'or *a-r-nɨq'-wa-ra*
DEF-tractor DEF-CAUS-drive-IPFV-MSD
z-dɨr-wa-nda(z)
1SG.ERG-know-IPFV-OPT(-PST)
'If I only knew how to drive the tractor!'[120]

[119] Glossed and transliterated from *Ой, андан алгъа ёлейим*, and translated from Russian 'Да чтобы я раньше его умерла!'.

[120] Glossed and transliterated from *Атрақтор арныӄуара здыруа-нда(з)*,and translated from Russian 'Эх, если бы я умел (знал) водить трактор (больше ничего не хотел бы)'

The Interrogative modal function

The Interrogative mood indicates that the clause is a question (Nikolaeva 2016: 71) and interrogative affixes or clitics are found in 46 of the 56 languages of the data, while most of these are described as interrogative particles. The interrogative markers in the Turkic languages are often suffixed to the verb but they are also technically particles (Ragagnin 2022: 245), and the interrogative particle *–mi* has been borrowed into Juhuri (Authier 2012: 222).

(215) Chechen (Aliroev 1999: 89) (my glossing)
ħo *bazar-a* *v-öd-ij?*
2SG bazaar-ALL V-go.PRS-Q
'Are you going to the bazaar?'[121]

(216) Adyghe (Rogava & Keraševa 1966: 255) (my glossing)
pɛəne *maqe,* *arep,* *qe-ʔʷ-a?*
accordion sound really PROX-spread-Q
ħawmi, *awɛtew* *se* *qə-s-ɛe-χʷ-a?*
or, that 1SG PROX-1SG-IN-be-Q
'Can the sound of accordions really be heard? Or is it as it seems to me?'[122]

(217) Laz (Lacroix 2009: 140) (original glossing)
tkva *deve* *ko-dzir-i-t-i?*
2PL camel PV-see-AOR-1/2PL-Q
'Have you seen the camel?'

The Imperative and Prohibitive modal functions

The Imperative mood indicates that the clause is a direct command (Bybee et al. 1994: 179; Aikhenvald 2010: 1). In about a third of the world's languages the second person singular imperative 'coincides with either the verb root or the stem' (Aikhenvald 2010: 18), as imperatives are often zero-marked (Timberlake 2007: 326). Nevertheless, 43 of the 56 languages in the data have explicit imperative affixes, which often encode plural imperatives or 3rd person imperatives, which are often analysed as jussive mood affixes instead, cf. the jussive mood. Non-specific imperative affixes are found in more than 20 of these 42 languages, e.g. Lezgian (Haspelmath 1993: 129), Hinuq (Forker 2013), Ghodoberi (Kibrik 1996: 49; Saidova 2004: 94), Tsez (Alekseev & Radžabov 2004: 143), Bagvalal (Dobrushina

[121] Glossed and transliterated from *Хьо базара воьдий?*, and translated from Russian 'Ты идешь на базар?'.

[122] Glossed and transliterated from *Пщынэ макъэ, арэп, къэIу-а? Хьауми, аущтэу сэ къысщэхъу-а?*, and translated from Russian 'Доносятся ли звуки гармошки, или так мне кажется?'.

et al. 2001: 73), Andi (Salimov 2010: 189, 198) and Karata (Magomedbekova 1971: 131-132). The Andic languages are worth mentioning, as transitive and intransitive imperatives are described as being formed by different suffixes in these languages (Dobrushina et al. 2001: 96).

(218) Chamalal (Bokarev 1949a: 94) (my glossing)
qχːʼajd-a / *qχːʼajd-be*
prepare-IMP.TR / prepare-IMP.ITR.SG
'Prepare (it)!' / 'Prepare yourself!'[123]

Imperative affixes are almost completely absent in the Kartvelian and Northwest Caucasian languages with the exceptions of Old Georgian (Fähnrich 1994: 85) and the polite imperative in Ubykh (Fenwick 2011: 128-129, 145), while in Megrelian the imperative and second person aorist are morphologically identical (Harris 1991b: 348). Imperative prefixes are noticeably rare in the Caucasus, as they appear to only be documented in Tat (Suleymanov 2020: 136), Juhuri (Authier 2012: 124), Talysh (Schulze 2000: 24) and Ubykh (Fenwick 2011: 129). Rostovtsev-Popiel describes the prefix *ko-* in Megrelian as an imperative marker (Rostovtsev-Popiel 2020: 562), while most other authors analyse this shared Zan prefix as an affirmative marker (Harris 1991b: 360; Holisky 1991: 437; Boeder 2000: 284-285; Öztürk & Pöchtrager 2011: 95), cf. section 6.6. A few verbs form imperatives with fossilised class prefixes in Aghul (Magometov 1970: 141) and Tabasaran (Babaliyeva 2013: 176-177). Doubled causative suffixes are used to express orders in Nogai (Csató & Karakoç 1998: 338).

(219) Aghul (Magometov 1970: 234) (my glossing)
bawa-s *ji-p,* *idžej* *xis-u* *balnicːa-ji*
mum-DAT IMP-tell, well be.INF-FUT hospital-ERG
'Tell mum that it will be fine in the hospital'[124]

(220) Şirvan Tat (Suleymanov 2020: 163) (reglossed)
un *ödömin=ä* *bi-yorn-ind* *bă_tän=män!*
that human=OBL IMP[125]-bring-2PL near=1SG.POSS
'Bring that person before me!'

[123] Glossed and transliterated from *къайда* and *къайдбе*, and translated from Russian 'приготовляй'/'готовься'.

[124] Translated from Russian 'Скажи маме, что в больнице будет хорошо'.

[125] Suleyman glosses it as MOD 'modal'.

The Prohibitive mood indicates that the clause is a prohibition or a negative command and it is typically the negative counterpart of the imperative mood (Aikhenvald 2010: 165-167; Bybee 1985: 173). Palmer argues that it is more 'illuminating' to label the prohibitive mood as a 'negative imperative' (Palmer 2001: 20), and although Aikhenvald mainly uses the term negative imperative, she also makes numerous convincing arguments that prohibitives are formally and syntactically distinct from imperatives in many languages (Aikhenvald 2010: 190-191). In the Caucasus most languages have specific prohibitive forms that are formally distinct from the imperative (Ganenkov & Maisak 2020: 116; Testelets 2020: 521), which supports the notion that the prohibitive and imperative are separate functions.

(221) Standard Dargwa (Van den Berg 2001: 176-177) (reglossed)
zigar=ma-r-ik'-ud, *di-la* *r-ig-usi* *xunul,*
moan=PROH-F-AUX-PROH[126], 1SG.OBL-GEN F-love-ADJ wife.ABS
w-ik'-uli *saj* *ɢurban*
M-say-CVB be.M Kurban
''Don't complain, my dear wife,' said Kurban'

(222) Talysh (Miller 1953: 143) (my glossing)
ti *də* *mə* *má-vo!*
2SG with 1SG.OBL PROH-go
'Don't go with me!'

The Kartvelian prohibitives are excluded from this thesis, as they are conventionally described as particles and not as verbal affixes, while Rostovtsev-Popiel describes the prohibitive *nu(m)-/ni(m)-* in Megrelian as a prefix (Rostovtsev-Popiel 2020: 562).

(223) Megrelian (Rostovtsev-Popiel 2020: 562) (gloss adapted from source)
iro *si* *nu[m]-Ø-ragad-an-k* *šxwa-s*
always 2SG PROH-2SG.S-talk-SM-2SG.S other-DAT
xolo *ko-Ø-u-rčkil-i*
ADD AFF[127]-3SG.IO-OV-listen-IMP
'Don't talk all the time yourself, listen also to others'

[126] Van den Berg glosses this suffix as FUT2, while this prohibitive form should rather be analysed as the circumfix *ma-…-ud*, cf. Forker (2020a: 298) and Daniel (2019: 88).

[127] Glossed as IMP by Rostovtsev-Popiel.

The Circassian prefix *mə-* is worth discussing, as it has a prohibitive function in all Circassian languages (Rogava & Keraševa 1966: 174; Smeets 1984: 300-301; Kumaxov 2006: 225; Konuk 2022: 525). It also negates non-finite verb forms, imperatives, optatives and interrogatives (Smeets 1984: 300-301, 314-320; Kumaxov 2006: 225; Kumakhov & Vamling 2009: 49; Konuk 2022: 525), which suggests that it should instead be analysed as a non-indicative negative prefix. The prefix *mə-* therefore fails to qualify as a prohibitive affix, as it simply is a non-indicative negation.

The Subjunctive/Irrealis modal functions

The Subjunctive or Irrealis moods are two modal categories that are both connected, albeit not identical, and they are evidently difficult to define as there is no clear consensus in the literature (Mauri & Sansò 2016). The subjunctive mood has traditionally been defined by its function in Latin and Romance languages (Palmer 2001: 108-111), which is problematic since the Latin subjunctive was clearly multifunctional. Bybee et al. (1994: 212) define it as 'verb forms or markers that obligatorily occur in certain types of subordinate clauses', which is unsatisfactory since the subjunctive can occur in main clauses as well (Palmer 2001: 108; Mauri & Sansò 2016: 174). The relationship between the irrealis and subjunctive mood is discussed by Palmer (2001), who concludes rather unsatisfactorily that there is 'not always a clear distinction' between the concepts indicative/subjunctive and realis/irrealis, and while it is 'basically the same' he also states that there are sufficient differences to discuss them separately (Palmer 2001: 145).

(224) Georgian (Bolkvadze & Kiziria 2023: 189) (my glossing)
net'av disert'acia-s v-c'er-d-e
if.only dissertation-DAT 1.S-write-IPFV-SUBJ
'Were I writing the dissertation'[128]

(225) Classical Armenian (Schmitt 2007: 133-134) (my glossing)
erkin-kʿ ew erkir anc ʿ-c ʿ-en,
heaven-NOM.PL and earth pass-AOR.SUBJ-3PL
ew ban-kʿ im očʿ anc ʿ-anic ʿ-en
and word-NOM.PL 1SG.POSS NEG pass-PRS.SUBJ-3PL
'Heaven and earth should pass, but my words shall not pass'[129]

The term subjunctive has conventionally been used for most languages of the Caucasus until recently, whereas the realis/irrealis distinction appears to have

[128] This is the original translation, although 'if only I were writing the dissertation' seems like a more appropriate English translation.

[129] Translated from German 'Himmel und Erde werden vergehen, aber meine Worte vergehen nicht'.

originated in Australian and Papuan linguistics in the 1970s (Bybee et al. 1994: 236). Smeets (1984) appears to be one of the first to use irrealis in an English-language Caucasian context, although labels such as irrealis-optative clauses, i.e. 'нереально-желательное предложение', and irrealis conditional clauses, i.e. 'нереально-условное предложение', are found already in Jakovlev (1940: 25) and Bokarev (1949b: 277) respectively.

The irrealis mood is defined in relation to the realis as 'distinguishing between actual and non-actual events' (Bybee et al. 1994: 236), which is further implied by Timberlake's distinction between indicative and subjunctive in Spanish as 'the indicative refers to a fact' and 'the subjunctive a possibility or potential' (Timberlake 2007: 327). Although Palmer makes a distinction between realis and irrealis and the notions 'factual' and 'non-factual', he states that the former are typically used to describe the latter (Palmer 2001: 149). Since the subjunctive and irrealis moods tend to indicate events that are both non-actual, i.e. counterfactual or hypothetical, or non-actualised, i.e. possible or potential, they are perhaps best analysed as instances of a general or non-specific irrealis mood.

Consequently, the multifunctionality described above is a defining feature of the subjunctive mood, as moods with a specific function should not be labelled as subjunctive. The irrealis mood could have a specific function, if it only indicates that an event is non-actual or non-factual. I have nonetheless chosen to merge these two moods into a subjunctive/irrealis mood category for comparative purposes as the terms are largely in complementary distribution. The relationship between the subjunctive and the conditional mood is also somewhat complicated, as e.g. Standard Dargwa typically uses the subjunctive mood for the conditional modal function as well (Musaev 2002: 72).

The Conditional modal function

The Conditional mood indicates that an event or a state, i.e. the *consequence* or *apodosis*, is conditioned by a *contingency* or *protasis* (Timberlake 2007: 321). The protasis typically corresponds to a subordinate 'if'-clause, while the apodosis is indicated by a finite conditional mood.

(226)

a. French (L'Huillier 1999: 124) (my glossing)
je viendrais [apodosis], *si je le pouvais* [protasis]
1SG come-COND.1SG, if 1SG 3SG.M.ACC can-IPF.1SG
'I would come if I could'

b. Georgian (Hewitt 1995: 586) (reglossed)
gušin rom e-c'vim-a [protasis]
yesterday if PrV-rain-PLUP.3SG,
šin da-v-rč-eb-od-i [apodosis]
at.home PV-1.S-remain-SM-IPFV-1/2SG.COND
'If it had rained yesterday, I would have stayed at home'

c. Georgian (Hewitt 1995: 584) (reglossed)
tu i-kux-eb-d-a [protasis],
if PrV-thunder-SM-IPFV-COND.3SG,
agretve i-elv-eb-d-a [apodosis]
also PrV-lighten-SM-IPFV-COND.3SG
'If it were to thunder, it would also lighten'

The pattern found in example (226) is not used in all Georgian conditional constructions, as the apodosis can also be modally 'unmarked' or both the protasis and the apodosis can be in the conditional mood as in example (226) (Hewitt 1995: 583-584). The protasis in example (226) is not modally neutral, as it too is conditioned, which demonstrates that the conditional mood by definition indicates the presence of an implicit or explicit protasis. I have therefore chosen to define the conditional mood as the explicit marking of the apodosis as seen in e.g. French. Some languages, such as Russian, Polish and Finnish, mark both the protasis and the apodosis in the same manner (Timberlake 2007: 325; Van Olmen & Van der Auwera 2016: 378; Karlsson 2018: 315-317), which would also qualify as conditional mood as these languages still explicitly mark the apodosis, cf. example (226c). The function of the conditional mood is at first glance identical to the conditional converb (cf. section 6.10.2), but there is a formal distinction that is explicit in some languages in constructions of the type 'if X (conditional converb), then Y (conditional mood)', where the conditional converb indicates the protasis and the conditional mood indicates the apodosis. The conditional mood in Turkic languages should therefore rather be analysed as conditional converbs as they

consistently mark the subordinate 'if'-clause, i.e. the protasis (Johanson 2022a: 57), cf. examples (227a-c).

(227)

a. North Azerbaijani (Širaliev 1971: 119) (my glossing)
ax, bir Bəhram-ı öz-üm-ə cəlb
oh, one Bahram-ACC self-1SG-DAT attract
edə bil-sə-jdi-m
do know-COND-COP.PST-1SG
daha qəm-im ol-maz-dı
more grief-1SG become-NEG.AOR-PST
'Oh, if I could attract Bahram for myself, my grief would be no more'[130]

b. Karachay-Balkar (Aliev 1973: 202) (my glossing)
börü qart bol-sa, džuburan-čï bol-a-dï
wolf old be-COND ground.squirrel-AN be-PRS-3SG
'If a wolf gets old, it becomes a molecatcher'[131]

c. Kumyk (Abdullaeva et al. 2014: 458) (my glossing)
eger sen biraz artïq kultur-alï bol-ĸan
if 2SG a.little more culture-ADJ be-PTCP.PRF
ed-ing bu-sa,
COP.PST-2SG be-COND,
ingilis-li bulan söjl-ep bol-ažaq ed-ing
English-ADJ with speak-CVB be-FUT COP.PST-2SG
'If you had been a little more cultured, you would have been able to speak with an Englishman.'[132]

Conditional modal suffixes are found in Megrelian (Harris 1991b: 345) and Laz (Holisky 1991: 421, 431, 435) while the conditional mood is expressed by combining perfective preverbs and the imperfect in Georgian (Hewitt 1995: 238-239). The irrealis mood in many Nakh-Dagestanian languages, e.g. Avar and Tsakhur, encodes the apodosis, which means that it at least partly functions as a conditional mood (Kibrik & Testelets 1999: 262; Forker 2020b: 259). The conditional mood described in Lezgian should rather be analysed as a conditional

[130] Translated from Russian 'Ах, если бы я смогла привлечь Бахрама, не было бы у меня горя'.

[131] Glossed and transliterated from *Бёрю къарт болса джубуранчы болады*, and translated from Russian 'Если волк состарится, то становится кротоловом'. The translation of *džuburan-čï* is lit. 'ground squirrel-er' or 'suslik-er'.

[132] Glossed and transliterated from *Эгер сен бираз артыкъ кулътуралы болгьан эдинг буса, ингйлисли булан сёйлеп болажакъ эдинг*, and translated from Russian 'Если бы ты был более культурным, ты бы смог говорить с англичанином'.

converb, as it encodes the protasis and not the apodosis (Haspelmath 1993: 394). The elaborate conditional verb forms found in the Dargic languages all encode the protasis, and although they conjugate for tense and person they are described as heading dependent clauses (Forker 2020a: 330), which means that they also should be analysed as conditional converbs. The conditional verb forms in the Dargic languages are therefore morphologically mood-like but functionally converb-like, which is interesting as it demonstrates the distinction between conditional mood and conditional converbs is not clear-cut. However, Haspelmath's definition of converbs as being non-finite would instead argue that these conditionals are simply subordinate mood.

The term 'eventual' mood has been used by Authier and Babaliyeva to describe moods in Juhuri and Tabasaran, which largely overlaps with the conditional modal function (Authier 2012: 187; Babaliyeva 2013: 192-194, 195-196). An irrealis conditional mood has furthermore been described in Kubachi (Magometov 1963: 207), cf. the irrealis conditional converb in section 6.10.2, but Magometov also analyses it as a past conditional which is probably a better label.

The Potential modal function

The Potential mood indicates that the validity of an event or a state is 'probable', 'possible', or 'potential', and is typically indicated by 'may', 'might', or 'probably' in English (Bybee 1985: 178). The label potential is also often used to indicate that 'the subject is able to' (Bybee 1985: 168), which should be analysed as a separate function. Since these two functions are indiscriminately referred to as the potential mood in various descriptions, it is difficult to assess which function is the 'core' potential mood. The two functions could be differentiated by introducing a distinction between *potential/probability* and *potential/ability*, cf. Palmer's term *abilitive* (Palmer 2001: 10).

In the Caucasus the potential mood exclusively refers to the function potential/ability, so I have chosen to operationalise the potential modal function as indicating ability or possibility to do something, thereby corresponding to 'can', 'may' or 'be able to'. I have found one language with an affix explicitly expressing the function potential/probability, i.e. Abzakh Adyghe (Paris 1989: 188; Konuk 2022: 231).

(228) Abzakh Adyghe (Konuk 2022: 231) (reglossed)
zə *tawrəχ* *gʷere* *qə-p-fe-s-ʔʷete-n*
one story any PROX-2SG.OBL-BEN-1SG.ERG-tell-PROB
'I will probably tell you a story/I would like to tell you a story'

Nichols chooses to replace the previous label potential in Ingush with 'inceptive' (Nichols 2011: 484), while the function of the mood is 'virtually synonymous to constructions with *mog* 'can'' (Nichols 2011: 491), thus having at least partly a potential/ability function, which also becomes apparent from the examples given.

Alekseev & Radzabov use the term 'possibilitive' to describe the suffix *-ɬ* in Tsez (Alekseev & Radzabov 2004: 147), which appears to have a potential/ability function, while previous Russian-language literature uses the term potential (Imnaišvili 1963: 171). Creissels (2010) lists a potential in Northern Akhvakh, but it is analysed as future tense by Magomedbekova (1967: 88-89).

(229) Tsez (Imnaišvili 1963: 229) (my glossing)
mi *idu* *Ø-uqˤɬ-o* *nesi* *žekʼu-r*
2SG at.home I-hide-IMP DIST.OBL man-LAT
Ø-ikʷadā-ɬ-čʼi-ru-χor
I-be.seen-POT-FUT.NEG-PTCP.PST-MNR.CVB
'Hide yourself at home, in such a way that the man will not be able to see you'[133]

The Hortative/Exhortative and Jussive modal functions

The Hortative, Exhortative or Jussive mood indicates that the 'speaker is encouraging or inciting someone to action' (Bybee et al. 1994: 179) and typically corresponds to constructions of the type 'let's X'. The term Cohortative mood, which indicates that the speaker is included in the encouragement, is also frequently encountered, and the term Adhortative mood has been used to describe 'ought to'-constructions in Udi (Schulze 1982: 165). The hortative is similar to the imperative, where the hortative is an 'invitation' rather than an order (Timberlake 2007: 328), and a theoretical distinction is made between first person (co-)hortatives, 'true' second person imperatives and third person jussives (Nikolaeva 2016: 76).

(230) Kryts (Authier 2009: 147) (gloss adapted from source, retransliterated)
dust *xi-day!*
friend be-HORT.2PL.INCL
'Let us be friends!' (cohortative)

(231) Ubykh (Fenwick 2011: 131) (original glossing)
ɐ-j-kʲʼɜ-jɐqʷʼ
3SG.ABS-PV-go-HORT
'Let him come' (exhortative)

Although many languages lack a specific hortative mood, they often express hortative functions with other moods, e.g. Eastern Armenian, where the subjunctive future also conveys a hortative function (Dum-Tragut 2009: 239).

[133] Translated from Russian 'спрячься дома так, чтобы не смог увидеть (тебя) тот человек'.

(232) Eastern Armenian (Dum-Tragut 2009: 239) (original gloss)
gn-ank^{h} *tun!*
go-SUBJ.FUT.1PL house.NOM
'Let's go home!'

The Debitive and Necessitative modal functions

The Debitive, Necessitative or Obligative moods are closely related moods indicating that an action is necessary or obliged, typically corresponding to constructions of the type 'need to X' and 'must X' (Bybee et al. 1994: 177; Palmer 2001: 27-28). The distinction between the necessitative and the debitive/obligative moods is potentially an artefact of English language linguistics, as there are no languages in the Caucasus with both necessitative and debitive affixes, implying that the distinction is largely terminological or that one of these moods is expressed by modal verbs instead. To complicate matters further, the aforementioned mood in Turkic is labelled necessitative and obligative synonymously (Johanson 2022a: 39), which is unsurprising since it covers both functions in e.g. North Azerbaijani (Ragagnin 2022: 253).

(233) North Azerbaijani (Ragagnin 2022: 253) (my glossing)
götür-məli-yəm
take-DEB-1SG
'I must take'

The debitive mood in Eastern Armenian is primarily expressed by periphrastic constructions, but there is also the specific future participle suffix *–ik^{h}* whose function is inherently debitive (Dum-Tragut 2009: 207). Necessitative affixes have been described in North Azerbaijani (Ragagnin 2022: 253), Kumyk (Abdullaeva et al. 2014: 317), Abkhaz (Chirikba 2003a: 47), Shapsug Adyghe (Smeets 1984: 278), Aghul (Magometov 1970: 143), Sanzhi Dargwa (Forker 2020a: 254-255) and Northern Talysh (Schulze 2000: 24), while debitive/obligative affixes have been described in Eastern Armenian (Dum-Tragut 2009: 207), Tabasaran (Babaliyeva 2013: 198), Archi (Kibrik 1994a: 327), Kryts (Authier 2009: 146) and Itsari Dargwa (Sumbatova & Mutalov 2003: 92).

(234) Eastern Armenian (Dum-Tragut 2009: 207) (reglossed)
t^{h}argman-v-el-ik^{h} *ēdžer-ə*
translate-PASS-INF-PTCP.FUT/DEB page-PL-DEF
nš-v-ac *en* *karmir-ov*
mark-PASS-PTCP.RES COP.3PL red-INS
'The pages that must be translated are marked with red (colour)'

The Mirative modal function

The Mirative or Admirative mood indicates that 'the information contained in the utterance is new or presumed to be unexpected to the addressee' (Squartini 2016: 59), typically relating to an 'unprepared mind, new information, and speaker's surprise' (Aikhenvald 2004: 8). The nature of the mirative has been discussed as either evidential, modal or a separate grammatical category (Aikhenvald 2004: 8; Squartini 2016: 60). I have chosen to analyse it as a mood, since it primarily encodes subjective experience of an event. Affixes expressing a mirative function have been described in Abkhaz (Chirikba 2003a: 48), Ubykh (Fenwick 2011: 118), Khinalug (Kibrik 1994b: 393) Lak (Friedman 2020: 230) and Northern Akhvakh (Creissels 2010: 110).

(235) Northern Akhvakh (Creissels 2010: 122) (gloss adapted from source, retransliterated)

č'-īdeɬːi	*gutła*	*mene,*
sow-SIM.CVB	COP.NEG.M	2SG
m-aqːʷ-ideɬːi	*gutła*	*mene,*
NH-dig-SIM.CVB	COP.NEG.M	2SG
qːʼ-ōnutła	*čugu*	*w-oqʼ-uwa?*
eat-INF	why	M-come-MIR.M

'You are not here when we sow, you are not here when we dig the ground, why are you coming to eat?'

The Dubitative modal function

The Dubitative mood indicates that the speaker expresses 'an element of doubt that the event described in the proposition occurred or will occur' (Bybee 1985: 179). The dubitative is occasionally described as an evidential due to its function being 'evidential in nature', cf. the dubitative in Algonquian languages (Junker, Quinn & Valentine 2018: 435). However, the dubitative should be analysed as a mood since it arguably belongs within the traditional definition of mood as conveying the 'speakers' (subjective) attitudes and opinions' (Bybee et al. 1994: 176). Dubitative affixes and clitics are found in a small number of languages in the Caucasus, e.g. Adyghe (Rogava & Keraševa 1966: 187), Lak (Friedman 2020: 229), Archi (Kibrik 1994a: 330; Chumakina 2020: 302) and Hinuq (Forker 2013: 427).

(236) Archi (Kibrik 1994a: 330) (my glossing)

o<w>-qi-ši	*edi-čugu*		*za-ri*
<I.SG>do-POT-CVB	IV.SG.be.PFV-DUBIT		1SG-ERG
jamu-t	*ari*	*sːanʁi*	
DEM-IV.SG.ABS	work(IV).ABS	yesterday	

'I doubt that I would have done the work yesterday'

The Permissive modal function

The Permissive mood indicates that the speaker permits a state or an event (Palmer 2001: 10), which makes the permissive mood functionally similar to the jussive mood. The term 'non-curative' has been used to describe a functionally identical mood in Itsari Dargwa (Sumbatova & Mutalov 2003: 98) and the permissive mood has also been described in Xaidaq (Temirbulatova 2004: 172-173). In Nogai, the causative also appears to express permissive functions (Karakoç 2022: 360).

(237) Itsari Dargwa (Sumbatova & Mutalov 2003: 98) (reglossed)
dehni *wajhat* *b=uʕ-ikːa*
child-PL for.a.walk HPL=go.PFV-PERMIS
'Let the children go for a walk'

Other modal functions

Abkhaz and Abaza have a potential mood that Hewitt describes as a 'suffix characterising the wretched status of any of the verb's arguments but especially of the subject' (Hewitt 1989: 51; Lomtatidze & Klychev 1989: 109), which I suggest could be referred to as a Commiserative mood, while this might not be a mood at all. Cf. the Adyghe and Kabardian particle (-)*gʷɕe* expressing regrettable or unfortunate events (Rogava & Keraševa 1966: 306) and 'compassion, empathy, condolence' (Kumaxov 2006: 365).

(238) Abaza (Lomtatidze & Klychev 1989: 109) (my glossing)
d-gʷərɣʲa-gʷša-p'
3SG.H.ABS-rejoice-COMMIS-STAT.PRS.FIN[134]
'(s)he will rejoice, poor thing!'

The Apprehensive mood indicates 'the undesirability of an event, and the need to avert it' (Timberlake 2007: 329). Apprehensive affixes have been described in Mehweb (Dobrushina 2019b: 154) and Northern Akhvakh (Creissels 2018: 168).

(239) Mehweb (Dobrushina 2019b: 154) (original glossing)
d-arʔ-a *mura,*
NH.PL-gather:PFV-IMP.TR hay
zab *d-aq'-a-la*
rain NH.PL-do:PFV-IRR-APPR
'Collect the hay, it might rain.'

[134] Lomtatidze & Klychev gloss the suffix *-p'* as finite, stative present, while O'Herin glosses it as present indicative.

A Deliberative mood has been described in Khwarshi (Khalilova 2009: 255). According to Khalilova it only occurs in questions, but it primarily appears to convey some sort of indirect question or wondering.

(240) Khwarshi (Khalilova 2009: 152) (original glossing)
idu *ito-q'e-k* *Ø-uh-alu* *tɬin*
this.ABS when-QUES-QUES[135] I-die-DELIB QUOT
b-eč-un *izzu*
HPL-be-UWPST that.PL.(P).ABS
'They wondered when he would die.'

Kibrik has described an Approbative mood in Archi as 'Event P is true and the speaker approves of P but at the same time there is another event that contradicts P' (Kibrik 1977: 211; Kibrik 1994a: 331).

6.4. Evidential functions

Evidentiality is a grammatical category 'whose primary meaning is source of information' (Aikhenvald 2004: 3). It is found to some extent in all three endemic language families of the Caucasus but most prominently in the Nakh-Dagestanian languages (Friedman 2018: 134-138; Forker 2018a: 492). Evidentiality is also generally present in the Turkic languages, where it has traditionally been described as *indirectivity* (Johanson 2018). The situation in Azerbaijani is somewhat complicated, likely by the long-standing influence of Persian as the language of prestige, since the evidential suffix *–mIš* has largely lost its evidential function unless it is reduplicated (Johanson 2018: 514, 518-519). The best described Turkic evidential system in the data is likely Nogai (Karakoç 2022: 361), while the suffix *–GAn* has an evidential function in Karachay-Balkar (Johanson 2018: 514) and to some extent in Kumyk (Abdullaeva et al. 2014: 335). It is also found in Eastern Armenian (Dum-Tragut 2009: 199) and Juhuri (Authier 2012: 227).

Evidentiality is a diverse and multi-faceted category, and it is not always easy to group the various evidential forms that are present in the languages of the Caucasus, as they tend to overlap and have slightly different nuances of meaning. Many older grammatical descriptions lack the concept of evidentiality altogether, which means that the overview presented in this thesis is by no means exhaustive.

[135] This glossing is puzzling, but it is Khalilova's glossing.

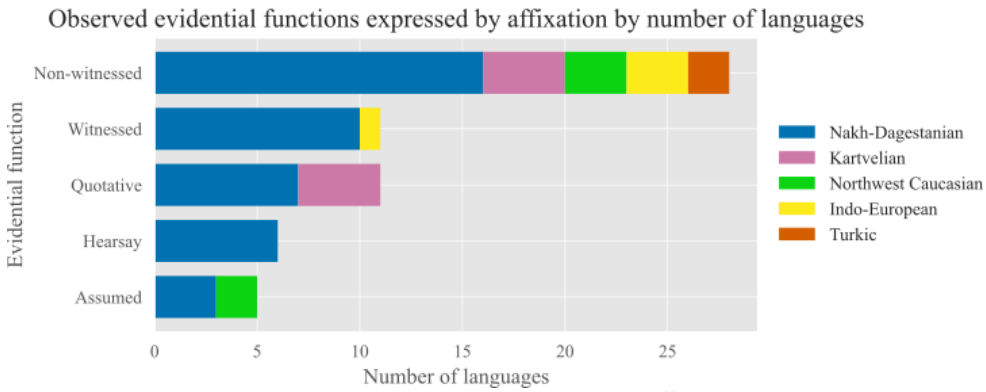


Figure 6.4: All observed evidential categories expressed through affixation in the data according to number of languages.

The Non-witnessed/Indirect evidential function

The Non-witnessed or Indirect evidential indicates that the event was not witnessed directly by the speaker, which typically makes it functionally identical to Aikhenvald's *non-firsthand* or *non-eyewitness* evidential (Aikhenvald 2004: 25). Aikhenvald also mentions the *non-visual sensory* evidential, which is different from the non-witnessed evidential as it covers information retrieved from the other senses (Aikhenvald 2004: 63), whereas the source of the non-witnessed or indirect evidential is by definition not specified. Non-witnessed/indirect evidentials typically occur as a binary distinction with the witnessed/direct evidential in the TAM systems in Nakh-Dagestanian (Forker 2018a: 498; Polinsky 2020: 23) and Kartvelian (Harris 1991a: 51), although the terms evidential (non-witnessed) and non-evidential (witnessed) have been used within the Kartvelian tradition.[136] The Georgian perfect typically expresses past events that have not been witnessed by the speaker (Boeder 2000: 285), and the perfect in Eastern Armenian has also been described as indicating non-witnessed past events (Kozintseva 2000: 410; Dum-Tragut 2009: 214).

(241) Bezhta (Forker 2018a: 411) (gloss adapted from source)

c'ohor	*žüy-ü-ʔ-tɬ'ä:*	*bitło-ʔ*	*Ø-etɬ'e-na*
thief	window-OBL-IN-TRANS	house-IN	I-go-UWPST

'The thief has entered the house through the window'

[136] Cf. also the Georgian evidential term *turmeobiti* for the Georgian perfect, which is derived from *turme* 'evidently, apparently' (Hewitt 1995: 93).

(242) Georgian (Bolkvadze & Kiziria 2023: 266) (my glossing)
čven-s mezobel-s axali mankana u-q'id-i-a
our-DAT neighbour-DAT new car.NOM OV-buy-PFV-3SG
'Our neighbour (apparently) has bought a new car'

(243) Eastern Armenian (Kozintseva 2000: 411) (reglossed)
her-s patm-um ēr, or im bolor
father-1SG.POSS tell-PTCP.IPFV COP.PST.3SG, that my all
pap-er-n ays kulay-ic en gini xm-el
ancestor-PL-DEF this bowl-ABL COP.3PL wine drink.PTCP.PFV
'My father told that all my ancestors had drunk wine from this bowl'

The Inferred or Inferential evidential is closely connected to the non-witnessed evidential, as it indicates that the referenced event is 'based on visible or tangible evidence, or result' (Aikhenvald 2004: 63), but the event itself has not been observed. Inferentials have been described in Abkhaz and Abaza (Chirikba 2003a: 47; O'Herin 2020: 478), but they are functionally more or less identical with the indirect and non-witnessed evidentials in Nakh-Dagestanian and Kartvelian (Chirikba 2003b: 245-246; Aikhenvald 2004: 29). This analysis is further supported by the presence of a non-witnessed evidential in Adyghe (Rogava & Keraševa 1966: 182). Inferential evidential forms have also been described in Ingush and Avar, but they are not formed by dedicated affixes (Molochieva & Nichols 2018: 39; Forker 2018b: 199). The category inferential evidential is therefore mainly relevant in evidential systems with more than two categories.

(244) Abkhaz (Chiribka 2003a: 47) (reglossed, retransliterated)
á-laʂara-[a-a]χʲ, á-mca-[a-a]χʲ d-an-áa-j-Ø,
DEF-light-ALL DEF-fire-ALL 3SG.F.ABS-TEMP-come-AOR.NFIN
lə́-la-kʷa Ø-q'apʃ-dzá jə́-q'a-n,
3SG.F.POSS-eye-PL 3PL.NH.ABS-red-ADV 3PL.NH.ABS-be-PST
d-tɕʷ'ə́wa-zaarən
3SG.F.ABS-cry-INFER
'When she came up to the light, to the fire, her eyes were very red; apparently, she had been crying'

The Witnessed evidential function

The Witnessed or Visual evidential indicates that the referenced event or state 'covers information acquired through seeing' (Aikhenvald 2004: 63). The category generally overlaps with the notion of *direct evidential* in the Caucasus and the witnessed/direct evidential tends to be integrated into the TAM system (Forker 2018a: 502). A witnessed/non-witnessed evidential distinction is found in all living Kartvelian languages, where series I and II typically express witnessed events and series III and IV express non-witnessed events (Harris 1991a: 51; Tuite 1998: 205),

while Boeder explicitly states that the aorist 'is neutral: it doesn't tell whether the speaker witnessed the fact or not' (Boeder 2000: 286). However, the aorist in Eastern Armenian has been described as conveying witnessed past events (Dum-Tragut 2009: 231).

(245) Chechen (Molochieva 2010: 217) (gloss adapted from source)

ahw	*dwa-miel-i*	*sa(n)*	*chai*
2SG.ERG	away-drink-PFV.WPST	1SG.GEN	tea

'You have drunk my tea'

(246) Tsez (Comrie & Polinsky 2007: 338; Khalilova 2011: 36) (gloss adapted from source)

a. *obiy-ā* *madina-r* *k'icu* *y-is-si*
father-ERG Madina-LAT strawberry(II) II-buy-WPST
'Father bought strawberries for Madina (the speaker saw this)'

b. *obiy-ā* *madina-r* *k'icu* *y-is-no*
father-ERG Madina-LAT strawberry(II) II-buy-UWPST
'Father bought strawberries for Madina (the speaker did not see this)'

The Quotative function

The Quotative indicates that the speaker quotes 'reported information with an overt reference to the quoted source' (Aikhenvald 2004: 64), and it typically reports direct speech verbatim (Aikhenvald 2018: 43). Quotative enclitics are found in the Tsezic languages Tsez (Forker 2018a: 506), Khwarshi (Khalilova 2009: 257), Hinuq (Forker 2013: 434), while a quotative suffix has been described in Archi (Kibrik 1994a: 332; Chumakina 2020: 313). Further examples of quotative particles or affixes in Nakh-Dagestanian are Bagvalal *-ēna* (Daragan & Maisak 2001: 177), Chamalal *-daq* (Magomedova 2004: 58) and Karata *-tł'ːe* (Magomedbekova 1971: 173).

(247) Archi (Chumakina 2020: 313) (original glossing)

ju-w-mi-r-ši		*bo-li*
that-I.SG-SG.OBL-CONT-ALL		say.PFV-EVID
un	*daki*	*w-eːˁ-t'o-r*
2SG.ABS	why	I.SG-come-POT.NEG-QUOT

'They said to him: "Why wouldn't you come?"'

(248) Karata (Magomedbekova 1971: 173) (my glossing)

wudu-w	*harge*	*wu-k'-alda-tł'ːe,*	*anł-a*	*di-ja*
DEM-I	here	I-be-UWPST-QUOT,	hear-PST	1SG.OBL-DAT

'I heard it was said that he was here'[137]

[137] Translated from Russian: '[Я] слышал, что он был здесь, мол'.

Quotative particles are found in all Kartvelian languages, where Georgian and Megrelian even encode which grammatical person is quoted (Harris 1991b: 381-382; Hewitt 1995: 614). Arhavi Laz has a similar system where the third person particle *–ya* occurs in free variation with the first person *–ma* for first person referents, while Lacroix only encountered one example of the second person particle *–şo* (Lacroix 2009: 725-726). The first person quotative particle *–metki* was also present in Old Georgian (Fähnrich 1994: 214). Svan has the quotative enclitic particle *-ədž*/*-idž*, which always attaches to the word immediately before the verb (Tuite 1997: 40).

(249) Georgian (Boeder 2000: 279) (gloss adapted from source)
mic'a-ši-o *okro-s* *kila* *iq'-o-o*
earth-IN-QUOT gold-GEN jar.NOM be-AOR.3SG-QUOT.3SG
'In the earth, there was a gold jar, it is said'

(250) Megrelian (Harris 1991b: 382) (gloss adapted from source)
zɣva-ša *gegnvoʔot(i)-mak*
sea-ALL throw-QUOT.1SG
''I threw it into the sea,' I said'

(251) Laz (Lacroix 2009: 725) (reglossed)
guy-çkimi-ten *"p'eya* *min* *oren-ma?"*
heart-1SG.POSS-INS DELIB who be.PRS.3SG-QUOT.1SG
b-zop'on-t'i
1SG-say-IPF
'I said to myself (lit. with my heart) 'Who is it then?''[138]

The Hearsay evidential function

The Hearsay evidential indicates that the utterance is 'reported information with no reference to those it was reported by' (Aikhenvald 2004: 64). The non-witnessed or indirect evidential is frequently described as indicating information retrieved from hearsay, but the indirect evidential does not exclusively refer from information retrieved from hearsay. The term 'narrative' has been used for the enclitic particles with a similar function in Hinuq and Khwarshi, as it is primarily used in narratives (Khalilova 2009: 237; Forker 2013: 313-314; Forker 2018a: 506). Forker (2018a) also describes the particle *kʷ'an* as an enclitic hearsay evidential based on examples from Magometov (1982), but Daniel et al. (2019) analyse it as a quotative non-enclitic particle. Affixes or cliticised particles expressing a hearsay evidential function have been described in a few Dagestanian languages, e.g. Avar (Charachidzé 1981: 135; Forker 2018a: 507; Forker 2018b: 201), Lezgian

[138] Translated from French: 'En moi-même, je me disais : 'Qui est-ce donc ?''.

(Haspelmath 1993: 148), Tsakhur (Schulze 1997: 50), Tsez (Khalilova 2011: 43; Forker 2018a: 506), Khwarshi (Khalilova 2009: 237) and Hinuq (Forker 2013: 313-314).

(252) Lezgian (Haspelmath 1993: 148) (gloss adapted from source)
qːe *sobranie* *že-da-lda*
today meeting be-FUT-HSAY[139]
'They say there will be a meeting today'

(253) Tsez (Khalilova 2011: 43) (reglossed)
žin *č'agu* *joł-tłaχ* *ža* *baɦarči*
today alive be-HSAY[140] he brave.young.man
'They say he is still alive'

The Assumed/Assumptive evidential function

The Assumed or Assumptive evidential indicates that the referenced event is 'based on evidence other than visible results: this may include logical reasoning, assumption, or simply general knowledge' (Aikhenvald 2004: 63), and it typically corresponds to the adverb 'probably' in past tense contexts. The assumed evidential has previously been analysed as an 'assumptive mood' in e.g. Kabardian (Kumaxov 2006: 220; Kumakhov & Vamling 2009: 46). Although not recognised as such, the suffix *-ʒ-ən* glossed as 'ITER-PROB' in Abzakh Adyghe appears to convey an assumptive function (Konuk 2022: 235), cf. example (254), and it is possibly cognate with the Standard Adyghe assumptive *-ʁen* (Rogava & Keraševa 1966: 186), cf. the Kabardian example (255).

(254) Abzakh Adyghe (Konuk 2022: 235) (reglossed)
təʁʷɐsɜ *jə-ʔʷɜf* *jə-wəxə-ʁɐ-ʁɜ-mɜ*
yesterday 3SG.POSS-work 3SG.ERG-finish-PFV-PST-COND.CVB
wənɜ-m *ʒɜw* *qɐ-kʔʷɜ-ʒən-əj*
house-OBL early PROX-go-ASS(ITER.PROB)-PFV
'If he had finished his work yesterday, he would probably have been home early'[141]

[139] Haspelmath uses the gloss EVID 'hearsay evidential'.

[140] Khalilova glosses it as NARR 'narrative'.

[141] Translated from French 'S'il avait fini son travail hier, il serait probablement rentré tôt à la maison'.

(255) Kabardian (Kumakhov & Vamling 2009: 47) (reglossed)
də-kʷ'e-ʁen-ɕ
1PL.ABS-go-ASS-IND
'We probably walked'

The suffixes or enclitics *-tːeħari* in Aghul and *-kːar* in Lak convey similar evidential functions that appear to be assumptive (Magometov 1970: 155; Friedman 2020: 229). The suffix *-ɢˁaˁ* in Tsakhur is described as an assumed evidential, but no examples are given (Ibragimov 1990: 129-130; Schulze 1997: 50).

(256) Aghul (Magometov 1970: 155) (my glossing)
k'i-tːeħari
die-ASS
'Apparently, he died'[142]

6.5. Intentionality-marking functions

Intentionality is a grammatical category that encodes speaker intentions and whether actions are intentional or not. It is found in a limited number of Nakh-Dagestanian and Northwest Caucasian languages. There are generally two types of affixes that indicate intentionality in the data, i.e. affixes indicating intentional actions and affixes indicating unintentional actions. Affixes indicating intentional actions have been described in Khwarshi (Khalilova 2009: 195), Hinuq (Forker 2013: 201-202, 298), Lak (Friedman 2020: 222), Andi (Salimov 2010: 194), Karata (Magomedbekova 1971: 175) and Ubykh (Fenwick 2011: 118). Intentional forms often interact with future reference, as Hinuq, Khwarshi, Lak and Ubykh all have intentional future affixes (Khalilova 2009: 195; Fenwick 2011: 118; Forker 2013: 201-202; Friedman 2020: 222). The intentional future in Hinuq does not entail the certainty of the future event however, as it can be combined with the dubitative enclitic *–m*, cf. example (257).

(257) Hinuq (Forker 2013: 202) (gloss adapted from source)

di-qo	*hag*	*χu*	*r-ac'-eɬ-an=e=m*
1SG.OBL-CONT[143]	that	meat(V)	V-eat-POT-INT.FUT=Q=DUBIT

'Will I be able to eat that meat?'

[142] Translated from Russian 'кажется, умер'.

[143] Forker glosses *–qo* as AT 'location 'at''.

(258) Ubykh (Fenwick 2011: 130) (gloss adapted from source)
zɜ-fɨ:ʃʷɜ-ʁɜ *ʃɨ-Ø-ɕɜ-kʲɜ-n[ɜ]-ɜw-mɜ*
one-eating.place-LOC 1PL.ABS-3SG.OBL-PV-enter-PL-INT.FUT-CVB
jɜ-ʃ-f-ɜw
NULL.ABS-1PL.ERG-eat-INT.FUT
'Let's (lit. 'we will') go into a restaurant and eat'

The intentional future suffix *-alaha* in Khwarshi and the intentional suffix *-(a)ru* in Hinuq can be used with past reference if combined with past tense auxiliaries (Khalilova 2009: 196; Forker 2013: 298), which indicate that intentionality is a separate grammatical category in these languages.

(259) Khwarshi (Khalilova 2009: 196) (gloss adapted from source)
ise *soyro* *b-ez-alaha* *b-eč-i,*
that.ERG horse(III) III-buy-INT III-be-WPST
os *m-uⁿ-ɬo*
money (III) III-be.enough-COND.CVB
'He was going to buy a horse, if there were enough money'

The second category of affixes indicating unintentionality is primarily found in the Northwest Caucasian languages, e.g. in Kabardian (Kumaxov 2006: 157; Kumakhov & Vamling 2009: 190), Adyghe (Rogava & Keraševa 1966: 281), Shapsug Adyghe (Smeets 1984: 260), Abkhaz (Chirikba 2003a: 38) and Abaza (Lomtatidze & Klychev 1989: 117).

(260) Adyghe (Rogava & Keraševa 1966: 281) (my glossing)
ɬ'ə-m *ħe-r* *ʔetɕ'e-wətɕ'a-ʁ*
man-OBL dog-ABS UNINT-kill-PST.PFV
'The man unintentionally killed the dog'[144]

The potential suffix in Khwarshi and Hinuq can also convey unintentional or accidental actions (Khalilova 2009: 306; Forker 2013: 503). The case of the agent in these unintentional constructions is typically not ergative, cf. the involuntary agent case function described in section 5.1.1.2., which also holds true for the potential constructions, cf. example (261).

(261) Khwarshi (Khalilova 2009: 306) (reglossed)
di-qo *zihe* *kok-l-i*
1SG.OBL-CONT cow eat-POT/UNINT-WPST
'I could make the cow eat'/'I made the cow eat accidentally'

[144] Glossed and transliterated from *ЛIым хьэр IэкIэ-укIагъ*, and translated from Russian 'Мужчина невольно убил собаку'.

6.6. Polar functions

Polarity is a grammatical category that is primarily realised as negative affixes in the Caucasus. 52 of the 56 languages in the data have been described as having negative verbal affixes, and negative affixes are found in all five language families of the Caucasus. Negative affixes are seemingly lacking in Iron Ossetic (Erschler 2020: 657) and Bats (Dešeriev 1953; Holisky & Gagua 1994), while Georgian and Svan lack negative affixes except for the largely derivational privative/negative participles, e.g. Georgian *da-u-vic’q’-ar-i* ‘unforgettable’ (Hewitt 1995: 433) and Svan *u-maːr-a* ‘unprepared’ (Schmidt 1991: 531).

(262) Classical Armenian (Van Damme 2004: 103) (my glossing)

ew	*ardzak-el*	*z-do-sa*	*nawtʿ-is*[145],
and	send.away-INF	ACC–DEM.PL-ACC	hungry-ACC.PL
čʿ-kam-im			
NEG-want-PRS.1SG			

‘and I will not send them away hungry’

(263) Budukh (Authier 2010: 155) (reglossed)

ye-z	*ts’eˤ*	*(şıma)*	*sorkú*	*ha<va>ts’ar-da-b*
1PL-DAT	goat(NH)	how	slay.DEB	know<NH>-PRS.NEG-NH

‘We do not know how to slay a goat’

(264) Karachay-Balkar (Aliev 1973: 106) (my glossing)

alma	*tereg-in-den*	*uzaq*	*tüš-me-j-di*
apple	tree-3SG.POSS-ABL	far	fall-NEG-PTCP.PRS-3SG
de-j-di-le			
say-PTCP.PRS-3SG-PL			

‘They say that the apple does not fall far from the tree’[146]

The Northwest Caucasian languages make a distinction between finite and non-finite negation, as the Circassian languages have the non-finite negative prefix *mə-* and the finite negative suffix is *–(e)p* in Standard Adyghe, Abzakh Adyghe and Shapsug Adyghe (Rogava & Keraševa 1966: 202; Smeets 1984: 283; Paris 1989: 198) and *-q’əm* in Kabardian (Kumakhov & Vamling 2009: 49).

[145] Van Damme gives the form *նութիս* (nutʿis), which most likely is a misreading of *նաւթիս* (Meillet 1936: 94).

[146] Glossed and transliterated from *Алма терегинден узакъ тюшмейди дейдиле*, and translated from Russian ‘Говорят, яблоко от яблони далеко не падает’.

(265) Adyghe (Rogava & Keraševa 1966: 243) (my glossing)
maʂʷ'a-m *wə-rə-mə-dʒegʷə,*
fire-OBL 2SG.ABS-PV-NEG(NFIN)-play
p'etɕ'e-tɕ'ə-me
PV-leave-COND
qə-p-fe-wəbətə-ʐə-ɕt-ep
PV-2SG.ABS-POT-catch-REP-FUT-NEG(FIN)
'Don't play with fire, if it flares up, you will not be able to put it out again (lit. if it leaves out of your hands, you will not be able to catch it)'[147]

An important sub-group of negative verbal affixes is prohibitive affixes that indicate the negative imperative, and they are particularly common in the Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 116), cf. the prohibitive mood in section 6.3.

The Affirmative polar function

Affirmative or Confirmative affixes are also found in a small number of Kartvelian and Northwest Caucasian languages in the Caucasus, e.g. Adyghe (Rogava & Keraševa 1966: 238), Shapsug Adyghe (Smeets 1984: 276), Megrelian (Harris 1991b: 360; Boeder 2000: 284-285; Rostovtsev-Popiel 2020: 563) and Laz (Holisky 1991: 437; Öztürk & Pöchtrager 2011: 95).

(266) Laz (Öztürk & Pöchtrager 2011: 95) (gloss adapted from source)
a. *ko-b-dzir-i*
AFF-1.S-see-PST.1SG
'I certainly saw (it)'
b. *do-p'-t'ax-i*
AFF-1.S-break-PST.1SG
'I certainly broke (it)'

(267) Shapsug Adyghe (Smeets 1984: 277) (gloss adapted from source)
sə-kʷ'e-pa-ʁ
1SG.ABS-go-AFF-PFV
'I did go away'

[147] Glossed and transliterated from *МашIом уры-мы-джэгу, пIэкIэкIымэ къыпфэубытыжьыщт-эп*, and translated from Russian 'Не играй с огнем, если (он) разгорится, (ты) не сможешь его потушить (букв.: «если он уйдет из твоих рук, то не сможешь его поймать!»)'.

6.7. Person and number-marking functions

The person agreement affixes indicate the arguments of a verb, i.e. subjects, direct objects and indirect objects. Person-marking affixes are common across the Caucasus, but person agreement is less widespread and usually not as complex in the Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 123). Person typically interacts with number, and singular and plural were the only grammatical numbers found in these languages, as none of the Indo-European languages of the Caucasus have retained their dual number, since it was lost already in Classical Armenian (Meillet 1936: 93), Parthian and Middle Persian (Durkin-Meisterernst 2014: 230), cf. section 5.2. This section only describes person-marking affixes and not person agreement, as agreement would also include gender or noun class, while I discuss gender/noun class affixes in section 5.5.

6.7.1. Person and number-marking of subjects

All three persons are not equally expressed by affixation in the Caucasus, as 41 languages have some form of first-person affixes, 37 languages have second-person affixes and only 32 languages have third-person affixes.

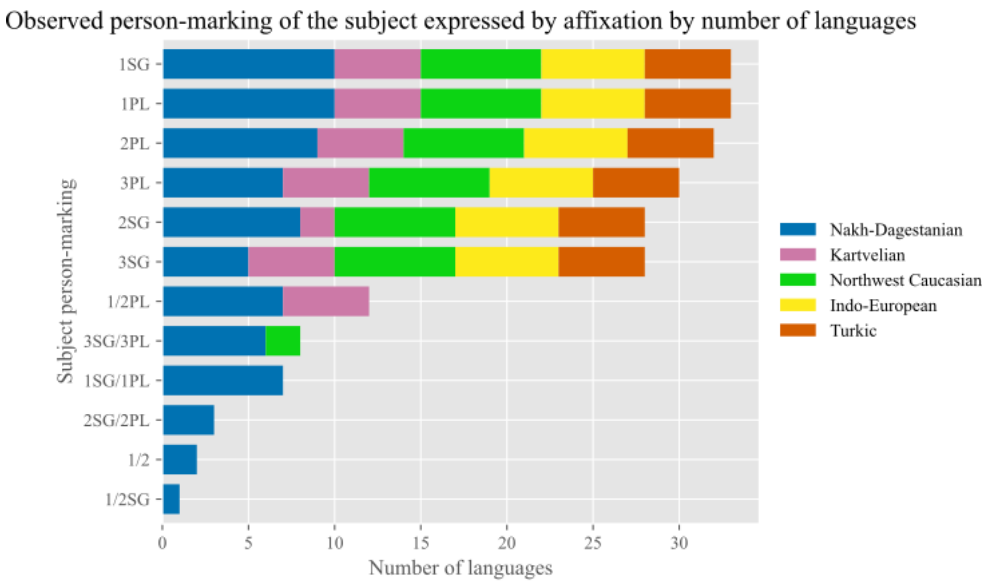


Figure 6.5: Person-marking of subjects expressed through affixation in the data according to number of languages.

This could potentially be explained by the tendency for particularly the third person singular to be zero-marked in verb paradigms, but this is not universally the case, as

it is the second person singular subject that is zero-marked in certain Georgian verb paradigms (Tuite 1998: 73).

(268) Georgian (Bolkvadze & Kiziria 2023: 150) (my glossing)[148]
a. *da-v-c'er-Ø*
PV-1.S-write-PRS.1/2SG.S
'I will write'
b. *da-Ø-c'er-Ø*
PV-2.S-write-PRS.1/2SG.S
'You (SG) will write'
c. *da-c'er-s*
PV-write-PRS.3SG.S
'He/she will write'

The interaction between person, number, gender and various TAM categories is a well-known phenomenon, but many languages of the Caucasus have dedicated person affixes that only encode verbal arguments. This is particularly true for the Northwest Caucasian languages (Arkadiev & Lander 2020: 407-408), while the situation for the Kartvelian languages is slightly more complex as the personal prefixes are generally stable formwise while the person suffixes vary according to the associated TAM category (Tuite 1998: 73).

(269) Kabardian (Kumaxov 2006: 190) (my glossing)
a. *Ø-s-o-ɦ*
3SG.ABS-1SG.ERG-DYN.PRS-carry
'I am carrying it'
b. *Ø-w-o-ɦ*
3SG.ABS-2SG.ERG-DYN.PRS-carry
'You are carrying it'
c. *Ø-j-e-ɦ*
3SG.ABS-3SG.ERG-DYN.PRS-carry
'He/she is carrying it'

It is also worth mentioning that many Nakh-Dagestanian languages exhibit various forms of person syncretisms, as e.g. Lak has identical first and second person affixes both in singular and plural (Ganenkov & Maisak 2020: 123) and Ingush indicates both the first and second person plural with the prefix *d-* (Nichols

[148] I have not glossed the implicit zero-marked 3rd person object in these examples, as the same subject-marking prefixes without an implicit object are used for what is known in Georgian as medio-passive verbs such as *t'irili* 'to cry', thus *v-t'iri* 'I cry' (Makharoblidze 2012: 83).

2011: 143-144).[149] It appears that only Udi distinguishes all three persons in singular and plural (Harris 2002: 27-28; Maisak 2018: 125).

(270) Udi (Harris 2002: 29-30, 91) (original glossing)

a. *eɣel* *nut'* *šam-k'-al-zu*
sheep.ABS NEG slaughter-LV-FUT-1SG
'I will not slaughter a sheep'

b. *mähl-in-a* *xod* *nut'* *boš-t'-al-le*
yard-OBL-DAT tree.ABS NEG bury-LV-FUT-3SG
'She will not plant a tree in the yard'

c. *xinär-i* *aš-l-ax* *b-e-q'un*
girl-GEN work-OBL-DAT do-AOR[150]-3PL
'They did the girl's work'

The Indo-European languages of the Caucasus tend to differentiate all three persons in singular and plural, while the Turkic languages typically mark all persons except the third person singular, which in certain contexts is indicated by the suffix *–dI(r)* (Ragagnin 2022: 252; Berta & Csató 2022: 328; Karakoç 2022: 363).

[149] The combined first and second plural forms in Ingush, Chechen and some Dargic languages occupy the same slot as the noun class prefixes, which

[150] Harris glosses it as AORII.

6.7.2. Person and number-marking of objects

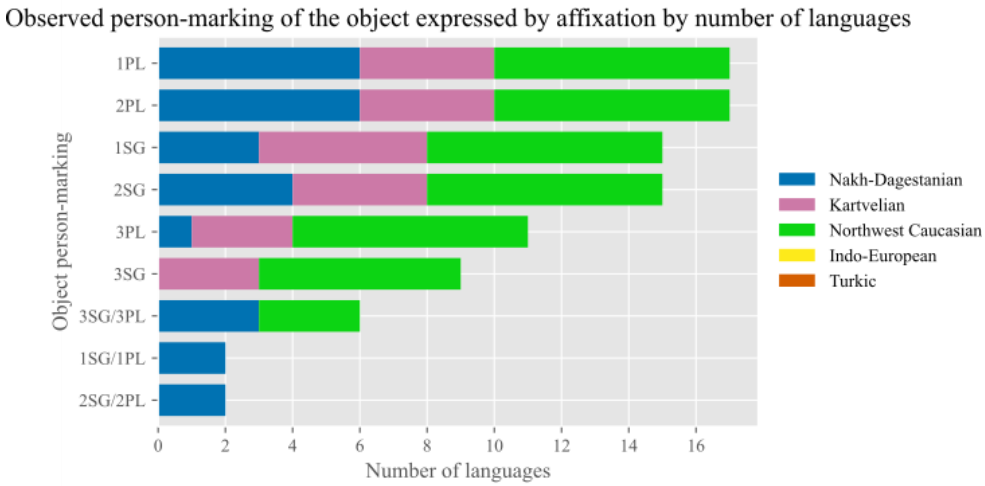


Figure 6.6: Person-marking of direct objects expressed through affixation in the data according to number of languages.

All Kartvelian languages indicate the direct object and indirect object arguments through affixation (Tuite 1998: 21; Boeder 2005: 25), while the prefix slot can only contain either a subject prefix or an object prefix.

(271) Georgian (Boeder 2005: 28) (reglossed)

m-i-nd-a	*g-a-k'oc-o,*	*m-a-k'oc-o*
1SG.O-OV-want-PRS.3SG	2SG.O-NV-kiss-OPT	1SG.O-NV-kiss-OPT

'I want to kiss you, [and] you to kiss me'

All Northwest Caucasian also have prefixes indicating the person of the direct object, i.e. the transitive absolutive, and the indirect object, which are therefore glossed as absolutive and oblique by e.g. Arkadiev and Lander (2020: 407-40). The glossing below follows Kumakhov and Vamling (2009) instead, in order to demonstrate the marking of subject, object and indirect object in Kabardian.

(272) Kabardian (Kumaxov 2006: 197) (my glossing)

a. *w-aj-s-t-a-ɕ*
2SG.O-3PL.IO-1SG.S-give-PST-IND
'I gave you to them'

b. *s-je-p-t-a-ɕ*
1SG.O-3SG.IO-2SG.S-give-PST-IND
'you gave me to him'

c. *fə-zə-j-t-a-ɕ*
2PL.O-1SG.IO-3SG.S-give-PST-IND
'he/she gave you to me'

Verbal affixes marking the person of the direct object, i.e. the transitive absolutive, are also found in a handful of Nakh-Dagestanian languages, e.g. Bats (Hauk 2020: 52), Itsari Dargwa (Sumbatova & Mutalov 2003: 99) and Tabasaran (Babaliyeva 2013: 199-200). The Dargic languages have particularly intricate patterns of affixes indicating both the agent and the patient of the verb (Sumbatova 2020: 174-177). Although Nakh-Dagestanian languages have ergative case alignment, the person agreement in many of these languages is rather accusative, e.g. as in Tabasaran (Harris 2002: 178) and Mehweb (Daniel 2019: 75).

(273) Tabasaran (Babaliyeva 2013: 209) (reglossed)

kːun-du-zuz	*uvu,*	*ʁaχ-ur-za-vu*
love-PRS-1SG.DAT	2SG	marry-FUT/GNOM-1SG.S-2SG.O

'I love you, I will marry you'

In Kubachi, the transitive verb agrees with the noun class but not the person of both the ergative and the absolutive (Magometov 1963: 151; Vamling & Tchantouria 1991: 225-230), as the prefix agrees with the absolutive and the suffix agrees with either the ergative or the dative case, cf. example (274).

(274) Kubachi (Vamling & Tchantouria 1991: 225) (gloss adapted from source)

a.

abadil	*w-alχːun-ni-sa-j*	*gal*
mother(II).ERG	I-feed-PTCP-AUX.PRS-II	son(I).ABS

'(The) mother feeds (her) son'

Kubachi (Magometov 1963: 263) (my glossing)

b.

gal-ij	*jūsːe*	*j-ikːul-sa-w*
boy(I)-DAT	girl(II).ABS	II-love-PRS-I

'The boy loves the girl'

6.8. Valency-changing functions

This section covers a wide range of connected grammatical phenomena that have been variously described as *valency*, *voice*, *diathesis* and *transitivity* (Zúñiga & Kittilä 2019: 3). I have also chosen to add the Caucasiologist concept of *version*, as it covers largely the same functions as applicatives (Arkadiev & Lander 2020: 409; Testelets 2020: 510). Valency is defined as relating to the 'the number of arguments a predicate takes: semantically, syntactically, or morphologically' (Zúñiga & Kittilä 2019: 3). Grammatical voice and diathesis are inherently related as diathesis relates to the 'specific mapping of semantic roles (SRs) onto grammatical roles (GRs)', whereas voice is 'a grammatical category whose values correspond to particular diatheses marked on the form of predicates' (Zúñiga & Kittilä 2019: 4). Transitivity is a less sharply defined concept as it can be defined as 'multi-parameter notion that comprises different facets of clauses, including semantic and syntactic valency, but also agentivity, affectedness, and referentiality of different arguments' and it is not always distinguished from valency (Zúñiga & Kittilä 2019: 3). Since this section covers such heterogeneous concepts and phenomena, the descriptions and analyses are tentative and admittedly unsatisfactory, particularly with regards to the concept of version or applicatives. Zúñiga and Kittilä (2019) make a distinction between processes that 'change semantic valency', e.g. causatives and applicatives, and those that 'change syntactic valency', e.g. passives and antipassives.

6.8.1. Transitivity-changing functions

Transitivity has been described as 'the most basic distinction in valency classes' and the distinction between transitive and intransitive clauses appears to be universally valid (Malchukov 2015: 76). The two universal clause types are *intransitive clauses*, i.e. clauses 'with an intransitive predicate and a single core argument which is in S (intransitive subject) function', and *transitive clauses*, i.e. clauses 'with a transitive predicate and two core arguments which are in A (transitive subject) and O (transitive object) functions' (Dixon & Aikhenvald 2000: 2). Although transitivity is a potentially universal valency class, it is not always morphologically marked. The Kartvelian languages distinguish transitive from intransitive forms both by stem vowel alternation, i.e. ablaut, and in certain contexts by adding intransitive suffixes (Harris 1991a: 18; Tuite 1997: 11).

Table 6.3: Comparison of transitive and intransitive forms of the verb *-t'Vx-* 'return' in Svan (Tuite 1997: 12-13).

Svan (Lower Bal dialect)	**Transitive**	**Intransitive**
Aorist (2SG)	*a-t'ɔx*	*a-t'ex*
Imperfect (2SG)	*t'ix*	*t'ex-en-(w)*

Transitivity is also explicitly encoded by means of affixation in some Nakh-Dagestanian languages and in the Iranian languages of the Caucasus (Abaev 1964: 51; Schulze 2000: 24; Authier 2012: 156). Iron Ossetic has previously been described as distinguishing transitive and intransitive verbs with different past tense suffixes, but Erschler claims that this is simply a tendency (Erschler 2020: 658).

6.8.2. Causative and anticausative functions

Causatives constitute a type of valency-increasing verb forms that introduce an agent into the argument structure (Zúñiga & Kittilä 2019: 15), and they typically indicate that an event is caused by someone or something. Causatives are common across the Caucasus, and causative affixes are found in all Kartvelian, Northwest Caucasian, and Turkic languages of the Caucasus (Boeder 2005: 47; Arkadiev & Lander 2020: 409; Ragagnin 2022: 250; Berta & Csató 2022: 329).

(275) Old Georgian (Harris 1985: 220) (reglossed)
taq'uanis-v-a-cem-in-e *mas* *q'ovel-n-i*
worship-1SG-NV-give-CAUS-AOR 3SG.DAT all-PL-NOM
k'ac-n-i
man-PL-NOM
'I shall cause all men to worship him'

(276) Abkhaz (Chirikba 2003a: 22) (gloss adapted from source)
jə-d-sə-r-c'a-wá-jt'
3SG.NH.ABS-3PL.OBL-1SG.ERG-CAUS-learn-DYN.PRS-FIN
'I teach them (lit. I make them learn it)'

(277) Kumyk (Abdullaeva et al. 2014: 291) (my glossing)
ol *ulan-ïn-a* *каʁïz* *jaz-dïr-ʁan*
3SG son-3SG.POSS-DAT letter write-CAUS-PST.PFV
'He made someone write a letter to his son'[151]

Causative affixes are also found in many but not all Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 120). There are even double causatives in languages such as Chechen (Komen, Molochieva & Nichols 2020: 342), which is also true for e.g. Kabardian (Kumakhov & Vamling 2009: 35-36).

[151] Glossed and transliterated from *Ол уланына кагьыз яздыргъан*, and translated from Russian 'Он заставил (кого-то) написать письмо своему сыну'.

(278) Akhvakh (Magomedbekova 1967: 102) (my glossing)
de-de *wacː-o-de* *di-be-da*
1SG-ERG brother-OBL-ERG 1SG-GEN-REFL
qχuri *b-etɬ'-ā-de*
field(III) III-plough-CAUS-PST.1SG
'I made my brother plough my field'[152]

(279) Chechen (Komen, Molochieva & Nichols 2020: 342) (original glossing)
naana-s *beer-ana* *ħöga* *šura* *mala-j.a-it-u*
mother-ERG child-DAT 2SG.ALL milk drink-J.CAUS-CAUS-PRS
'The mother lets you let/make/have the child drink milk'

Anticausatives are valency-reducing verb forms and are therefore the opposite of causatives, as an anticausative 'removes an agent from the verbal semantics' (Zúñiga & Kittilä 2019: 41). Anticausative affixes are generally not described as such in the Caucasus, but Eastern Armenian has been described as having both causative and possibly anticausative suffixes (Dum-Tragut 2009: 199). However, the presence of anticausatives in Eastern Armenian is worth questioning, as the suffix *–v* also indicates passive and reflexive constructions (Dum-Tragut 2009: 199).

(280) Eastern Armenian (Dum-Tragut 2009: 345) (gloss adapted from source)
šokolad-ə *hešt* *ē* *hal-v-um*
chocolate.NOM-DEF easily be.PRS.3SG melt-ANTIC-PTCP.PRS
'Chocolate melts easily'

Pazar Laz has also been described as having the anticausative prefix *i-* but it is also used in impersonal passive constructions (Öztürk & Pöchtrager 2011: 58-59). The suffix *–nad* in Tsez has also been described as expressing an anticausative function, although it is glossed as *detransitive* (Comrie 2000: 366-367), cf. example (281).

(281) Tsez (Comrie 2000: 367) (reglossed)
patʼi *ker-ā* *y-esa-nay-xo*
Fatima.ABS river-IN II-wash-DETR-PRS
'Fatima is washing in the river'

[152] Translated from Russian '[Я] брата заставил свое поле вспахать'.

6.8.3. Voice functions

Voice is generally not an important grammatical category in the languages of the Caucasus, and it is only truly relevant for the Turkic (Ragagnin 2022: 250; Berta & Csató 2022: 329) and the Kartvelian languages (Tuite 1998), where all languages have affixes indicating passive voice. The Kartvelian passives are somewhat unusual as they can often be expressed by either prefixes or suffixes (Tuite 1998: 92; Boeder 2005: 39; Bolkvadze & Kiziria 2023: 171-173), or both as the passive circumfix in Laz (Öztürk & Pöchtrager 2011: 59), while these passive prefixes are sometimes reanalysed as 'detransitivising' 'decausative' affixes (Testelets 2020: 511; Rostovtsev-Popiel 2020: 560).

(282) Svan (Schmidt 1991: 511) (my glossing)
keː-χ-ten-aːn-da *bobš*
PV-3SG.O-born-PASS-IPFV.PST child.NOM
'She bore a child (lit. a child was born to her)'

The Turkic passive suffix *–Il* and the reflexive suffix *–In* are often presented as simply passive (Širaliev 1971: 106; Ragagnin 2022: 250) or 'as two varieties of the same suffix' (Seegmiller 1996: 24), while the following examples from Karachay-Balkar indicate that they should be differentiated, cf. (283).

(283) Karachay-Balkar (Aliev 1973: 121) (my glossing)
a. *kölek* *džuw-ul-a-dï*
shirt wash-PASS-PRS-3SG
'The shirt is being washed'
b. *xasan* *džuw-un-a-dï*
Xasan wash-REFL-PRS-3SG
'Xasan is washing himself'[153]

Eastern Armenian and Classical Armenian also have a voice distinction (Meillet 1936: 107; Dum-Tragut 2009: 199), which is somewhat complicated by the presence of the medio-passive voice in Classical Armenian (Meillet 1936: 125; Van Damme 2004: 71).

(284) Classical Armenian (Meillet 1936: 96) (my glossing)
or *kočʿ-ecʿ-eal* *ēr* *i* *hreštak-ē-n*
who call-PASS-PTCP be.IPF.3SG by angel-ABL-DEF
'[Jesus], who was (so) named by the angel'[154]

[153] Both glossed and transliterated from *Кёлек джууулады* and *Хасан джууунады*, and translated from Russian 'Рубашка стирается' and 'Хасан купается'.

[154] My translation, from Luke 2:21.

At least two Nakh-Dagestanian languages have been described as having a voice distinction, Udi (Schulze 1982: 178) and Bats (Holisky & Gagua 1994: 197), although the passive suffix in Udi is not productive (Harris 2002: 3) and Hauk describes Bats as 'lacking a more traditional passive' (Hauk 2020: 253). While antipassive constructions have been described in e.g. the Dargic languages (Sumbatova 2020: 188), I have only found antipassive affixes in three Tsezic languages, Hinuq (Forker 2013: 331), Bezhta (Kibrik & Testelets 2004: 273; Comrie, Xalilov & Xalilova 2015: 321) and Hunzib (Van den Berg 1995: 110).

(285) Hunzib (Van den Berg 1995: 110) (gloss adapted from source)
kid *q'uti-lɑː* *j-ũcu-laː-r*
girl(II) trunk-OBL.DAT II-hide-ANTIP-AOR
'The girl hid in the trunk'

6.8.4. Reflexive and reciprocal functions

Reflexivity and reciprocity are two closely related grammatical phenomena that have been described as 'valency-reducing' (Kulikov 2011: 384-385; Dixon & Aikhenvald 2000: 7) or 'argument-identifying' (Malchukov 2015: 101). Reflexives and reciprocals are sometimes subsumed under the wider category of grammatical voice (Zúñiga & Kittilä 2019: 151), but I have decided to treat them as a separate valency-related category as they both encode coreference. The 'canonical' or 'direct' reflexive typically encode events where both the agent and the patient correspond to one referent (Zúñiga & Kittilä 2019: 153) or 'where the Subject is co-referential with the Direct Object' (Kulikov 2011: 384). The reciprocal instead 'consists in several referents simultaneously corresponding to both the A and the P' (Zúñiga & Kittilä 2019: 153), and reciprocals typically correspond to constructions with 'each other' in English. Reflexive and reciprocal verbal affixes are found in all Northwest Caucasian languages (Arkadiev & Lander 2020: 407-408) and all Turkic languages of the Caucasus (Ragagnin 2022: 250; Berta & Csató 2022: 329; Karakoç 2022: 360). The Turkic reciprocal forms are referred to as 'cooperative-reciprocal' (Johanson 2022a: 48), as they also convey a comitative-like function, cf. the comitative applicative/version in section 6.8.5.

(286) Adyghe (Kumakhov & Vamling 2009: 89, 91) (reglossed)
c. *a-r* *Ø-ze-płə-žə-ʁ*
3SG-ABS 3.ABS-REFL-look-REP-PFV
'He looked at himself'
d. *a-xe-r* *Ø-zəre-łəʁʷə-ʁə-x*
3SG-PL-ABS 3.ABS-REC-see-PFV-PL
'They saw each other'

(287) Kumyk (Abdullaeva et al. 2014: 289, 373) (my glossing)

a. *song jül-ün-üp, žuw-un-up, gij-in-ip,*
after shave-REFL-CVB wash-REFL-CVB dress-REFL-CVB
čïq-ma aj-la-n-dï-kïz
go.out-INF go.round-PASS-REFL-PST-2PL
'After having shaved, washed and dressed yourselves, you got ready to go out'[155]

b. *olar birbiri-ne kömek-le-š-e*
3PL each.other-DAT help-PASS-REC-PRS
'They help each other (lit. they are helped by each other)'[156]

Reflexive affixes are also found in Eastern Armenian (Dum-Tragut 2009: 199), Juhuri (Authier 2012: 156). The Eastern Armenian reflexive suffix *–v* is identical to the passive suffix, which also means that reflexive and passive forms can only be distinguished by the syntax and the presence of explicit agent marking, e.g. the postposition *коʁmicʰ* (Dum-Tragut 2009: 356), cf. (288).

(288) Eastern Armenian (Dum-Tragut 2009: 356) (reglossed, retransliterated)

a. *hayr-s sapʰričʰ-i mot sapʰr-v-ecʰ*
father-1SG.POSS barber-DAT at shave-REFL-AOR.3.SG
'My father had (himself) shaved at the barber's'

b. *hayr-s sapʰr-v-ecʰ sapʰričʰ-i koʁm-icʰ*
father-1SG.POSS shave-PASS-AOR.3SG barber-DAT side-ABL
'My father was shaved by (lit. from the side of) the barber'

Reflexive affixes appear to be completely absent in Nakh-Dagestanian languages, while reciprocal prefixes or proclitics have been described in Chechen and Ingush (Čokaev 1970: 140; Nichols 2011: 252).

(289) Ingush (Nichols 2011: 252) (original glossing)
qoana suoca vwaashagh-qieta jish
tomorrow 1SG.INS REC-meet.INF possibility
xugjii hwa?
be.FUT.J=Q 2SG.GEN
'Can we meet tomorrow?/Can you meet with me tomorrow?'

[155] Glossed and transliterated from *Сонг юлюнюп, жувунуп, гийинип, чыкъма айландыгъыз*, and translated from Russian 'Потом, побрившись, умывшись, одевшись, вы приготовились выйти'.

[156] Glossed and transliterated from *Олар бир-бирине кёмеклеше*, and translated from Russian 'Они помогают друг другу'.

The situation in the Kartvelian languages is complicated by the presence of the subjective version described in section 6.8.5 below, which could be analysed as a 'reflexive applicative' (Tuite 2024). I have decided to analyse them as reflexive, as although the subjective version is defines as indicating that the indirect object rather than the direct object 'is coreferential with the subject' (Harris 1991a: 46), the subjective version clearly has a reflexive valency-reducing function. Fähnrich furthermore gives examples from Old Georgian where the prefix *i-* is used in reflexive constructions with reference to the direct object (Fähnrich 1994: 205).

6.8.5. Version or applicative functions

The concept of version is a term that has mainly been used for applicatives in Kartvelian (Tschenkéli 1958: 243; Boeder 2005: 34; Tuite 2024) and the Northwest Caucasian languages (Rogava & Keraševa 1966: 261; Colarusso 1992: 96-97; Arkadiev, Lander & Bagirkova 2024). The nature of the information that is encoded by version is quite different in the two language families, as Kartvelian-type versions can be easily distinguished from the Northwest Caucasian-type versions. Four versions are found in most Kartvelian languages, the subjective, objective, indirect objective and neutral versions (Boeder 2005: 34-37).

(290) Georgian (Boeder 2005: 36) (my glossing)
a. *v-u-k'rep* *vašl-s*
1.S-OV-pick.PRS apple-DAT/ACC
'I pick an apple for him'
b. *v-i-k'rep* *vašl-s*
1.S-SV-pick.PRS apple-DAT/ACC
'I pick an apple for myself'

The subjective version indicates that the subject is carrying out an action to or for itself (Hewitt 1995: 170), while the objective version indicates that the action affects or is carried out for an object (Hewitt 1995: 177). The indirect objective version is somewhat less common, and it is used in passive constructions to indicate when an action involving a direct object is carried out for an indirect object (Hewitt 1995: 204). The neutral version is used in some instances to indicate transitive meanings without specifying the beneficiary (Hewitt 1995: 170). There is also a locative version in Kartvelian, which Tschenkéli refers to as the 'superessive' version (Tschenkéli 1958: 243), and it indicates that the event is carried out on or in an object (Hewitt 1995: 184). The locative version is identical to the neutral version in Georgian and Megrelian (Rostovtsev-Popiel 2020; Bolkvadze & Kiziria 2023: 182), while Svan and Laz make a formal distinction between the neutral and locative versions (Tuite 1997: 26; Öztürk & Pöchtrager 2011: 55).

As Tuite (2024) compares the Kartvelian version system to applicatives in a wider linguistic framework, he finds that only the objective and locative (or 'superessive')

versions have clear parallels in the benefactive and locative applicatives (Tuite 2024: 917-924). I have therefore encoded the objective version and the benefactive applicative as a fused 'benefactive-objective' since they have the same grammatical function. The subjective version is particularly interesting as it can be analysed as a 'reflexive benefactive' or a 'reflexive applicative', since the implicit indirect object 'is coreferential with the subject' (Harris 1991a: 46; Tuite 2024: 934). This demonstrates how difficult it is to separate different categories of valency-changing functions, as they often overlap.

(291) Svan (Tuite 1997: 26) (original glossing)

a. *dina* *qæn-s* *æ-b-em*
girl.NOM bull-DAT NV-tie-SM
'The girl ties up the bull'

b. *dina* *qæn-s* *i-b-em*
girl.NOM bull-DAT SV-tie-SM
'The girl ties up her own bull'/'The girl ties up the bull for herself'

c. *dina* *mu-s* *qæn-s* *x-o-b-em*
girl.NOM father-DAT bull-DAT 3.O-OV-tie-SM
'The girl ties up her father's bull/'The girl ties up the bull for her father'

d. *dina* *megæm-s* *qæn-s* *x-a-b-em*
girl.NOM tree-DAT bull-DAT 3.O-LOC-tie-SM
'The girl ties the bull to a tree'

The Northwest Caucasian versions are more diverse, but generally share three versions, namely the benefactive, malefactive and comitative versions or applicatives (Arkadiev & Lander 2020: 409). O'Herin labels these constructions as 'postposition incorporation' (O'Herin 2002: 213), which would make them comparable to spatial preverbs (cf. section 6.11.1). The comitative version indicates that an event is carried out 'together' or 'with' an explicit object (Colarusso 1992: 97; Kumakhov & Vamling 2009: 31), and there is even a version in Adyghe and Kabardian that indicates a 'joint action of different subjects' (Kumakhov & Vamling 2009: 31). Arkadiev & Lander also add the instrumental applicative, which is found in Abkhaz (Hewitt 2010: 125), Abaza (O'Herin 2002: 217), Abzakh Adyghe (Konuk 2022: 308) and Shapsug Adyghe (Smeets 1984: 261).

(292) Abzakh Adyghe (Konuk 2022: 260) (reglossed)
hʁləʁʷ-ər *ʃʒʒəjʒ-m* *sə-rʒ-ʔʷə-pʃʒtə-ʁ*
bread-ABS.DEF knife-OBL 1SG.ERG-INS-TERMIN[157]-cut-PST.PFV
'I have cut the bread with the knife'

[157] Labelled as 'locatif au bout' by Konuk, which would literally mean 'I have cut the bread at the end with the knife'.

(293) Abaza (O'Herin 2002: 214-215) (adapted from source gloss)

a. *ahʷa* *r-zə-s-tɛpa-b*
sword 3PL.OBL-BEN-1SG.ERG-make-FUT2
'I will make them a sword.'

b. *j-s-tʃʷə-j-ɣətɛ-d*
3SG.NH.ABS-1SG.OBL-MAL-3SG.M.ERG-steal-DYN
'He stole it from me.'

c. *q'apɮan-j* *asjat-j* *mʕʷa*
Kaplan-ADD Asiat-ADD way
j-hə-c-ʕakʷəl-t'
3PL.ABS-1PL.OBL-COM-set.out-DYN
'Kaplan and Asiat set out on the way with us'

Finally, there is also a locative version/applicative in all Northwest Caucasian languages (Fenwick 2011: 116; Arkadiev & Lander 2020: 390-391; Konuk 2022: 323; Arkadiev, Lander & Bagirkova 2024: 889-891). The locative version/applicative in Kabardian is described as a spatial preverb by Kumaxov (Kumaxov 2006: 262), as the function of locative applicatives and spatial preverbs clearly overlaps. Arkadiev and Lander (2020) also mention the 'involuntative' applicative (Arkadiev & Lander 2020: 409-410), cf. section 6.5.

(294) Kabardian (Kumakhov & Vamling 2009: 91) (reglossed)
we *ɕ'ə-m* *wə-ɕə-ɬ-Ø-ɕ*
2SG ground-OBL 2SG.ABS-LOC-lie-PRS-IND
'You are lying on the ground'

6.9. Non-finite functions

The Participial function

There are three primary categories of non-finite verb forms that are found in the affixal data set, i.e. participles, infinitives and masdars. Participles are found in all five language families, and they are often further divided according to tense as e.g. present, past and future participles. However, the exact definition of what constitutes a participle varies between the families, as participles typically accompany auxiliary verbs to from periphrastic forms in Nakh-Dagestanian (Ganenkov & Maisak 2020: 117), while participles in Northwest Caucasian languages can often convey entire relative clauses (Arkadiev & Lander 2020: 417). The latter usage of the term follows Haspelmath's definition of participles as 'nonfinite verb forms specialised for adnominal subordination' (Haspelmath 1995a: 7). The function of participles is therefore often twofold, and adnominal relative

clauses are also frequently expressed by means of participles in Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 122).

(295) Kabardian (Kumakhov & Vamling 2009: 63-64)
(reglossed, retransliterated)
a. *a-bə* *ž-j-e-ʔe*
3SG-OBL LOC-3SG.ERG-DYN.PRS[158]-say
fe *fə-kʷ'e-w*
2PL 2PL.ABS-go.PRS-PTCP
'He says that you are leaving'
b. *(wə)* *wi-gʷəʁe-ɕ*
2SG 2SG.ERG-think.PRS-IND
(de) *də-kʷ'-a-we*
1PL 1PL.ABS-go-PFV-PTCP
'You think that we (have) left'

(296) Bagvalal (Maisak & Tatevosov 2001b: 296) (gloss adapted from source)
maħammad-i-r *awal* *dž-āla-łō-b*
Mahammad-OBL-ERG house do-POT-PTCP.FUT-NH
b-uk'a-b-o *ekʷ'a*
NH-be-NH-PTCP.PST[159] be.PRS
'Mahammad has been intending to build a house (they say)'[160]

(297) Eastern Armenian (Dum-Tragut 2009: 158) (gloss adapted from source, retransliterated)
inčʰ *kʰuyr-s* *teʁapʰox-ʏ-el* *ē* *Erevan,*
since sister-1SG.POSS move-REFL-PTCP.PFV be.PRS.3SG Yerevan
hačax *karot-um* *em* *nran*
often miss-PTCP.PRS be.PRS.1SG 3SG.DAT
'Since my sister has moved to Yerevan, I often miss her (lit. I am often missing her)'

The concept of *gerunds* is closely connected to participles, but it is problematic as its function is ambiguous and Haspelmath states that the term is 'unsuitable for general use' (Haspelmath 1995a: 45). The term gerund is sometimes applied to present participles, e.g. English, or to adverbial constructions such as the Russian concept of *деепричастие*, i.e. 'forms of the verb that substitute for co-ordinate or

[158] Kumakhov & Vamling segment it as *-je-* 'S3SG', while I follow Paris (1989: 201) and Arkadiev & Lander (2020: 418) in analysing *-e-* as a separate dynamic/present prefix.

[159] Kibrik et al. gloss is as CONV 'converb', but its function is identical to a past participle in this context.

[160] Translated from Russian 'Магомед(, говорят,) собирался строить дом'.

adverbial clauses' (Wade et al. 2020: 385). Since 'gerund' in a Russian-language context is both termwise and functionwise identical to the concept of *converbs* (Weiss 1995: 241), I have decided to instead analyse verb forms described as 'gerunds' as either participles, when used in periphrastic constructions with an auxiliary, or as converbs when used to indicate subordination. Gerunds typically express adverbial functions such as 'while', 'when' and manner in Russian (Wade et al. 2020: 390-391), which corresponds to simultaneous, temporal and manner converbs in many languages of the Caucasus (cf. section 6.10).

The Infinitive function

Infinitives constitute a category of non-finite verb forms that are often poorly defined and the term *infinitive* has been applied to 'rather different sorts of syntactic entities' (Noonan 2007: 67). The function of the infinitive is typically to indicate a 'complement clause with (roughly) irrealis meaning' and a 'purpose clause' (Haspelmath 1995a: 28). Noonan defines the infinitive as 'verb-like entities that do not bear syntactic relations to their notional subjects; i.e. their subjects do not take nominative case marking or condition verb agreement (where otherwise appropriate for subjects), nor are they marked in the genitive case, as a subject of a nominalization might be marked' (Noonan 2007: 67).

Infinitives should be distinguished from converbs (cf. section 6.10), as converbs express coordination and subordination while infinitives indicate complementation. In the Caucasus, infinitive affixes are found in all Turkic (Dehghani 2000: 121; Abdullaeva et al. 2014: 368; Ragagnin 2022: 253; Berta & Csató 2022: 333; Karakoç 2022: 356), all Indo-European languages (Meillet 1936: 127; Schulze 2000: 22; Dum-Tragut 2009: 201; Authier 2012: 233; Erschler 2020: 660; Suleymanov 2020: 152) and all Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 114).

(298) North Azerbaijani (Širaliev 1971: 92) (my glossing)

mən	*heç*	*nə*	*öyrən-mək*	*istə-m-ir-əm*
1SG	not.any	what	learn-INF	want-NEG-PRS-1SG

'I do not want to learn/find out anything'[161]

(299) Northern Talysh (Schulze 2000: 71) (gloss adapted from source)

av	*ayo*	*damänd-e*	*ba*	*do*	*peš-e*
3SG.ABS	there	start.PST-AUX.3SG	to	tree	climb-INF

'There he started to climb on a tree'

(300) Northern Akhvakh (Creissels 2009b: 118) (original glossing)

di-tła	*ʕĩk'o*	*b-ix:-urutła*	*ĩd-itła*
1SG-DAT	hen	NH-catch-INF	be.able-PFV.NEG

'I was not able to catch the hen'

[161] Translated from Russian 'Я ничего не хочу узнавать'.

The Masdar function

Almost all Nakh-Dagestanian languages also have Masdars, which are non-finite verb forms that are similar to verbal or deverbal nouns, and masdars are found in all three endemic language families of the Caucasus (Polinsky 2020: 16; Arkadiev & Lander 2020: 441). The 'non-finite' suffix *-i* in Ubykh is difficult to categorise as it appears to neither be an infinitive nor a masdar (Fenwick 2011: 109). The complementary distribution of masdars and infinitives in language families such as Kartvelian and Northwest Caucasian is problematic from a comparative perspective, as they clearly express similar albeit not identical functions. The relationship between the concepts verbal nouns and masdars is equally troublesome, as the terminology is typically language family dependent. The Kumyk verbal noun suffixes *–mAG* and *–(V)w* (Abdullaeva et al. 2014: 365-366; Berta & Csató 2022: 33) would have been analysed as masdars in a Nakh-Dagestanian context. I have therefore decided to merge masdars and verbal nouns into one joint category, as the terms are often used interchangeably. However, there are rare instances where languages have been described as having both a masdar and a separate verbal noun, e.g. Itsari Dargwa (Sumbatova & Mutalov 2003: 123-125).

(301) Laz (Lacroix 2009: 162) (reglossed)

hak	*skan*	*o-dzir-u*	*şen*	*mo-p-t-i*
here	2SG.GEN	MSD-see-MSD	for	PV-1SG-come-AOR

'I came here to see you (lit. for your seeing)'

(302) Kryts (Authier 2009: 311) (reglossed, retransliterated)

za	*k'iy-в-ar*	*č'in*	*v-ar-idž*
1SG.GEN	heart-SUPER-ELA	dance(F)	F-do-MSD

sa-d-χun-udž
PV-INAN-forget-PRS
'I have forgotten to dance (lit. from on my heart I forget dance doing)'

(303) Adyghe (Rogava & Keraševa 1966: 329) (my glossing)

a-w	*we*	*qə-we-s-ʔʷe-n*
DEM-ADV	2SG	PROX-2SG.ABS-1SG.ERG-say-MSD

s-ɣ'e-rep
1SG.ERG-know-PRS.NEG
'I don't know what to tell you'[162]

[162] Glossed and transliterated from *Ау о къыосIон сшIэрэп*, translated from Russian 'Не знаю, что тебе сказать'.

6.10. Converbs or adverbial subordinators

Converbs are defined by Haspelmath as 'a nonfinite verb form whose main function is to mark adverbial subordination' (Haspelmath 1995a: 3) or 'a verb form which depends syntactically on another verb form, but is not its syntactic actant' by Nedjalkov (1995: 98), thereby usually indicating how a subordinate clause relates to a main clause. The question of whether converbs are non-finite by definition therefore differs between Haspelmath and Nedjalkov, which van der Auwera summarises as a difference between Haspelmath's converb in *sensu stricto* and Nedjalkov's converb in *sensu latiore* (Van der Auwera 1998: 277-280). Since this thesis has a functional approach, I have adopted Nedjalkov's wider definition of converb, thus including finite subordination, as many Northwest Caucasian converb forms are ostensibly finite with TAM, gender and person agreement. The concept of converb in this sense was introduced within Russian linguistics as *деепричастие*, and has been widely used for Nakh-Dagestanian, Turkic, Mongolic and various other Central and Northern Asian language families (Haspelmath 1995a: 3). The equivalent English term *converb* has therefore been applied to a wide range of functions in Nakh-Dagestanian languages (Ganenkov & Maisak 2020: 115) and quite recently also for similar constructions in Northwest Caucasian languages (Chirikba 2003a; Arkadiev & Lander 2020: 417). The number and function of these affixed converbs vary significantly within these families, but some of the most common converbs mirror the meanings of the English conjunctions 'while', 'after', 'before', 'when', 'as soon as', 'if', 'because' and 'although' (König 1995: 64; Forker 2020a: 321). All Northwest Caucasian languages included in this thesis have a substantial inventory of verbal affixes that convey typical converb functions.

(304) Kabardian (Kumakhov & Vamling 2009: 61-63)
(original gloss, retransliterated)

a. [*a-r* *mezə-m* *ɕə-kʷ'e-r*] *we* *w-o-ɕ'e*
he-ABS wood-ERG when-go-ABS you S2SG-DYN-know
'You know [when he goes to the wood]'

b. [*we* *wə-zerə-mə-tə-m]* *wədz* *q'-o-č'e*
2SG S2SG-where-NEG-stand-OERG[163] grass OR[164]-DYN-grow
'The grass grows [where you don't stand]'

c. [*fe* *pismo-r* *dəʁʷase* *f-txə-ʁa-me*]
you letter-ABS yesterday S2PL-write-PLUP-COND
se *a-r* *nobe* *q'ə-s-ʔerəhe-at*
I it-ABS today OR-S1SG-receive-PRF2
'[If you had written the letter yesterday], I would have received it today'

[163] Kumakhov & Vamling's gloss OERG 'oblique ergative'.

[164] Kumakhov & Vamling's gloss OR 'orientation'.

These verbal affixes are somewhat different from conventional converbs though, since they are typically finite with person agreement and occur both before and after the root, which is not the case in Turkic and Nakh-Dagestanian languages. This might suggest that some of these verbal affixes are not true converbs, but I have decided to analyse them as such due to their converb-like functions.

Caucasian converb inventories typically range from 6 to 16 converbs, but it is important to consider that converbs are often inadequately described in older Nakh-Dagestanian and most Northwest Caucasian descriptions, so the actual inventories in some of these languages are possibly larger than suggested in previous literature. Converbs are primarily found in the Nakh-Dagestanian, Northwest Caucasian and Turkic languages of the Caucasus, but converbs or converb-like constructions are also found in Laz (Anderson 1963; Holisky 1991; Lacroix 2009), Juhuri (Authier 2012) and Ossetic (Erschler 2020: 660). Converbs are conventionally divided into *specialised* and *contextual* (or *polyfunctional*) converbs following Nedjalkov (1995), where the specialised converbs can be further subdivided into temporal and non-temporal converbs (Nedjalkov 1995: 107). Specialised converbs are distinguished from contextual converbs as the specialised converbs have 'one or two adverbial meanings' while contextual converbs 'have three or more adverbial meanings' (Nedjalkov 1995: 106), which subsequently raises questions about what constitutes a distinct adverbial meaning. Nedjalkov also distinguishes a third category of narrative converbs that 'express a coordinative connection' (Nedjalkov 1995: 106), but this third category is seldomly distinguished in contemporary literature, which has led me to merge the contextual converbs with the narrative converbs.

The distinction between specialised and contextual converbs is observable and generally valid in Nakh-Dagestanian languages (Creissels 2010: 107), but this distinction is less strict outside of the Caucasus, which makes Bisang suggest that it is rather a continuum from completely contextual or general converbs to the most specialised converbs (Bisang 1995: 156).

6.10.1. Temporal specialised converbs

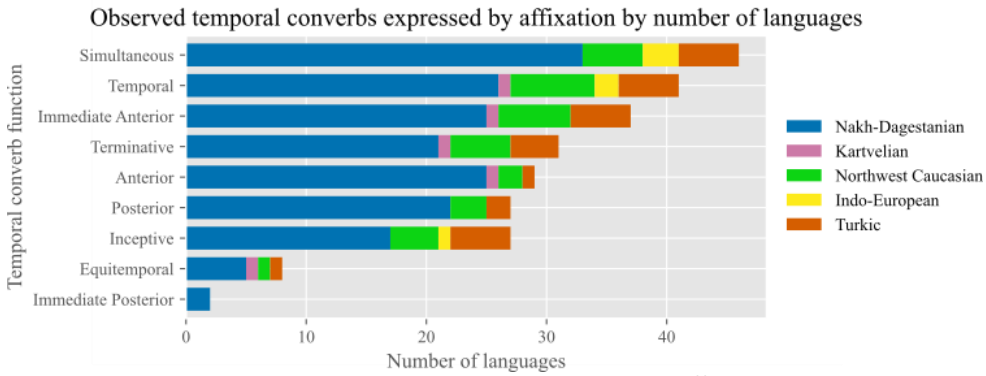


Figure 6.7: The most common temporal converbs expressed through affixation in the data according to number of languages.

The Simultaneous converb function

The Simultaneous converb indicates that the subordinate clause takes place simultaneously with the main clause, and typically corresponds to the conjunction 'while' (Nedjalkov 1995: 107). The simultaneous converb can be difficult to distinguish from gerunds and participles in some languages although the function is similar, as the terminology differs between the descriptive traditions of specific language families (Haspelmath 1995a: 2). This is further complicated by the semantic overlap between the function of the simultaneous converb and the function of the manner converb, although the theoretical difference is that the simultaneous converb optimally refers to two separate actions while the manner converb describes the same singular event as in the main clause (König 1995: 65).

Even though the simultaneous converb is categorised as a specialised temporal converb, the simultaneous converb function is often expressed by using contextual converbs in many languages, which supports Bisang's claim that the distinction between specialised and contextual converbs is rather a continuum, as the simultaneous converb appears to be one of the least specialised converbs.

(305) Tindi (Magomedova 2012: 180) (my glossing)

[*eči-la*	*qχaná:-kʲ'a*]	*w-áχ:a*	*t'a-wo*	*o-w*
apple-SUPER	eat-SIM.CVB	I-out	run-WPST	DEM-I

'While biting into (lit. on) an apple, he ran out'[165]

[165] Glossed and transliterated from *Эчила къанáкI'а вáхъа тIаво ов*, and translated from Russian 'Надкусывая яблоко, он выбежал'.

(306) Kabardian (Kumaxov 2006: 163) (my glossing)
[*ɕ'ale* *c'əkʷ'-əj-t'ə-r* *čaj* *je-fe-xʷ*]
boy small-CRD-three[166]-ABS tea 3SG.OBL-drink-SIM.CVB
ja *ane-r* *šəgʷəxʷə-m* *je-pseɬə-ɬ'-a-ɕ*
their mother-ABS driver-OBL 3SG.OBL-talk-DIR-PFV-IND
'While the (three) boys were drinking tea, their mother bargained with the driver'[167]

(307) North Azerbaijani (Širaliev 1971: 145) (my glossing)
kərim [*mürəbbə* *götür-ə* *götür-ə*] *danış-dı*
Kerim murabba take-SIM.CVB take-SIM.CVB speak-PST
'While helping himself to some murabba (fruit preserve), Kerim began to speak'[168]

(308) Juhuri (Authier 2012: 270) (gloss adapted from source)
şimi [*kekül* *xuruz=e* *timor* *sox-de*]
Shimi crest cockerel=DAT stroke do-SIM.CVB
mu=gu *ez* *zen* *xüşde*
EVT=say ABL wife REFL
'While stroking the crest of the cockerel, Shimi says to his wife…'[169]

The processual participle *–is* in Eastern Armenian can occur in 'temporal non-finite' clauses that are in progress (Dum-Tragut 2009: 205), which means that the suffix *–is* has a function similar to a temporal converb, while the processual aspect of the subordinate clauses and the examples given by Dum-Tragut suggest that its function is closer to a simultaneous converb.

(309) Eastern Armenian (Dum-Tragut 2009: 205) (reglossed)
armen-ə [*erek* *vazel-is*] *vnas-ecʰ* *otkʰ-ə*
Armen-DEF yesterday run-SIM.CVB hurt-AOR.3SG foot-DEF
'Armen hurt his foot when he was running (lit. while running) yesterday'

[166] Kumaxov only translates it as plural, while I can only analyse this suffix as the numeral 'three' in this position.

[167] Glossed and transliterated from *ЩIалэ цIыкIуитIыр шай ефэху, я анэр шыгухум епсэлъыIащ* and translated from Russian 'Пока мальчики пили чай, мать торговалась с ямщиком'.

[168] Translated from Russian 'Керим заговорил, накладывая себе варенья'.

[169] Translated from French 'Shimi, en caressant la crête du coq, dit à sa femme…'.

The Temporal converb function

The Temporal converb indicates that the subordinate clause is temporally connected to the main clause, whereas the precise temporal relation depends on the tense-aspect-mood of the subordinate clause, generally corresponding to the conjunction 'when' (Haspelmath 1995b: 427).

(310) Chechen (Komen, Molochieva & Nichols 2020: 351) (original gloss)
[*muusa c'a-ca=v-eʔa-ča*] *ji-lxi-ra zaaraa*
Musa HOME-NEG=V-come-TEMP.CVB J-cry-WPST Zara
'Zara cried when Musa did not come home'

(311) Kabardian (Kumakhov & Vamling 2009: 61) (reglossed)
[*a-r mezə-m ɕə-kʷ'e-r*] *we w-o-ɕ'e*
3SG-ABS wood-OBL TEMP.CVB-go-ABS 2SG 2SG-DYN-know
'You know when he goes to the wood'

(312) Nogai (Karakoç 2005: 194) (reglossed)
[*men bölme-ge kir-gende*]
1SG room-DAT/ALL enter-TEMP.CVB
ol oltïr-ïp turï e-di
3SG sit-CVB AUX.IPFV be-PST
'When I entered the room, he was already sitting there'[170]

Comrie, Forker and Khalilova (2012: 162-164) seem to merge the temporal converb with the simultaneous converb, which could be questioned from a functional perspective. The Khwarshi suffix *-uq'artł'a* is described as a simultaneous converb, while following Khalilova's previous analysis it becomes apparent that it is a temporal converb if example (313) is considered, as its semantics makes a simultaneous interpretation impossible.

(313) Khwarshi (Khalilova 2009: 88) (original glossing)
[do Ø-uh-uq'artł'a] mižul žib.žibis co-n
1SG.ABS I-die-TEMP(.CVB) 2PL.LAT each.GEN name-ADD
qʷa-yin himon goli
write-PFV.CVB thing be.PRS
'When I die, there is a thing for three of you, with your names written (on it).'

Khwarshi merits further attention, as the 'anterior III converb' in Khalilova's terminology appears to express that the converbal clause is anterior to the main

[170] Translated from German 'Als ich ins Zimmer eintrat, saß er gerade da (hatte sich schon hingesetzt)'.

clause but that both events occur within the same 24 hours (Khalilova 2009: 402). This roughly corresponds to the construction 'on the day when', which potentially constitutes a separate specialised converb, while also having a temporal-like function.

(314) Khwarshi (Khalilova 2009: 402) (gloss adapted from source)
[*žu* *λar* *Ø-ot'uq'-dow-quł*]
that.ABS guest(I) I-come-GNT[171].PTCP-DAY.CVB
nišoho[172] *allahise* *ise* *žik'ol*
night.AD Allah.ERG that.OBL man-LAT
b-ešt'-in *ħono* *eⁿš*
III-send-UWPST three apple(III)
'On the night when that guest came, God sent him three apples.'

Temporal converb constructions are expressed by the suffix *–şi* in Laz, which coincides with the genitive suffix, or with the suffixed postposition *–k'ule*/*–(ş)kule* 'after' (Holisky 1991: 460; Lacroix 2009: 775), which is interesting as this suggests that these forms function as specialised converbs in Laz.

(315) Laz (Lacroix 2009: 208) (reglossed)
[*o-çk'om-u-şi*] *hamu-s* *tsk'ai-ş* *yei-s*
PV-eat-AOR.3SG-TEMP.CVB DEM-DAT water-GEN instead-DAT
şarap'i *ko-gy-u-b-u*
wine AFF-PV-OV-pour-AOR.3SG
'When he had eaten, she poured him wine instead of water'

The Immediate Anterior converb function

The Immediate Anterior converb indicates that the subordinate clause takes place in time immediately before the main clause, typically corresponding to the conjunction 'as soon as' (Haspelmath 1993: 385-386). The immediate anterior converb is surprisingly common across the Caucasus, as affixes expressing an immediate anterior function are found in all branches of Nakh-Dagestanian, Northwest Caucasian and Turkic in the Caucasus.

[171] Khalilova's original gloss for GNT 'general tense'.

[172] As nouns functioning as temporal adverbs appear to come first in Khwarshi phrases, *nišoho* likely belongs to the main clause rather than the subordinate clause, which is relevant for the interpretation of the converb.

(316) Avar (Alekseev et al. 2012: 349) (my glossing)
[*kaʁat* *c'al-un* *b-aqχː-ara-b-go*],
letter(NH) read-CVB NH-finish-PTCP.PST-NH-IMANT.CVB
co-jaw *karim-i-d-e-χun* *aħt'-ana*
one-ABS.M Karim-OBL-SUPER-LAT-DIR shout-AOR
'As soon as they had finished reading the letter, one of them shouted to (lit. towards) Karim'[173]

(317) Lezgian (Haspelmath 1993: 7) (gloss adapted from source)
[*zun* *kʷ'al-äj* *fe-ji-waldi*],
1SG.ABS house-IN.ELA go-PTCP.AOR-IMANT.CVB
ahmed *ata-na*
Ahmed come-AOR
'As soon as I left the house, Ahmed came'

(318) North Azerbaijani (Širaliev 1971: 248) (my glossing)
[*maro-nun* *səs-in-i* *eşid-incə*],
Maro-GEN voice-3SG.POSS-ACC hear-IMANT.CVB
yer-ləri-nə *qayıt-dı-lar*
place-3PL.POSS-ALL/DAT return-PST-3PL
'As soon as they heard Maro's voice, they returned to their places'[174]

The immediate anterior converb in Circassian is worth mentioning as it is formed by the circumfix *zerə-…-əw* in Kabardian (Kumaxov 2006: 475) and *zere-…-ew* in Adyghe (Rogava & Keraševa 1966: 177) and Abzakh Adyghe (Paris 1989: 239).

(319) Kabardian (Kumaxov 2006: 475) (my glossing)
[*wə-zerə-ne-sə-ž-əw*], *q'e-psałe*
2SG-IMANT.CVB-DIST-go.to-REP-IMANT.CVB PROX-speak
'Call (lit. word hither) as soon as you get there!'[175]

Holisky describes the postposition *–steri* in Laz as having an immediate anterior function (Holisky 1991: 460), cf. example (320), which therefore could be an example of a specialised Kartvelian converb. Lacroix does not describe the postposition *ster(i)* as being either suffixed or having an immediate anterior function (Lacroix 2009: 783-784).

[173] Glossed and transliterated from *Кагъат цIалун бахъарабго, цояв Каримидехун ахImIана*, and translated from Russian 'Как только прочитали письмо, один из них крикнул Кариму'.

[174] Translated from Russian 'Как только они услышали голос Маро, вернулись на свои места'.

[175] Glossed and transliterated from *У-зэры-нэсыж-у къэпсалъэ* and translated from Russian 'Позвони, как только доедешь'.

(320) Laz (Holisky 1991: 460) (reglossed)
[*oxordža-na* *i-d-u-steri*],
wife-COMP SV-leave-AOR.3SG-IMANT.CVB
kimoli-k *i-dušun-u*
husband-ERG SV-think-AOR.3SG
'Just as the wife left, the husband thought…'

The Anterior converb function

The Anterior converb indicates that the subordinate clause takes place before the main clause, typically corresponding to the conjunction 'after' (Sheyanova 2019: 238; Forker 2020b: 261). Schulze (1997) describes the 'past genitive' suffix *–na* in Tsakhur, which appears to carry an anterior converb function (Schulze 1997: 32), cf. similar suffixes in Lezgian, Aghul and Budukh. The anterior converb is also expressed by affixation in all branches of Nakh-Dagestanian, while it is generally lacking in the Northwest Caucasian and Turkic languages. Dumézil states that the Ubykh circumfix *d(ʁ)ɜ-…-t'ɨn* had the meaning 'après que' (Dumézil 1931: 94), which would constitute an anterior converb, while both Dumézil and Fenwick translate all examples of this converb with temporal 'when'-constructions.

(321) Tabasaran (Babaliyeva 2013: 276) (gloss adapted from source)
[*dars-ar* *häzur* *ʁ-ap'-q^{h}an*],
lesson-PL prepared AOR-make.PTCP.AOR-ANT.CVB
häšim *ʁär-i-q^{h}-na* *ʁu-š-u*
Hashim dell-OBL-POST-LAT AOR-go-AOR
'After preparing his homework, Hashim went to (lit. behind) the valley'[176]

(322) Tsez (Comrie, Forker & Khalilova 2012: 165) (gloss adapted from source)
[*bitor* *b-ezu-nosi*], *neɬā-z* *haqār* *retɬ*
thither III-look-ANT.CVB 3SG.OBL-GEN mouth.IN.LAT meat
kur-xo *zow-no*
throw-IPFV.CVB be-PFV.CVB
'After looking that way, (it) threw the meat into its mouth'

The converb *–Ip* in Kumyk has been described as conveying an anterior function, whereas *–Ip* should rather be analysed as a non-specialised converb as in the other Turkic languages, since it also conveys simultaneous functions (Abdullaeva et al. 2014: 373). Laz again stands out among the Kartvelian languages as the suffixed postposition *–k'ule/–(ş)kule* 'after' in Laz functions as both an anterior and temporal converb (Anderson 1963: 89; Holisky 1991: 460; Lacroix 2018: 856), which further indicates that these constructions should be analysed as specialised converbs in Laz. The double function of *–k'ule/–(ş)kule* as both temporal and anterior also

[176] Translated from French 'Après avoir préparé les devoirs, Hachim alla au vallon'.

demonstrates that it is, in principle, impossible to differentiate temporal constructions from anterior constructions for perfective past actions, as there is no semantic difference between ‘when I had eaten’ and ‘after I had eaten’.

(323) Laz (Lacroix 2009: 203) (reglossed)
[*epei* *vaxti* *gola-xt-u-şkule*]
enough time PV-pass-AOR.3SG-ANT.CVB
padişai *sotxa* *harbi-şa* *i-gzal-u*
sultan somewhere war-ALL SV-go-AOR.3SG
‘After some time had passed, the sultan went to war somewhere’

The Terminative/Limitative converb function

The Terminative or Limitative converb indicates that subordinate clause terminates the main clause, typically corresponding to the conjunction ‘until’ (Forker 2020b: 261), cf. the similar nominal function of the terminative case (section 5.1.2.2). The described function of the terminative converb varies as it occasionally includes both ‘until’ and ‘as long as’ (Johanson 1995: 319; Forker 2013: 237; Johanson 2022a: 57), but the latter function is expressed by a separate converb in some languages of the Caucasus which suggests that they are two different functions (cf. the equitemporal converb function).

(324) Avar (Alekseev et al. 2012: 354) (my glossing)
[*dun* *χʷ-eze-ʕan*], *di-ca* *ħalt’i* *ha-b-ila*
1SG.ABS die-INF-TERM.CVB 1SG-ERG work do-NH-FUT
‘I will work until I die’[177]

(325) Kumyk (Abdullaeva et al. 2014: 378) (my glossing)
[*mart* *čïq-majlï*], *dert* *čïq-mas*
March go.out-TERM.CVB sorrow go.out-FUT.NEG
‘The anxiety will not pass (lit. leave) until March passes’[178]

(326) Ubykh (Fenwick 2011: 171) (gloss adapted from source)
[*ɐ́-zʷʒ* *Ø-zʒ-wʒ-ni-wːtʷ’-ɐ́j-ʃɐχʲʒ*],
DEF-sky 3SG.ABS-REFL-PV-3SG.ERG-take.out-ITER-TERM.CVB
ɐ́-mkʲʒ-n *ʃî-kʲ’ʒ́-n[ʒ]-ʒːmiːt*
DEF-road-OBL 1PL.ABS-go-PL-FUT-NEG
‘We will not go until the sky clears up’

[177] Glossed and transliterated from *Дун хвезегIан, дица хIалтIи гъабила*, and translated from Russian ‘Пока я не умру, я буду работать’.

[178] Glossed and transliterated from *Март чыкъмайлы, дерт чыкъмас*, and translated from Russian ‘До тех пор, пока март не пройдёт, беспокойство не пройдёт’.

The Posterior converb function

The Posterior converb indicates that the subordinate clause takes place after the main clause, typically corresponding to the conjunction 'before' (Haspelmath 1995b: 428). Posterior converbs are noticeably less common than the anterior converbs, but they are still found in all Nakh-Dagestanian branches with Lak being a possible exception. Posterior converbs are generally not found in the Northwest Caucasian languages, but possible posterior converbs have been described in Abkhaz (Hewitt 2010: 195), Abaza (Lomtatidze & Klychev 1989: 145) and Abzakh Adyghe (Paris 1989: 198, 240).

(327) Chechen (Molochieva 2010: 22) (reglossed)
[*muːsa* *ħa-v-aːl-ie*]
Musa.ABS(V) PROX-V-come-POSTR.CVB
dʕa-j-axa-ra *zaːra*
DIST-J-go-WPST Zara.ABS(J)
'Zara left before Musa came'

(328) Abkhaz (Hewitt 2010: 195) (my glossing)
[*akrə-b-f-aːndza*], *b-napə* *dzʷdzʷa!*
anything-2SG.F-eat-POSTR.CVB 2SG.F.POSS-hand wash.IMP
'Before you eat, wash your hand(s)!'[179]

There are also specialised posterior converbs in Kumyk (Abdullaeva et al. 2014: 376) and North Azerbaijani (Širaliev 1971: 145), while the latter primarily indicates anterior clauses by combining non-specialised converbs and adverbs (Ragagnin 2022: 257).

(329) Kumyk (Abdullaeva et al. 2014: 376) (my glossing)
[*jurt-ʁa* *bar-ʁanča*], *qurdaš-lar* *tüken-le-ge* *qara-ma*
village-ALL go-POSTR.CVB friend-PL shop-PL-DAT look-INF
šaʁar-ʁa *čïq-dï-lar*
city-ALL go.out-PST-3PL
'Before going to the village, the friends went out to the city to go shopping (lit. look at shops)'[180]

[179] Glossed and transliterated from *Акрыбфаанза, бнапы зәзәа!*.

[180] Glossed and transliterated from *Юртгъа баргъанча, къурдашлар тюкенлеге къарама шагъаргъа чыкъдылар*, and translated from Russian 'До отъезда в село, друзья вышли в город, чтобы посмотреть магазины'.

The Inceptive converb function

The Inceptive converb indicates that the action of the main clause was incepted by the time of the action in the subordinate clause, therefore indicating an ongoing action that started at an earlier point in time specified by the subordinate clause, typically corresponding to the conjunction 'since' or 'from the moment…' (Creissels 2010: 126; Sheyanova 2019: 240). Converbs with an inceptive function are generally found in the Nakh-Dagestanian languages and the Turkic languages of the Caucasus, while they are potentially found in Abkhaz (Aristava 1968: 66; Hewitt 1989: 51; Hewitt 2010: 195), Abaza (Genko 1955: 152), Abzakh Adyghe (Paris 1989: 238) and Ubykh (Fenwick 2011: 171).

(330) Standard Dargwa (Abdullaev 1971: 363) (my glossing)
[*nuša kanikul-t-a-s d-ata-ib-la*]
1PL summer.holiday-PL-OBL-DAT 1PL-let.go-AOR-INCEP.CVB
žaga-ti burħ-n-i d-aš-uli sa<r>i
beautiful-PL day-PL-ABS NH.PL-come.IPFV-PTCP be<NH.PL>
'Since we were sent off on summer holiday, beautiful days have been coming'[181]

(331) Tabasaran (Babaliyeva 2013: 276) (gloss adapted from source)
[*gaga ʁa-k'-i-qʰanmina*]
father AOR-die-PTCP.AOR-INCEP.CVB
dada yas-na-ʔ imi
mother mourning-OBL-IN be.still
'The mother is (still) in mourning since the father died'[182]

(332) Karachay-Balkar (Aliev 1973: 331) (my glossing)
[*kün tij-genli*], *üč saʁat bol-a-dï*
sun rise-INCEP.CVB three hour be-PRS-3SG
'It has been three hours since the sun rose'[183]

[181] Glossed and transliterated from *нуша каникултас датаибла жагати бурхIни дашули сари*, and translated from Russian '[С] тех пор как нас отпустили на каникулы, установились хорошие дни'.

[182] Translated from the French translation: "Depuis que le père est mort, la mère est en deuil".

[183] Glossed and transliterated from *Кюн тийгенли, юч сагъат болады*, and translated from Russian 'С тех пор как взошло солнце, прошло три часа'.

(333) Abkhaz (Hewitt 2010: 195) (my glossing)
[*a(b)ra* *ʃʷə-nχ-oz-ižtaj*], *(a)k'ər* *c'-wa-jt'*
here 2PL-live-PST.NFIN-INCEP.CVB long pass-DYN.PRS-FIN
'Quite a time has passed (lit. is passing) since you used to live here'[184]

The converb *–Demun* in Juhuri also seems to express an inceptive converb function as Authier translates it as 'since' (Authier 2012: 273).

(334) Juhuri (Authier 2012: 273) (gloss adapted from source)
divāʕ *odomi* [*ez* *dede* *bi-remun*] *usol=i*
nature man ABL mother be-INCEP.CVB dirty-COP3
'The nature of man is vile from/since birth'[185]

The Equative and 'Equitemporal' converb functions

The Equative and 'Equitemporal' functions are two closely connected converb functions which equate the subordinate clause with the main clause. The equative function quantitatively equates two clauses, typically corresponding to 'as much as'. The 'equitemporal' function indicates that the duration of the action in the main clause is equal to the action in the subordinate clause, thus it typically corresponds to the conjunction 'as long as'. As the latter converbal function appears to lack a term in the literature, I choose to introduce the term *equitemporal* to distinguish this function from other converbs. Although the equitemporal converb has been linked with the terminative converb, it is apparent that the functions are quite different, as the equitemporal converb implies that *the subordinate clause is true as long as the main clause is true* while the terminative converb implies that *the main clause is only true until the subordinate clause is true*.

The exact relationship between the equative and equitemporal converbs is unclear and they should be treated separately. The only potential equative converb affix I have found is the Standard Dargwa clitic *–cad* (Musaev 2002: 81), but it primarily attaches to nouns. Equitemporal converb affixes are often poorly described and potential examples of equitemporal converbs are e.g. *-cadħi* in Standard Dargwa (Abdullaev 1971: 369; Musaev 2002: 136), *-satːi(nna)* in Itsari Dargwa (Sumbatova & Mutalov 2003: 120), *-tːahːan*/*-tːehːen* in Aghul (Magometov 1970: 151). The North Azerbaijani converb *–(y)InǰA* has also been described as having an equitemporal function (Širaliev 1971: 144).

[184] Glossed and transliterated from *А(б)ра шәынхозижьҭеи (а)кыр ԥуеит*.

[185] Translated from the French translation: "La nature de l'homme est vile depuis sa naissance".

(335) Standard Dargwa (Musaev 2002: 136) (my glossing)
[*neš* *ʁaj-r-ik'-u-cadħi*],
mother word-F-speak-PTCP-EQTEMP.CVB
rursi-ra *ʁaj-r-ik'-i*
daughter-ADD word-F-speak-HAB.PST.3SG
'As long as (lit. as much time as) the mother was speaking, the daughter spoke as well'[186]

(336) Itsari Dargwa (Sumbatova & Mutalov 2003: 193) (reglossed)
[*du* *r=išː-ib-li* *r=iɣal-satːinna*],
1SG F-sleep.PFV-PRET-CVB F-stay-EQTEMP.CVB
čaʿχla.marka-l *b=us-ib-li-di*
heavy.rain-ERG NH-rain.IPFV-PRET-CVB-PST
'All the time I was asleep, a thunderstorm was going on'

The Immediate Posterior converb function

The Immediate Posterior or the Imminent converb indicates that the subordinate clause takes place immediately after the main clause, typically corresponding to the conjunction 'just before' (Creissels 2010: 128). The immediate posterior converb is therefore the posterior equivalent of the immediate anterior converb, which is far more common in the languages of the Caucasus. Immediate posterior converb affixes are surprisingly rare as they have potentially only been described in Chechen (Jakovlev 1940: 268) and Akhvakh (Creissels 2010: 128).

(337) Chechen (Jakovlev 1940: 268) (my glossing)
[*ber* *sama-d-al-lalie*], *d-öšu-ra* *as*
child wake.up-D-LV-IMPOSTR.CVB D-read.IPFV-PST 1SG.ERG
'I was reading just before the child woke up'[187]

(338) Northern Akhvakh (Creissels 2010: 128) (gloss adapted from source)
[*išʷada* *guč'-idałːa*], *šĭ-łː-e* *čelada* *lãgi*
shepherd M-arrive-IMPOSTR.CVB bear-NH-ERG another sheep
b-eχ-e *m-āwi*
NH-take-CVB.NH NH-go.UWPST.NH
'Just before the shepherd arrived, the bear took another sheep away'

[186] Glossed and transliterated from *Неш гъайрикIуцадхIи, рурсира гъайрикIи*, and translated from Russian 'Сколько (времени) мать говорила, столько и дочь говорила'.

[187] Glossed and transliterated from *Бер самадаллале доьшура ас*, and translated from Russian 'Перед тем самым моментом, как ребенок просыпался, я читал'.

6.10.2. Non-temporal specialised converbs

The non-temporal specialised converbs are clearly more diverse than the temporal converbs, but the number of non-temporal specialised converbs found across the language families is approximately the same. The most frequent non-temporal converbs are presented in order of occurrence in the languages of the Caucasus below.

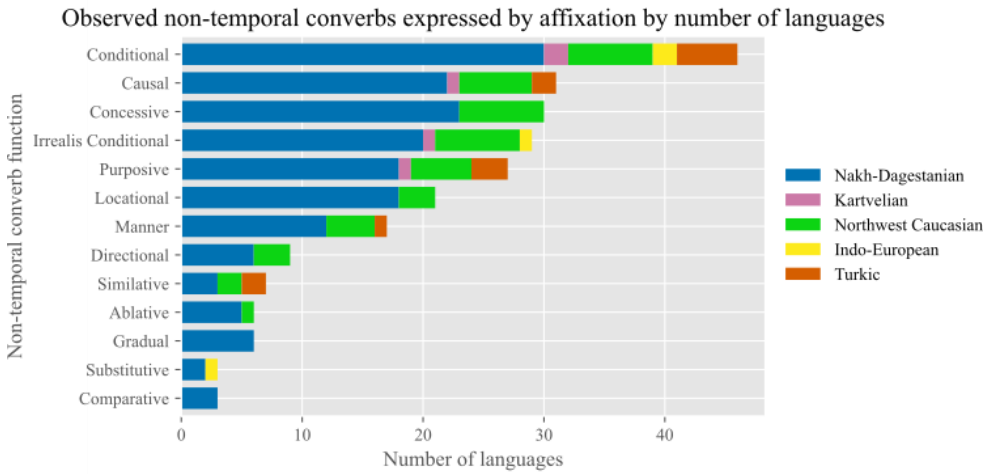


Figure 6.8: The most common non-temporal converbs expressed through affixation in the data according to number of languages.

The Conditional converb function

The Conditional converb indicates a conditional clause, i.e. the *protasis* of a conditional construction (Timberlake 2007: 321), thus typically corresponding to the conjunction 'if' in English (Nedjalkov 1995: 107-108). The conditional converb is the most common non-temporal converb in the Caucasus and it is related to the conditional *mood*, cf. section 6.3, but they have intrinsically different functions, as the conditional mood can constitute an independent main clause whereas the conditional converb always indicates a conditional subordinate clause.

(339) Chechen (Molochieva 2010: 24) (gloss adapted from source)

[*i*	*ħa-v-oːʁ-aħ*]		
3SG.ABS(V)	PROX-V-come-COND.CVB		
vai	*čai*	*mie-r*	*d-u*
1PL.INCL	tea(D)	drink-PTCP.FUT	D-be.PRS

'We would drink tea, if he would come'

(340) Khinalug (Dešeriev 1959: 193) (my glossing)
[*muχtar* *mäktäb-ir* *la-kʷː-to-qʼi*]
Muxtar school-ALL DIST-go.IPFV-PRS-COND.CVB
zɪ-m *hinä-škili* *la-kʷːi-d-mä*
1SG-also 3SG-COM DIST-go.IPFV-I.SG-IND
'If Muxtar goes to school, then I will also go with him'[188]

The conditional 'mood' conventionally described in the Northwest Caucasian languages should rather be analysed as conditional converbs (Arkadiev & Lander 2020: 443), as they encode the protasis and not the apodosis.

(341) Kabardian (Kumakhov & Vamling 2009: 63) (reglossed)
[*fe* *pismo-r* *dəʁʷase* *f-txə-ʁa-me*]
2PL letter-ABS yesterday 2PL.ERG-write-PST.PFV-COND.CVB[189]
se *a-r* *nobe* *qʼə-s-ʔerəhe-at*
1SG 3SG-ABS today PROX-1SG.ERG-receive-PFV
'If you had written the letter yesterday, I would have received it today'

Conditional verb forms in many Nakh-Dagestanian languages are also conventionally described as conditional 'mood', but since they refer in most cases exclusively to the protasis (e.g. Haspelmath 1993: 394; Forker 2020a: 331-334), these forms should be analysed as conditional converbs. The elaborate conditionals in some Dargic languages have both person and tense agreement (Magometov 1963: 205-210; Sumbatova & Mutalov 2003: 102-103; Forker 2020a: 331-334), demonstrating how these conditionals are morphologically mood-like but functionally converb-like.

(342) Sanzhi Dargwa (Forker 2020a: 333) (gloss adapted from source)
[*sːusːul-la* *tʼult'* *b-erkʷ-itːel=ra*],
rye-GEN bread NH-eat.PFV-COND.CVB.2SG=ADD
at *bahlalla* *ʔaˁħ-ce* *ca-b* *žan-ni-j*,
2SG.DAT most.EMPH good-DD[190].SG COP-NH body-OBL-DAT
qʼarqʼala-li-j
body-OBL-DAT
'If (you) would eat bread made of rye, it is the best thing for you, for the body, for the organism'

[188] Glossed and transliterated from *Мухтар маьктаьбир лаккутокъи зэм хинаьшкили лаккуидмаь*, and translated from Russian 'Если Мухтар в школу идёт/пойдёт, то я тоже пойду'. Dešeriev gives the form *хинаъшкили* instead of *хIинаъшкили*, which must be an error, cf. Dešeriev (1959: 126).

[189] Kumakhov and Vamling glosses it as 'S2PL-write-PLUP-COND'.

[190] Forker's original gloss for DD 'definite description'.

Conditional subordinate clauses in Turkic languages are almost uniformly expressed by the suffix –*sA*, which can be followed by possessive person markers and the past tense copula (Johanson 2022a: 42, 57). The conditional suffixes in Turkic are therefore often analysed as conditional mood markers, but if the distinction between conditional converbs and conditional mood discussed above is considered, the suffix –*sA* rather has the function of a conditional converb. This is further supported by the use of possessive personal suffixes rather than the regular person-markers and the presence of other converbal constructions with person-marking in Turkic languages (Johanson 2022a: 42) as well as in Mongolic and Tungusic languages (Nedjalkov 1995: 119-121).

(343) Karachay-Balkar (Aliev 1973: 215) (my glossing)
[*qoj-la* *börü-nü* *kör-se-le*],
sheep-PL wolf-ACC see-COND.CVB-3PL
ajaq-lar-ïn *džer-ge-džer-ge* *qaya-dïla*
foot-PL-3PL.POSS earth-ALL/DAT-earth-ALL/DAT beat-PRS.3PL
'If sheep notice a wolf, they beat their hooves on the ground'[191]

Conditional clauses in Laz are marked with the suffix –*na*, thereby having a function comparable to a conditional converb, while Lacroix describes it as a 'multi-purpose subordinator' as it has multiple other functions as well (Lacroix 2018: 857). A similar situation is found in Megrelian, where the subordinator –*n(i)* marks conditional, temporal, purposive and complement clauses, while the suffix –*da* seem to exclusively mark conditional clauses (Vamling & Tchantouria 1993: 76).

(344) Laz (Lacroix 2009: 180) (reglossed)
[*ha* *dağı-s* *mutu* *ko-b-dzir-na*],
DEM mountain-DAT/LOC anything PV-1SG-see-COND.CVB
o-b-i-bxor-ya
PV-1SG-NV-eat-QUOT
'If I find anything (to eat) on this mountain, I will eat it'[192]

[191] Glossed and transliterated from *Къойла бёрюню кёрселе, аякъларын джерге-джерге къагъадыла*, and translated from Russian 'Если овцы заметят волка, то ногами об землю бьют'.

[192] Translated from French 'Si je trouve quelque chose [à manger] sur cette montagne, je le mange.'

(345) Megrelian (Vamling & Tchantouria 1993: 76) (gloss adapted from source)
[*amser* *tina* *c'eril-s*
tonight 3SG.NOM letter-DAT
do-Ø-č'ar-un-sə-da],
PV-3.S:3.O-write-SM-FUT.3SG-COND.CVB[193]
č'ume *mi-b-ɣ-en-t*
tomorrow PV-1.S:3.O-receive-SM-1PL.S
'If he writes the letter tonight, we will receive it tomorrow'

The Causal converb function

The Causal or Explicative converb indicates that the subordinate clause is the cause of the main clause, typically corresponding to the conjunction 'because' (Haspelmath 1995b: 431).

(346) Tsez (Comrie, Forker & Khalilova 2012: 186) (reglossed)
[*dā-de* *ħuł* *ħon-tłāɣor* *Ø-ik'i-zatł*]
1SG.OBL-APUD yesterday mountain-SUPER.DIR I-go-CAUSL.CVB
obuj-ā *debe-r* *micxer* *netłi-jā?*
father.OBL-ERG 2SG-LAT money give-WPST.Q
'Did father give you money because you went to the mountain with me yesterday?'

(347) Abzakh Adyghe (Konuk 2022: 494) (reglossed)
[*s-ʃhʁ* *wəzə-ʃtə-ʁɜ-təj*]
1SG.POSS-head ache-IFPV-PST-CAUSL.CVB
w*ənɜ-m* *sə-ḱʷ'ɜ-ʒə-ʁa-ʁ*
house-OBL 1SG.ABS-go-REP-PFV-PST
'I had gone home (again), because I had a headache (lit. my head was aching)'[194]

The causal converb should be compared to the causal case (cf. section 5.1.1.2), as the affixes are identical in some languages such as the causal case/converb suffix *–(a)tłeru* in Khwarshi (Khalilova 2009: 73). The causal converbs are found in all branches of Nakh-Dagestanian and potentially also Northwest Caucasian, while in the other families specialised causal converbs are likely only found in Laz (Anderson 1963: 109; Holisky 1991: 460). The situation in Laz is worth further investigation, as both Anderson and Harris describe the suffixed postposition *–šeni* as having a causal/purposive converb function, while Lacroix lists the same

[193] Vamling & Tchantouria glosses it as '3:3:write:FUT-if'

[194] Translated from French 'Du fait que j'avais mal à la tête, j'étais rentrée à la maison'.

postposition together with other suffixed postpositions but he does not describe it as suffixed (Lacroix 2009: 785).

(348) Laz

a. (Holisky 1991: 460) (reglossed)
[*bere* *k'itxeri-na* *t'u-šeni*], …
child studied-COMP be.IPFV.3SG-CAUSL.CVB
'because the child was educated, …'

b. (Lacroix 2009: 785) (reglossed)
[*doxtori-ş* *yer* *mendra-na* *t'u* *şen*]
doctor-GEN place far.away-COMP be.IPFV.3SG for
bere *k'at'a* *ndʁa-s* *var* *a-l-e-t'u-don*
child each day-DAT NEG LV-go-SM-IPFV.3SG-UWPST
'The young man could not go there each day, because the doctor's place was far away'

The Concessive converb function

The Concessive converb indicates that the main clause should be, but is not, contradicted by the subordinate clause, typically corresponding to the conjunction 'although' (Nedjalkov 1995: 107). Concessive converbs can also cover concessive conditional functions, as discussed by Haspelmath & König (1998: 566-568), then typically corresponding to the conjunctions 'even if' and 'even though' in English, e.g. the concessive converb in Hinuq (Forker 2013: 249). The defining difference between the concessive converb, even if it refers to concessive conditionals, and the irrealis/counterfactual converb, is that the main clause or apodosis is typically true in concessive constructions (Haspelmath & König 1998: 563-567), while it is untrue in irrealis/counterfactual conditional constructions.

(349) Northern Akhvakh (Creissels 2010: 133) (gloss adapted from source, retransliterated)
[*w-ošqː-aloʁola*] *čegaza* *χe* *b-otɬː-itɬa*
M-work.CONC.CVB nothing profit NH-occur-PFV.NEG
'Although I worked, I got no result'

(350) Hinuq (Forker 2013: 250) (gloss adapted from source)
[*de* *ɣʷere-z* *daru* *netɬ-ono*]
1SG.ERG cow(III).OBL-DAT medicine give-CONC.CVB
haw *b-uhe-s*
it III-die-PST
'Although I gave the cow medicine, it died'

(351) Ubykh (Fenwick 2011: 186) (gloss adapted from source)
[*ʁɜ-nɜ* *Ø-Ø-ʂiqʷʼɐ-w-nɜːjtʼ-gʲiːlɜn*]
3SG.POSS-mother 3SG.ABS-3SG.OBL-PV-enter-IPFV-CONC.CVB
ɐʁʷɜ *Ø-Ø-ʂiqʷʼɐ-w-ɜːmɨːt*
he.EMPH 3SG.ABS-3SG.OBL-PV-enter-FUT.NEG
'Although his mother was climbing up it, he himself will not'

Concessive converbs are generally found in both Nakh-Dagestanian and Northwest Caucasian languages, but the only description I have found of concessive converbs outside of these families in the Caucasus is the complex clitic *=ge=ş*[195] in Juhuri (Authier 2012: 267), which Authier labels as a 'concessive globale'. This is instead an example of a concessive conditional converb, which should rather be analysed as a sub-category of the irrealis conditional converb.

(352) Juhuri (Authier 2012: 267) (gloss adapted from source)
[*kelebebe=şmu=ş* *ez* *qovre* *växişde* *omo=ge=ş*],
grandfather=2PL=even ABL tomb rise-PTCP come.AOR=if=even
me *ez* *jige=y=me* *ni=jüm-üm*
1SG ABL place=EZ=1SG NEG.EVT=move-1SG
'Even if your grandfather rose from the grave, I would not move from my place'

The Irrealis/Counterfactual Conditional converb functions

The Irrealis or Counterfactual Conditional indicates that the subordinate clause is conditional and not true, typically corresponding to constructions of the type 'if Y would be true (which it is not), X would be true' (König 1995: 64). The irrealis/counterfactual conditional converb differs from concessive converb, which can convey concessive conditionals, as the main clause or apodosis is typically not true in irrealis/counterfactual constructions.

(353) Ingush (Nichols 2011: 305) (gloss adapted from source)
[*Suo-ga* *axcha* *d-alaa-rie*],
1SG.ALL money D-be-IRR.COND.CVB
so *Jivroop-ie* *ghogjar*
1SG Europe-ALL/LOC[196] go.J.COND
'If I had the money I'd go to Europe'

[195] Authier refers to it as a complex suffix, but still glosses it as a combination of the clitics *=ge* and *=(i)ş* (Authier 2012: 267).

[196] Nichols glosses this as ADV 'adverb form', while the suffixes *–ie*/*-a* rather express allative and locative functions in Ingush (Jakovlev 2001: 196-198), which makes Nichols labelling surprising since she also states that its function is to express 'most often location and goal' (Nichols 2011: 427).

(354) Ghodoberi (Kibrik 1996: 209) (gloss adapted from source)

[*min*	*manza*	*ĩh-u*	*bu=k'a-wara*],
2SG.ERG	food	make.PFV-CVB	NH=be.PFV-IRR.COND.CVB
išːe	*makʷatła*	*ba=k'-u-tłi-bu*	*ba=k'-učі*
we.EXCL	hungry	HPL=be-FUT-PTCP	HPL=be.PFV-NEG

'If you had made food, we wouldn't have been so hungry'

The modal suffixes *–je* and *–te* in Adyghe combine with the conditional converb suffix *–me* to indicate 'optative' conditional clauses, which Rogava & Keraševa refers to as desiderative particles, as they express 'desirable unrealised actions' (Rogava & Keraševa 1966: 305). This could also be analysed as a combination of mood and converb, further supporting the notion that converbs are not non-finite by definiton. Example (355) suggests that these constructions should be analysed as irreal converbs, as a desiderative interpretation becomes problematic for negated conditional clauses.

(355) Adyghe (Rogava & Keraševa 1966: 189) (my glossing)

ʂəpqe,	[*te*	*tə-ne-mə-sə-ʁa-ʁe-je.me*],	
true	1PL	1PL-DIST-NEG-reach-PFV-PST-IRR.COND.CVB	
tɛ'ale-r		*xekʷ'ade-ɛtə-ʁe*	
young.man-ABS		perish-FUT/IPFV[197]-PST	

'It is true, if we would not have reached him, the young man would have perished'[198]

Although irrealis conditional converbs are primarily found in Nakh-Dagestanian and Northwest Caucasian languages, the Megrelian and Laz particle *-k'o(n)* appears to indicate irrealis or counterfactual conditional subordinate clauses (Vamling & Tchantouria 1993: 76; Lacroix 2009: 805; Öztürk & Pöchtrager 2011: 81), which suggests that it functions as an irrealis conditional converb.

(356) Laz (Öztürk & Pöchtrager 2011: 81) (reglossed)

[*sk'ani-şe*	*v-ort'-i-k'o*]
2SG-ALL	1.S-be-PST.1SG-IRR.COND.CVB
dersi	*v-i-çaliş-a-Ø-rt'u*
lesson.NOM	1.S-NV-work-SUBJ-PRS.1SG-AUX.PST

'If I were you, I would study'

[197] The complex suffix *-ɛtə-ʁe* is described as a past imperfective (Rogava & Keraševa 1966: 212), while it is formally composed of the future suffix *-ɛtə* and the past suffix *-ʁe*, cf. similar constructions indicating conditional mood in Germanic and Romance languages.

[198] Glossed and transliterated from *Шъыпкъэ, тэ тынэмысыгъагъэ-е-мэ, кIалэр хэкIодэщтыгъэ* and translated from Russian 'Правда, если б мы не добрались до него, то он погиб бы'.

The Purposive converb function

The Purposive converb indicates that the subordinate clause is the purpose of the main clause, typically corresponding to the conjunction 'in order to' (König 1995: 65), as the main clause is carried out with the purpose specified in the subordinate clause. Affixes with a purposive function are found in a majority of the Nakh-Dagestanian languages, in most Northwest Caucasian and in some Turkic languages.

(357) Hinuq (Comrie, Forker & Khalilova 2012: 174) (original glossing)

haw	*ɣodes*	*kekir-ho*	*zoq'e-n*	*čeq-i-do*
3SG	daily	send-PRS	be-UWPST	forest-IN-DIR
[*inaħzek'u*		*r-ut'-ayaz*]		
mushroom(V)		V-gather-PURP(.CVB)		

'(They) send her daily into the forest to gather mushrooms'

(358) Abkhaz (O'Herin 2020: 487) (gloss adapted from source)

[*s-ɥiza*		*di-s-ba-ranɨ*]
1SG.POSS-friend		3SG.H.ABS-1SG.ERG-see-PURP.CVB[199]
a-kalak^{j}	*ax^{j}*	*s-ce-jt'*
DEF-town	to	1SG.ABS-go-AOR.IND.DYN

'I went to town to see my friend'

(359) Karachay-Balkar (Aliev 1973: 221) (my glossing)

al-ɣan-bïz	*onow-nu*	*qol-ubuz-ɣa*
take-PFV-1PL	power-ACC	hand-1PL.POSS-DAT/ALL
[*tïnč*	*džašaw*	*qur-arɣa*]
calm	life	build-PURP.CVB

'We have taken power into our own hands in order to build a calm life'[200]

The Locational/Locative converb function

The Locational, Local or Locative converb indicates the location of the main clause by means of a locational subordinate clause, typically corresponding to the conjunction 'where' (Creissels 2010: 121; Comrie, Forker & Khalilova 2012: 168; Forker 2020b: 261). The locational converb conveys the general location of an action without further specification, which should be compared to the related directional and ablative converbs mentioned below. Locational converbs appear to

[199] O'Herin describes the purposive in Abkhaz as a 'purposive mood' (O'Herin 2020: 487), while the function of the suffixes *-ranɨ*/*-razɨ* indicates that they rather should be analysed as purposive converbs.

[200] Glossed and transliterated from *Алгъанбыз оноуну къолубузгъа тынч джашау къураргъа* and translated from Russian 'Взяли власть в свои руки, чтобы построить спокойную жизнь'.

only be present in the Nakh-Dagestanian and Northwest Caucasian languages in the Caucasus.

(360) Chechen (Molochieva 2010: 26) (gloss adapted from source)
[*šaː voɦ-v-ož-čaɦ*] *b-aiʔi-na k'anta-s kui*
3SG.REFL DE-V-fall-LOC.CVB B-lose-PFV boy-ERG hat.ABS(B)
'The boy lost his hat at the place where he fell down'

(361) Xaidaq (Temirbulatova 2004: 202) (my glossing)
[*insan-t-al-li* *či-a-j-ig-na*]
human-PL-OBL-ERG SUPER-NEG-I.SG-see.IPFV-LOC.CVB
ʔiniq-cːi *diʔaˤn-w-iž-iw* *ca-j*
cave(IV)-INTER hide-I.SG-sit.PFV-PFV COP-I.SG
'He hid in that cave, where people could not see him'[201]

(362) Kabardian (Kumakhov & Vamling 2009: 61) (reglossed, retransliterated)
[*pɛaɛe-r zde-ɛəsə-r*] *ɛ'ale-m jə-łeʁʷ-a-ɛ*
girl-ABS LOC.CVB-sit-ABS boy-OBL 3SG.ERG-see-PFV-IND
'The boy saw where the girl sits'

The Manner converb function

The Manner converb indicates that main clause is carried out in the manner of the subordinate clause, typically corresponding to the conjunction 'as (if)' or the gerund in English and the *gérondif* in French (Authier 2009: 137). Since manner is often expressed by participles and gerunds in many languages, it is generally difficult to make a clear-cut distinction between manner converbs and gerunds. The function of the manner converb is closely related to the similative converb in the same way that the adverbial case is similar but not identical to the equative/similative case (cf. section 5.1.1.2). Manner converbs are found in half a dozen Nakh-Dagestanian languages and in some Northwest Caucasian languages, e.g. Abkhaz (Hewitt 2010: 191), Abaza (Genko 1955: 150; Lomtatidze & Klychev 1989: 142), Abzakh Adyghe (Konuk 2022: 353) and Ubykh (Fenwick 2011: 161).

(363) Chechen (Molochieva 2010: 24) (gloss adapted from source)
[*so ca-gu-čox*] *t'iex j-eli-ra zaːra*
1SG.ABS NEG-see-MNR.CVB by J-go-WPST Zara.ABS
'Zara passed by as if she did not see me'

[201] Glossed and transliterated from *инсанталли чиайигна, гIинихъцци дигIянвижив цай*, and translated from Russian '[С]прятался в таком месте - пещере, где людям не видно'.

(364) Kryts (Authier 2009: 57) (reglossed)
[*ula-ci* *uxvats'-ra*] *zina-z*
eat-SEQ sing-MNR.CVB 1SG.REFL-DAT
ku-kvar-yu-ni-zin
REDUP-walk-PRS.F-PST-1SG
'I was walking by myself, eating and singing'[202]

(365) Ubykh (Vogt 1963: 128; Fenwick 2011: 161) (gloss adapted from Fenwick)
[*ɐ-gʷɜgʷɜ-gʲɨːmsɜ*] *ɐ-kʲʾɜ-n*
3SG.ABS-shuffle-MNR.CVB 3SG.ABS-go-PRS
'He goes shuffling like an old man'

The Directional/Lative converb function

The Directional or Lative converb indicates that the main clause takes place in the direction towards the location of the subordinate clause, typically corresponding to the conjunction 'towards where' or 'whither'. The directional converb is hence functionally related to the locational and ablative converbs. The directional converb is only found in a small number of Nakh-Dagestanian languages, e.g. Chechen (Jakovlev 1940: 252-253), Ingush (Jakovlev 2001: 399), Xaidaq (Temirbulatova 2004: 203), Tsez (Alekseev & Radžabov 2004: 146), Beztha (Comrie, Khalilov & Khalilova 2015: 409) and Karata (Magomedbekova 1971: 174). The directional/lative converb is often segmentable as the locational converb with a directional/lative suffix, e.g. *–ča-ħa* in Ingush and *–guri-r* in Karata.

(366) Ingush (Jakovlev 2001: 399) (my glossing)
[*so* *v-axa-čaħa*], *duqa* *d-ar* *balx-aš*
1SG V-go-DIR.CVB much D-be.PST work-PL
'Where (lit. whither) I went, there was much work'[203]

(367) Karata (Magomedbekova 1971: 175) (my glossing)
[*b-is-am-gurir*] *t'am-a* *ho-b* *hedela*
III-find-PST.III-DIR.CVB throw-IMP DEM-III thing(III).ABS
'Throw that thing towards where (lit. whither) you found it'[204]

The prefixes *aχʲɨndza-* in Abkhaz and *axʲʔa–/ʔaxʲɨ-* in Abaza express the adverbial function 'whither' (Genko 1955: 108; Lomtatidze & Klychev 1989: 138; Hewitt 2010: 190), and could therefore be analysed as directional converbs.

[202] Translated from French 'Je me promenais pour moi, en mangeant et en chantant'.

[203] Glossed and transliterated from *Со вахáчахьа, дукха дар балхаш*, and translated from Russian 'В той стороне/местности, куда я съездил, было много работы.'.

[204] Translated from Russian '[Б]рось эту вещь туда, где нашел'.

(368) Abkhaz (Hewitt 2010: 190) (my glossing)
[*b-aχʲɨndza-co*] *z-dɨr-wajt'*
2F.SG-DIR.CVB-go 1SG.ABS-know-PRS.IND.DYN
'I know where (whither) you are going'[205]

(369) Abaza (Lomtatidze & Klychev 1989: 138) (reglossed)
[*j-axʲʔa-ʕ-ga-z*] *də-ca-t'*
3NH.ABS-DIR.CVB-1PL-take-PST.NFIN 3H.SG.ABS-go-AOR.IND
'(S)he went to where we had taken it/them'

The Similative converb function

The Similative converb indicates that the action of the main clause is carried out in a manner which is similar to the action of the subordinate clause, typically corresponding to the conjunctions 'as', 'like' or 'in the same way as…' (Creissels 2010: 134). Previous literature mentions the Comparative converb as corresponding to the conjunctions 'as if', 'as' or 'like' (König 1995: 72), but this is not an optimal term in a Caucasian context as there are truly comparative converbs in the Caucasus (cf. the comparative converb below).

(370) Northern Akhvakh (Creissels 2010: 134) (original glossing)
[*šuni* *gʷ-ēroqːe*] *žetɬa-la* *gʷij-a!*
yesterday do-SIMIL(.CVB) today-ADD do-IMP
'Do today as you did yesterday!'

(371) Kumyk (Abdullaeva et al. 2014: 379) (my glossing)
xajrulla, [*birew* *ur-ʁandaj*], *ujan-ïp* *get-di*
Xajrulla someone hit-SIMIL.CVB wake.up-CVB go-PST
'Xajrulla suddenly woke up, as if someone had hit him'[206]

The similative converb should potentially be grouped together with the manner converb as both categories encode the manner of the main clause, while the similative converb optimally refers to two different events and the manner converb referes to one single event. These categories do however appear to be in complementary distribution in the Nakh-Dagestanian and Northwest Caucasian language families. Fenwick describes the postposition *gʲɐtɛ'* 'as, like' in Ubykh as indicating 'subordination of manner' (Fenwick 2011: 174), while *-gʲɐtɛ'* seems to rather have a similative function and the converb *-gʲɨmsɜ* appears to have a function which is closer to a typical manner converb (Fenwick 2011: 161), cf. (365).

[205] Glossed and transliterated from *Бахьынзацо здыруеит.*

[206] Glossed and transliterated from *Хайрулла, бирев ургъандай, уянып гетди*, and translated from Russian 'Хайрулла внезапно проснулся, как будто кто-то его ударил'.

(372) Ubykh (Fenwick 2011: 174) (gloss adapted from source)
[*sɨ-nɜ-n*
1SG.POSS-mother-ERG
dɜ-sɨ-Ø-dɨ-ʁˤ-qʼɜ-n-gʲɐtɕʼɨ-n],
SUB-1SG.ABS-3SG.ERG-CAUS-be.born-PST-OBL-SIMIL.CVB-ADV
sɨ-tʼɜtʃʷɜːqʼɜ-ʃ-ɜwːt
1SG.ABS-naked-become-FUT
'I will become as naked as [when] my mother gave birth to me'

The Ablative converb function

The Ablative converb indicates that the location of the subordinate clause is the origin or source of the main clause, typically corresponding to the conjunction 'from where' or 'whence'. The ablative converb is functionally related to the locational and directional converbs, and the directional prefix described above in Abkhaz has an ablative counterpart, i.e. *aχʲɨntʷʼ-* 'from where, whence' (Hewitt 2010: 191).

(373) Abkhaz (Hewitt 2010: 191) (my glossing)
[*b-aχʲɨntʷʼ-aːwa*] *z-dir-wajtʼ*
2F.SG-ABL.CVB-come-PRS.NFIN 1SG.ABS-know-PRS.IND.DYN
'I know whence you are coming'[207]

Ablative converbs are potentially found in a small number of Nakh-Dagestanian languages, i.e. Chechen (Jakovlev 1940: 251), Ingush (Jakovlev 2001: 397), Xaidaq (Temirbulatova 2004: 202) and Tsez (Alekseev & Radžabov 2004: 146). As with the directional/lative converb, the ablative converb is often composed of the locational converb and an ablative suffix, while in Ingush it is even possible to combine the locational converb with both the lative and ablative suffixes, cf. (374)

(374) Ingush (Jakovlev 2001: 399) (my glossing)

a.	[*so*	*v-ena-čara*],	*atːa*	*d-ar*	*balx-aš*
	1SG	V-come-ABL.CVB	easy	D-be.PST	work-PL
	'Where I came from, work was easy'[208]				
b.	[*so*	*v-ena-čaħara*],	*atːa*	*d-ar*	*balx-aš*
	1SG	V-come-DIR.ABL.CVB	easy	D-be.PST	work-PL
	'In the direction of where I came from, work was easy'[209]				

[207] Glossed and transliterated from *Бахьынтəааyа здыруеит*.

[208] Glossed and transliterated from *Со веначара, атта дар балхаш*, and translated from Russian 'Там, откуда я пришел, работа легкая'.

[209] Glossed and transliterated from *Со веначахьара, атта дар балхаш*, and translated from Russian 'В той стороне, откуда я приехал, работа была легкая'.

The Gradual/Graduative converb function

The Gradual or Graduative converb indicates 'a gradual development of the adverbial clause situation that correlates with a development in the superordinate clause situation' (Haspelmath 1993: 387), thus roughly corresponding to the English 'the more …, the more …' (Creissels 2010: 134) or to constructions of the type 'as Y gradually happens, X becomes more true'. Gradual converbs are found in a small number of Nakh-Dagestanian languages, e.g. Lezgian (Haspelmath 1993: 387), Tabasaran (Babaliyeva 2013: 281), Tsakhur (Kibrik & Testelets 1999: 543), Tsez (Comrie, Forker & Khalilova 2012: 160), Northern Akhvakh (Creissels 2010: 134) and Mehweb (Sheyanova 2019: 247).

(375) Lezgian (Haspelmath 1993: 387) (gloss adapted from source)
[*čun* *χür-ü-z* *agaq'-irdawaj*]
we.ABS village-OBL-DAT reach-GRAD.CVB
rik'-i-k *qalabuluχ* *akat-zawa*
heart-OBL-SUB.ESS excitement appear-IPFV
'As we are approaching the village, the heart is getting (more and more) excited.'

(376) Northern Akhvakh (Creissels 2010: 134) (gloss adapted from source)
[*hudu-we* *tɬ'-ũːdałe*] *tał-ari*
DEM-M dance-GRAD.CVB get.tired-PFV
'The more he danced, the more he got tired'

The Substitutive converb function

The Substitutive or Contrastive converb contrasts the action of the subordinate clause with the action of the main clause, typically corresponding to the conjunction 'instead of' in English (König 1995: 64). There are only a few converbs with a substitutive function attested in the data, e.g. Beztha - *ałʁadā(l)* (Comrie, Khalilov & Khalilova 2015: 412), Tabasaran *–ayiz/-äyiz* which also has posterior and terminative functions (Babaliyeva 2013: 282-284), and *–Dei(n)jon*/*–jon* in Juhuri (Authier 2012: 276; Belayev 2020: 610). It is however important to point out that I have not found any examples of substitutive subordinate clauses in any descriptions of Lezgian (Haspelmath 1993; Ganieva 2007; Ganieva 2008; Babaliyeva 2007).

(377) Bezhta (Comrie, Khalilov & Khalilova 2015: 412) (my glossing)
[*qχow-ałʁadā(l)*] / [*čaχ-ałʁadā(l)*]
read-SUBST.CVB / write-SUBST.CVB
'instead of reading' / 'instead of writing'[210]

[210] Glossed, transliterated and translated from *хъов-алъгъа-дā(л)* 'вместо того, чтобы читать' and *чах-алъгъа-дā(л)* 'вместо того, чтобы писать'.

(378) Juhuri (Authier 2012: 276) (adapted from source gloss)
[*e=ijire* *guj,* *tertäʕdi=revoz* *gelin-giyov* *bi-reinjon*]
LOC=such strength haste=INS married be-SUBST.CVB
ä=qärqi *kovre* *deryoh* *domun-de* *xub=i*
LOC=depth blue.ATTR sea remain.in-INF good=COP.3SG
'Instead of getting forcibly and hastily married, it is better to drown (lit. remain) in the depths of the blue sea.'[211]

(379) Tabasaran (Babaliyeva 2013: 189) (original glossing)
[*uvu-z* *ʁ-äyiz*] *keškena* *uzu* *daždi-z*
2SG-DAT come-POSTR(.CVB) PTCL 1SG donkey.OBL-DAT
ʁu-š-niyi-š!
AOR-go-AOR.PST-IF
'If only I had married a donkey instead of marrying you!' [212]

Example (379) is peculiar however, as the substitutive *-äyiz* can just as well be interpreted as a posterior clause (which it is also glossed as), thus yielding the alternative reading 'if only I had married a donkey before marrying you'.

The Comparative converb function

The Comparative converb compares the subordinate clause to the main clause, typically corresponding to the English conjunction 'than', and has a function that is identical to the comparative case (cf. section 5.1.1.2). The term comparative has also been used for converbs corresponding to the English conjunctions 'as if', 'as' and 'like' (König 1995: 72, 76), which rather should be analysed as instances of similative or equative converbs (Creissels 2010: 134; Johanson 2022a: 50). Comparative converb functions have been described in Standard Dargwa (Musaev 2002: 140) and Tsakhur (Kibrik & Testelets 1999: 566-568).

(380) Tsakhur (Kibrik & Testelets 1999: 566-567) (gloss adapted from source)
zuhr-ē *gɨnej-bɨ* *jug=da* *haʔ-a*
Zuhra-ERG bread.IV-PL good=ADV.NHPL NH.PL.make-IPFV
[fāt'imat-ē *t'ele-bɨ* *haʔʷ-ī-le]*
[Fatima-ERG khinkal-PL NH.PL.make.PF-MSD-COMPR.CVB]
'Zuhra is better at making bread than Fatima is at making khinkal'[213]

[211] Translated from French 'Plutôt que d'être marié de force et à toute vitesse comme cela, il vaut mieux rester noyé au fond de la mer'.

[212] Translated from French 'Si seulement j'avais épousé un âne au lieu de me marier avec toi !'.

[213] Translated from Russian ' Зухра делает хлеб лучше, чем Фатима делает хинкалы'.

The comparative converb *–čuol/–čul* described in Chechen (Molochieva 2010: 26; Komen, Molochieva & Nichols 2020: 360) expresses that the subordinate clause is an 'unpreferred alternative' to the main clause, which Molochieva translates as 'rather than', thereby having a slightly different function, cf. (381).

(381) Chechen (Molochieva 2010: 26) (gloss adapted from source)

[*xi*	*mola-čul*]	*čai*	*mie-ra*	*d-u*	*as*
water	drink-COMPR.CVB	tea	drink-FUT	D-be.PRS	1SG.ERG

'I will drink tea rather than water'

6.10.3. Contextual or non-specialised converbs

Contextual converbs or polyfunctional are converbs that 'have three or more adverbial meanings' (Nedjalkov 1995: 106), thus having multiple converbal functions. I would suggest that the concept of contextual converbs should be replaced by the wider concept of *non-specialised converbs* as this also includes Nedjalkov's concept of narrative converbs. Non-specialised converbs would therefore comprise all converbs that are not clearly specialised. I will present the various contextual and non-specialised converbs found in the languages of the Caucasus below.

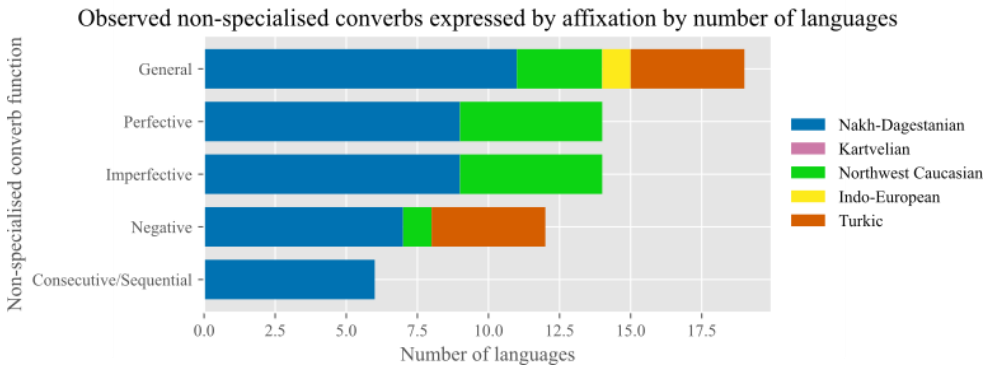


Figure 6.9: The most common contextual converbs expressed through affixation in the data according to number of languages.

The General converb function

The General converb indicates that the clause is subordinate, but it does not entail any specific information regarding the relationship between the subordinate clause and the main clause (Kustova 2019: 255). Nedjalkov's concept of *narrative converbs* appears to be largely identical with the function of the general converb (Nedjalkov 1995: 109). There are numerous examples of converbs that can either be analysed as general converbs or as polyfunctional converbs with separate functions, and the functional approach of this thesis makes Nedjalkov's definition

of contextual converbs as having three or more converbal functions inherently problematic. The Turkic converb suffix *-Ib/-Ip* is a good example, as it covers simultaneous, causal and manner functions in North Azerbaijani (Širaliev 1971: 143) but it also generally connects clauses (Ragagnin 2022: 257), while in Kumyk it covers anterior, simultaneous, causal and manner functions (Abdullaeva et al. 2014: 373-374). I have therefore analysed all *-Ib*/*-Ip* converbs as general converbs, which means that general converbs are found in all Turkic languages of the Caucasus (Ragagnin 2022: 257; Berta & Csató 2022: 332; Karakoç 2022: 363).

The Turkic *–Ib*/*-Ip* converb has previously been analysed as a narrative converb (Nedjalkov 1995: 109), which furthers the notion that the general converb and the narrative converb should be regarded as synonymous concepts. However, the term narrative converb can refer to the perfective converb in languages with a perfective/imperfective converb distinction, e.g. the Tsezic languages (Comrie, Forker & Khalilova 2012: 160). Bezhta lacks an imperfective converb (Comrie, Forker & Khalilova 2012: 167), which means that the 'perfective/narrative' converb in Bezhta should instead be analysed as a general converb. Since the general converb by definition is a non-specialised converb, I have chosen to simply gloss it as CVB below.

(382) North Azerbaijani (Širaliev 1971: 143) (my glossing)

a. [*mahru baş-ın-ı aşağı sal-ıb*] *otur-muş-du*
Mahru head-3SG.POSS-ACC down put-CVB sit-PFV-PST
'Mahru had sat down, lowering his head'[214]

b. [*ay sel-in həyasızlığ-ın-dan*
moon stream-GEN shamelessness-3SG.POSS-ABL
utan-ıb] *gizlə-n-miş-di*
be.ashamed.CVB hide-REFL-PFV-PST
'The moon had hid itself, ashamed by the impudence of the stream'[215]

c. [*kərim [o-nun baş-ın-ı sığalla-jıb*],
Kerim 3SG-GEN head-3SG.POSS-ACC stroke-CVB
üz-ün-dən öp-dü
face-3SG.POSS-ABL kiss-PST
'Kerim, having stroked his head, kissed him on the face'[216]

[214] Translated from Russian 'Махру сидел, опустив голову'.

[215] Translated from Russian 'Луна спряталась, стыдясь наглости потока'.

[216] Translated from Russian 'Керим, погладив его по голове, поцеловал в лицо'.

(383) Iron Ossetic (Bagaev 1965: 357) (my glossing)
bælccon *fæ-cæj-cɨd-is*
traveller PV-IPFV-go.PST-3SG
[*bæχ* *jæ* *fædɨl* *idadz-æj* *las-gæ*]
horse 3SG after rein-ABL pull-CVB
'The traveller walked away, leading a horse after him by the rein'[217]

The Perfective converb function

The Perfective converb is a subtype of the general converb that is only used for subordinate clauses that precede the action in the main clause (Kustova 2019: 256; Forker 2020a: 458), hence it typically encodes anteriority (Forker 2020b: 260). Even though the perfective converb is functionally similar to the anterior converb, they must be distinguished as the perfective converb is not a specialised converb (Kustova 2019: 255). All languages in the data with a perfective converb also have an anterior converb, which further indicates that they are two separate functions. The presence of 'past' and 'present' converbs in Abkhaz and Abaza should also be analysed as instances of perfective and imperfective converbs, as they have an identical function (O'Herin 2020: 471). There are also potential perfective converbs in a few other Northwest Caucasian languages, as the suffix *-jə/-əj* appears to have a past general converbal function (Rogava & Keraševa 1966: 173; Konuk 2022: 359), which might also be true for the Ubykh converbal suffix - *gʲɨ/-jɨ* (Fenwick 2011: 159).

(384) Bezhta (Comrie, Forker & Khalilova 2012: 182) (original glossing)
hogo *y-uɣo-s* [*žamilati* *äč'enayig-na* *ɨtna* *tɬi*
3SG II-die-PRS Zhamilat.ERG ninety-and six year
ömrö-nä *b-oh-na*]
life(III)-and III-do-PFV.CVB
'After Zhamilat lived for 96 years, she dies (lit. She dies, Zhamilat having lived for 96 years)'

(385) Abzakh Adyghe (Konuk 2022: 360) (reglossed)
[ʔɐlɜ-r *qɜwəʃ-əj]*
child-ABS.DEF wake.up-PFV.CVB
jə-ʃxən *jə-ʃxə-ʁ*
3SG.POSS-meal 3SG.ERG-eat-PST
'After waking up (lit. having woken up), the child ate his/her meal'[218]

[217] Glossed and transliterated from *Бælццон фæцæйцыдис бæх йæ фæдыл идадзæй ласгæ*, and translated from Russian 'Пугник шел, ведя за собой коня за повод'. The preverb *fæ-* indicates motion away from the speaker (Thorardson 2009: 68), hence 'walked away'.

[218] Translated from French 'Après s'être réveillé, l'enfant a pris son repas'.

The Imperfective converb function

The Imperfective converb is a subtype of the general converb which is only used for subordinate clauses that are simultaneous with the action in the main clause (Kustova 2019: 256). Imperfective converbs typically convey simultaneity and manner (Forker 2020b: 261), e.g. the imperfective converb in Lezgian (Haspelmath 1993: 380).

(386) Lezgian (Haspelmath 1993: 378) (reglossed)

aburu	[*sad-a=sad-a-w*	*cük^{w}-er*	*wugu-z*]
3PL.ERG	one-OBL=one-OBL-AD	flower-PL	give-IPFV.CVB
[*šad*	*ja-z*]	*sühbetar-zawa-j*	
glad	be-IPFV.CVB	talk-IPFV-PST	

'They were talking gladly (lit. being glad), giving flowers to each other'

The 'present' converbs in Abkhaz, Abaza and Abzakh Adyghe should therefore also be analysed as imperfective converbs (O'Herin 2020: 471; Konuk 2022: 352), cf. the perfective converb above. The Ubykh converbal suffix *-ҁɜ* is also a potential imperfective converb (Fenwick 2011: 160).

(387) Abkhaz (O'Herin 2020: 472) (reglossed)

[*ak'ri-fa-wa*]	*di-t^{w}'a-wp'*
something-eat-IPFV.CVB	3SG.H.ABS-sit-PRS.IND.STAT

'He is sitting eating'

The Negative converb function

The Negative converb indicates that the subordinate clause is negated, typically corresponding to constructions of the type 'without X-ing' (Johanson 2022a: 58). Negative converbal affixes are found in a few Nakh-Dagestanian languages, e.g. Avar (Forker 2020b: 262), Tindi (Magomedova 2012: 151), Rutul (Maxmudova 2001: 160), Itsari Dargwa (Sumbatova & Mutalov 2003: 239) and Khwarshi (Khalilova 2009: 208).

(388) Avar (Charachidzé 1981: 192; Forker 2020b: 262) (gloss adapted from source)

[*ʕemera-b*	*meχ:*	*b-a-č'ogo*],	*padišah*	*χ^{w}-ana*
much-NH	time	NH-go-NEG.CVB	king	die-AOR

'Not much time went by and the king died'

Negative converbs are also found in all Turkic languages of the Caucasus (Širaliev 1971: 144; Dehghani 2000: 126; Abdullaeva et al. 2014: 375; Ragagnin 2022: 254; Berta & Csató 2022: 332; Karakoç 2022: 363).

(389) Kumyk (Abdullaeva et al. 2014: 296) (my glossing)
[*sen gel-e de-p ešit-se-m*],
2SG come-CVB say-CVB hear-COND.CVB-1SG
[*awru-maj*] *öl-er ed-im*
be.ill-NEG.CVB die-CAUS COP-1SG
'If I had heard (it being said) that you were coming, I would die without being ill'[219]

The Consecutive/Sequential converb function

The Consecutive or Sequential converb indicates that one or more clauses are chained together as they typically have the same subject and the tense-aspect-mood is only realised on the last clause of the sequence of joined clauses (Babaliyeva 2013: 314). The consecutive converbs could be analysed as examples of narrative converbs (Nedjalkov 1995: 109). Consecutive converbs have been described in Ingush (Nichols 2011: 294) and the Lezgic languages of Tabasaran (Babaliyeva 2013: 315), Tsakhur (Schulze 1997: 52), Kryts (Authier 2009: 137) and Archi (Chumakina, Bond & Corbett 2016: 39).

(390) Tabasaran (Babalieyva 2013: 315) (gloss adapted from source)
dumu, [*ča-z a-yi buχara bačuk'*
3SG self.OBL-DAT be.in-PTCP Bukhara papakh (wool hat)
ulu<b>k'-nu],
put.on<NH>-SEQ.CVB
[*marcːi c'iyi čuχa qʰa<b>x-nu*],
clean new coat put.on<NH>-SEQ.CVB
[*arsran č'ul ʁidi<b>t'-nu*],
silver belt tie<NH>-SEQ.CVB
[*k'aru čekmyir alax-nu*],
black boot.PL put.on-SEQ.CVB
qʰana sumčri-z ʁu-š-u
again wedding.OBL-DAT AOR-go-AOR
'He put on his own Astrakhan (lit. Bukhara) hat, his clean new coat, his silver belt, his black boots and he went back to the wedding again'[220]

Molochieva describes the converb *–i* in Chechen as having an immediate anterior function, while it can only be used in 'chained clauses and in chained imperative constructions' (Molochieva 2010: 20), cf. the sequential converb *–ie* in Ingush (Nichols 2011: 294).

[219] Glossed and transliterated from *Сен геле деп эшитсем, аврумай оьлер эдим*, and translated from Russian 'Если бы я услышала, что ты приезжаешь, то, не болея, умерла бы'.

[220] Translated from French 'Il mit le chapeau d'astrakan qu'il avait, sa veste neuve et propre, sa ceinture en argent, et des bottes noires et repartit au marriage'.

(391) Chechen (Molochieva 2010: 20) (gloss adapted from source)

[muːsa	*ħa=ʔa-v-oːʁ-i*	
Musa.ABS(V)	PROX=ADD-V-come.IPFV-SEQ.CVB	
niaʔ	*ħa=ʔa-j-oːl-i*	
door.ABS(J)	PROX=ADD-J-open.IPFV-SEQ.CVB	
axča	*ħa=ʔa-oec-i]*	*dʕa-v-oed-u*
money.ABS	PROX=ADD-take-SEQ.CVB	DIST-V-go.IPFV-PRS

'Musa is coming and opening the door and taking the money and going away'

6.11. Spatial preverbs and spatial functions

Preverbs are a category of verbal prefixes that have a wide array of functions, but in the languages of the Caucasus they mainly encode spatial and aspectual information (Arkadiev & Lander 2020: 412; Testelets 2020: 507). They are found in all three endemic language families of the Caucasus (Ganenkov & Maisak 2020: 109; Arkadiev & Lander 2020; Testelets 2020) and in the Iranian languages (Belayev 2020: 606). Although spatial preverbs are the most widespread type of preverbs in the region, it is important to consider that Kartvelian preverbs systematically encode aspect, tense and spatial information simultaneously in a remarkably intricate manner (Boeder 2005: 32-34). The term spatial preverbs is not optimal however, as most Northwest Caucasian languages also encode spatial information by means of suffixes (Aristava 1968: 156; Smeets 1984: 274) and circumfixes (Rogava & Keraševa 1966: 289-291; Paris 1989: 185).

It is meaningful to divide the spatial reference of preverbs into orientation and direction, which is conventionally done for spatial cases (as seen in section 5.1.2), since spatial preverbs often convey functions that are identical to the functions of spatial cases. Spatial preverbs typically encode more specific grammatical functions than spatial cases, especially with regard to spatial orientations, as many Northwest Caucasian languages differentiate more than 20 preverb orientations (Rogava & Keraševa 1966; 112-134; Smeets 1984: 253-261; Kumaxov 2006: 260-268; Hewitt 2010: 114-121).

This study is not exhaustive due to the extremely rich preverb inventories in certain Northwest Caucasian languages as e.g. Abkhaz has been described as having 123 separate preverbs (Spruit 1986: 22-31). Kartvelian preverbs generally express deictic functions while the Megrelian and Laz spatial preverb inventories are almost as complex as the Northwest Caucasian systems (Boeder 2005; Reseck 2015; Öztürk & Pöchtrager 2011). Reseck (2015) also acknowledges the importance of describing spatial preverbs according to orientation and direction (Reseck 2015: 59), but her definition of orientation follows Shanidze as it only covers deictic functions (Reseck 2015: 95-96).

I therefore give a tentative categorisation of only the most common orientations and directions expressed by preverbs in the Caucasus, in analogy with the categorisation used for spatial cases. This is not unproblematic, as preverbs generally convey a wider array of functions than spatial cases, which necessitates the introduction of new orientational and directional categories. It is also worth discussing whether certain preverb orientations should be considered as instances of noun incorporation instead (Chiribka 2003: 43), as many of them are clearly derived from the noun they refer to, e.g. the Kabardian preverbs *ʔeɕʼe-* 'in/into the hands' and *ɕħerə-* 'on/onto the head' where the first components are identical to the nouns *ʔe* 'hand' and *ɕħe* 'head' (Kumaxov 2006: 268, 279). I will present the most common preverb orientations followed by all preverb directions, followed by discussion of the remaining preverb functions such as the Kartvelian aspectual preverbs.

6.11.1. Orientational spatial functions

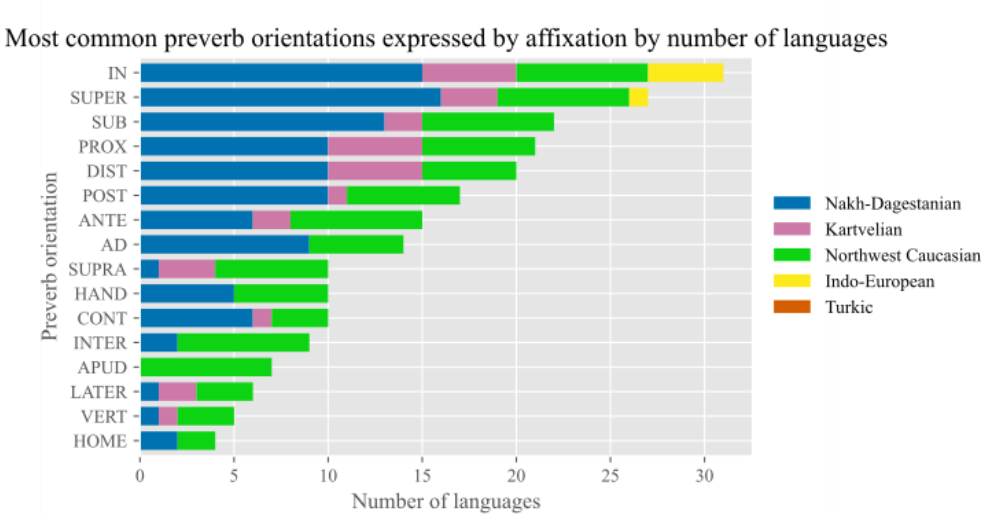


Figure 6.10: Non-exhaustive list of the most common preverb orientations expressed through affixation in the data according to number of languages.

The preverb orientation IN

The preverb orientation IN indicates that an action takes place in or inside something, typically expressing inward motions. For some languages it must be differentiated from the orientations INTER 'in a mass' (cf. the local case orientation INTER). The orientation IN is found in all language families of the Caucasus with

preverbs, thereby excluding the Turkic languages, and it is functionally identical to the local case orientation IN.

(392) Georgian (Bolkvadze & Kiziria 2023: 159) (my glossing)
čit'i *še-prin-d-a* *galia-ši*
bird IN-fly-PASS-AOR.3SG cage-IN
'A bird flew into the cage'

(393) Adyghe (Kumakhov & Vamling 2009: 35) (reglossed)
ɬ'ə-m *šə-r* *Ø-d-j-e-ɕe*
man-OBL horse-ABS 3SG.ABS-IN-3SG.ERG-DYN/PRS-lead
'The man leads the horse into (something)'

(394) Kryts (Authier 2009: 197) (gloss adapted from source, retransliterated)
χinib *k'ul-dʒ-a* *ʕa-b-č'-d-u*
woman house-OBL-IN IN-F-go-AOR-F
'The woman entered the house'

(395) Iron Ossetic (Erschler 2020: 682) (reglossed)
iugɐr *nɐ=wat-mɐ* *ɐrba-χəʒ-tɐ* *wɐd=nɐm* *a-bad*
if 1PL=room-ALL IN-climb-PST.2SG then=ALL.1PL MOM-sit.IMP
'If you've gotten into our room, sit with us'

The preverb orientation SUPER

The preverb orientation SUPER indicates that an action takes place on the surface of something, typically expressing motions 'on' or 'onto'. The orientation SUPER is the second most common preverb orientation and it is also found in all language families of the Caucasus except Turkic.

(396) Kabardian (Kumaxov 2006: 471) (my glossing)
txəɬə-r *st'olə-m* *tje-ɬ-Ø-ɕ*
book-ABS table-OBL SUPER-lie-PRS-IND
'The book lies on the table'[221]

As we can see in the Lezgian example below (397), the SUPER preverb is identical to the superessive case in some Lezgic languages.

[221] Glossed and transliterated from *Тхылъыр стIолы-м телъщ*, and translated from Russian 'Книга лежит на столе'.

(397) Lezgian (Haspelmath 1993: 169) (gloss adapted from source)
ali-di *kʷ'ač-i-n* *kap-ar*
Ali-ERG foot-OBL-GEN container-PL
kʷ'ač-er-a-l *al-uk'-na*
foot-PL-OBL-SUPER SUPER-put-AOR
'Ali put his shoes on his feet'

In Megrelian and Laz the preverb orientation SUPER combines with the delative preverb direction 'down' to express the combined superdelative function 'down onto a surface', e.g. Megrelian *do-* (Reseck 2015: 65), Pazar Laz *ce-* (Öztürk & Pöchtrager 2011: 102) and *gela-* (Lacroix 2009: 402).

(398) Megrelian (Reseck 2015: 68) (gloss adapted from source)
dža-še *uškur-k* *geito-l*
tree-ABL apple-ERG SUB.DE[222]-fall.AOR.3SG
do *dixa-s* *ku-do-l*
and earth-DAT AFF-SUPER.DE-fall.AOR.3SG
'An apple fell from the tree and it fell down on the ground'[223]

The preverb orientation SUB

The preverb orientation SUB indicates that an action takes place beneath something, typically corresponding to the adverbs 'under' or 'below'. The orientation SUB is found in all three endemic language families of the Caucasus.

(399) Megrelian (Reseck 2015: 78) (reglossed)
čuan-s *diška* *k-atu-(v)-u-t'ur-e*
pot-DAT wood AFF-SUB-1SG-OV-put.wood.into.fire-AOR
'I put wood (into the fire) under the pot'[224]

(400) Abkhaz (Hewitt 2010: 116) (my glossing)
aχəza *sə-c'a-la-jt'*
blanket 1SG.ABS-SUB-enter-AOR.IND.DYN
'I went under the blanket'[225]

[222] The preverb *g(e)ito-* is a complex preverb indicating a vertical downwards motion under a horizontal surface (Reseck 2015: 79, 95), confirmed by Revaz Tchantouria (p.c.).

[223] Translated from German 'Ein Apfel fällt vom Baum und fällt auf die Erde'. Reseck glosses both forms as aorist however, and the forms are confirmed to be aorist by Revaz Tchantouria (p.c.). The original glossing of *–l* is 'fall:3SG.AOR', while it is simply the root of 'to fall'.

[224] Translated from German 'Ich legte unter dem Topf Feuer nach'. Reseck glosses the verb as 'AFF-KPRV-1SG.S:VV:anzünd:AOR' while the meaning of *t'ur-* is translated as '[Holz] nachlegen, nachfeuern' by Fähnrich (2012), confirmed by Karina Vamling (p.c.).

[225] Glossed and transliterated from *Ахыза сыҵалеит*.

(401) Lezgian (Haspelmath 1993: 170) (reglossed)

kac	*stol-di-n*	*k'anik*	*ak-ax-na*
cat	table-OBL-GEN	under	SUB-go-AOR

'The cat went under the table'

The preverb directions PROX and DIST

The preverb orientation PROX indicates that an action is proximal or directed towards or to the speaker, typically corresponding to the adverbs 'here' or 'hither'. The proximal preverb orientation is found in all Kartvelian languages (Aronson 1990: 42; Holisky 1991: 436; Fähnrich 1994: 83; Tuite 1997: 23; Öztürk & Pöchtrager 2011: 103; Reseck 2015: 97).

(402) Georgian (Aronson 1990: 101, 108) (my glossing)

imit'om,	*rom*	*xval*	*ak*	*zurab-is*	*da*	*sopik'o*
because	that	tomorrow	here	Zurab-GEN	sister	Sopik'o

mo-v-a
PROX-come-FUT.3SG
'Because Zurab's sister Sopik'o will come here tomorrow'[226]

The term *cislocative* is sometimes used for this function in the Northwest Caucasian languages, while the Northwest Caucasian cislocative affixes tend to have other functions as well (Arkadiev & Lander 2020: 415). Proximal preverbs are found in all Northwest Caucasian languages (Rogava & Keraševa 1966: 112; Smeets 1984: 253; Lomtatidze & Klychev 1989: 138; Chirikba 2003a: 54; Kumaxov 2006: 260; Fenwick 2011: 111).

(403) Adyghe (Rogava & Keraševa 1966: 112) (my glossing)

ape-w	*wəne-m*	*q-je-kʷ'e-ɬ'e-ʒə-ʁe-r*
first-ADV	house-OBL	PROX-3SG.OBL-go-TERM-REP-PFV-ABS

ajdamər	*jə-ʂeweʐəj-ew*	*qepłan*	*ar-ə*
Aydamir	3SG.POSS-boy-ADV	Qaplan	DEM-COP

'Aydamir's son Qaplan was the first to return here to the house'[227]

The proximal preverbs are also found in the Nakh languages (Čokaev 1970: 123; Holisky & Gagua 1994: 184; Nichols 2011: 346) and in the Dargic languages (Magometov 1963: 176; Magometov 1982: 75; Van den Berg 2001: 32; Sumbatova & Mutalov 2003: 63; Temirbulatova 2004: 187-189; Forker 2020a: 220).

[226] Glossed and transliterated from *იმიტომ რომ ხვალ აქ ზურაბის და სოფიკო მოვა.*

[227] Glossed and transliterated from *Апэу унэм къ-екIолIэжьыгъэр Айдамыр ишъэожъыеу Къэплъан ары* and translated from Russian 'Первым вернулся сюда в дом сын Айдамыра Каплан'.

(404) Chechen (Molochieva 2010: 122) (gloss adapted from source)
soe-ga *ħa-ħiːži-ra* *k'ant* *kest-kesta*
1SG-ALL PROX-look.PFV.ITER-WPST boy.NOM often-REDUP
'The boy looked at me (repeatedly for a while)'

The proximal preverb function is in oppositional relation with the distal preverb orientation DIST, which indicates that an action is directed away from the speaker, typically corresponding to the adverbs 'there' or 'thither'. The orientation DIST is also found in all Kartvelian languages (Aronson 1990: 42; Holisky 1991: 436; Schmidt 1991: 505-506; Fähnrich 1994: 83; Reseck 2015: 97).

(405) Georgian (Bolkvadze & Kiziria 2023: 497) (my glossing)
adre *mi-vedit,* *magram* *bilet-eb-i* *uk've*
early DIST-go.AOR.1PL but ticket-PL-NOM already
ga-q'id-ul-i *iq'-o*
ELA-sell-PTCP.PST-NOM be-AOR.3SG
'We went there early, but the tickets were already sold out'

The term *translocative* is sometimes used to describe the distal preverb function in the Northwest Caucasian languages (Arkadiev & Lander 2020: 415). Distal preverbs are common in the Northwest Caucasian languages, as they have been described in Kabardian (Kumaxov 2006: 260), Adyghe (Rogava & Keraševa 1966: 114), Shapsug Adyghe (Smeets 1984: 260), Abkhaz (Aristava 1968: 151; Chirikba 2003a: 54) and Abaza (Genko 1955: 170; O'Herin 2020: 462).

(406) Abkhaz (Aristava 1968: 152) (my glossing)
d-na-z-ga-ra-nə
3SG.M.ABS-DIST-1SG.ERG-bring-MSD-PFV.CVB
sə-q'a-n
1SG.-be-PST.STAT
'I had to bring him there'[228]

As for the proximal preverbs, distal preverbs are found in the Nakh and the Dargic languages, and are typically found in dichotomous pairs of 'hither' and 'thither'. The Khinalug preverbs *kʰa(l)-/tʰal-* and *la-* should therefore also be analysed as expressing a proximal/distal opposition, although Kibrik and Khvtisiashvili both describe the distinction as being related to being 'on the same level' or 'the same horizontal plane' while also stressing the location of the speaker (Kibrik 1994b: 396; Khvtisiashvili 2013: 211-213). The deictic analysis is further supported by

[228] Glossed and transliterated from *дназгараны сыҟан*, and translated from Russian '[Я] должен был его (чел.) туда привести'.

Rind-Pawlowski, who analyses these forms as cislocative and translocative, i.e. proximal and distal (Rind-Pawlowski 2023: 92).

(407) Khinalug (Khvtisiashvili 2013: 208)[229] (reglossed, retransliterated)
a. *si* *zɪ* *cʷ'a* *č<kal>ʁ-i-šä-mä*
3SG 1SG house bring<PROX>-PFV-PST-IND
'He brought me home ([…] The speaker is inside the house currently)'
b. *zɪ* *cʷ'a* *la-c'ʁ-i-šä-mä*
1SG house DIST-go-PFV-PST-IND
'I went in the house ([…] which is not where the speaker is currently)'

The preverb orientation POST

The preverb orientation POST indicates that an action takes place behind or after something, typically corresponding to the adverbs 'behind' or 'after'. The preverb orientation POST is therefore identical to the local case orientation POST, cf. section 5.1.2.1. The orientation POST is found in all Northwest Caucasian languages, although it appears to be absent in Abkhaz, while among the Kartvelian languages it seems to only be present in Laz (Öztürk & Pöchtrager 2011: 104; Rostovtsev-Popiel 2016)

(408) Abzakh Adyghe (Konuk 2022: 329) (gloss adapted from source)
ʔɐlɜ-r *jɐ-nɜ-m* *ɬə-kʷ'ɐ-ʁ*
boy-ABS.DEF 3PL.POSS-mother-OBL POST-go-PFV
'The boy went behind his mother'[230]

(409) Laz (Öztürk & Pöchtrager 2011: 104) (my glossing)
ek'o-gut-in-u
POST-stand-AUGM-VN
'to walk behind someone'

POST preverbs are also generally found in the Nakh, Dargic and Lezgic languages. In Aghul, the POST orientation is used when e.g. putting on or taking off clothes from behind (Magometov 1970: 159), demonstrating that the preverb can also be used to express 'behind oneself'.

[229] Khvtisiashvili gives both forms as ending with *–sämä*, which is clearly an error (Dešeriev 1959; Kibrik 1994b).

[230] Translated from French 'Le garçon est allé derrière sa mère'.

(410) Aghul (Magometov 1970: 159) (my glossing)
dada *qa-c-une* *kːul*
father POST-put-PST.PFV fur.coat
'Father put on the fur coat'[231]

However, the POST orientation is more often used to express 'to chase after something/someone', cf. example (411).

(411) Sanzhi Dargwa (Forker 2020a: 218) (gloss adapted from source)
hitːi-b-uq-un=xːar,
POST-NH-go.PFV-PRET=CONC.CVB
hitːi-a-jt-eʁ-ib
POST-NEG-DIST-go.PFV-PRET
'Even though (the hare) ran after (the turtle), it did not reach it.'

The preverb orientation ANTE

The preverb orientation ANTE indicates that an action takes place in front of something, typically corresponding to the prepositions 'before' or 'in front of'. The orientation ANTE is found in all Northwest Caucasian languages (Genko 1955: 172; Rogava & Keraševa 1966: 121, 129; Smeets 1984: 259; Paris 1989: 184; O'Herin 2002: 17, 216; Chirikba 2003a: 43; Kumaxov 2006: 264; Fenwick 2011: 112-113; Konuk 2022: 328).

(412) Abaza (O'Herin 2002: 217) (gloss adapted from source)
j-a-mc'a-gəla-p'
3SG.NH.ABS-3SG.NH.ERG-ANTE-stand-STAT.PRS
'It stands before it'

The ANTE orientation has also been described in Megrelian (Reseck 2015: 95; Harris 1991b: 359) and Laz (Öztürk & Pöchtrager 2011: 105-106), and it is also found in most Dargic languages (Magometov 1963: 179-180; Van den Berg 2001: 32; Sumbatova & Mutalov 2003: 63; Temirbulatova 2004: 191; Forker 2020a: 218) and Aghul (Magometov 1970: 158).

(413) Laz (Öztürk & Pöchtrager 2011: 106) (my glossing)
k'ots'o-xed-u
ANTE-sit-VN
'to sit in front of'

[231] Translated from Russian 'отец надел шубу'.

(414) Sanzhi Dargwa (Forker 2020a: 218) (original gloss)
sa-r-b-uqː-a *il*!
ANTE-ABL-N-carry.PFV-IMP that
'Take it away! (from in front)'

The preverb orientation AD

The preverb orientation AD indicates that an action takes place by something, typically corresponding to the adverb and preposition 'by', 'at' or 'near', and is therefore identical to the adessive local case. The orientation AD is found in most Northwest Caucasian languages, e.g. Kabardian (Kumaxov 2006: 266-268), Adyghe (Rogava & Keraševa 1966: 132), Shapsug Adyghe (Smeets 1984), Abaza (Genko 1955: 172) and Ubykh (Fenwick 2011: 113).

AD preverbs are also present in numerous Nakh-Dagestanian languages, e.g. Chechen (Čokaev 1970: 122, 129), Ingush (Nichols 2011: 365), Lezgian (Haspelmath 1993: 167), Tabasaran (Babaliyeva 2013: 159), Rutul (Alekseev 1994a: 227), Kryts (Authier 2009: 123), Aghul (Magometov 1970: 158), Xaidaq (Temirbulatova 2004: 191) and Itsari Dargwa (Sumbatova & Mutalov 2003: 63).

(415) Lezgian (Haspelmath 1993: 157) (reglossed, retransliterated)
daʁustan-di-n *žehil-ar* *pamjatnik-di-z,*
Dagestan-OBL-GEN youth-PL monument-OBL-DAT
gah *ag-at-iz,* *gah* *qaq-at-iz,* *kilig-zawa*
now AD-go-IPFV.CVB now FORL-go-IFPV.CVB look-IPFV
'The Dagestanian youngsters are looking at the monument, while going back and forth (lit. now going near, now going away)'[232]

The preverb orientation SUPRA

The preverb orientation SUPRA (from Latin *supra* 'above') indicates that an action takes place above something, typically corresponding to the adverbs 'above' or 'over'. The orientation SUPRA may appear identical to the orientation SUPER, but many languages have separate preverbs for these two orientational functions. The orientation SUPRA is found in most Northwest Caucasian languages (Genko 1955: 76; Rogava & Keraševa 1966: 132-133; Chirikba 2003: 43; Kumaxov 2006: 268). In Ubykh the connection between the orientations SUPER and SUPRA becomes apparent, as it uses the preverb *bʁʲʒ-* for both orientations (Fenwick 2011: 112). The Circassian SUPRA preverbs are all based on the noun *ҽћe* 'head' as they can also have specific reference to the head (Kumaxov 2006: 268), which is also true for the Abkhaz SUPRA preverb *xə-* which is identical to *a-xə́* 'head' (Chirikba 2003: 43).

[232] Haspelmath gives the translation 'The Daghestanian youngsters are looking at the monument, now approaching, now moving away'.

(416) Adyghe (Rogava & Keraševa 1966: 132) (my glossing)
ʂʷəzə-r *čew-me* *qa-ʂħede-tɛ'-ze*
woman-ABS fence-OBL.PL PROX-SUPRA-get.over-SIM.CVB
wərəsbəj *dez* *qe-kʷ'-a-ʁ*
Urysbij near PROX-go-PFV-PST
'The woman, while getting over the fences, came to Urysbij'[233]

SUPRA preverbs are also found in Georgian (Hewitt 1995: 235; Bolkvadze & Kiziria 2023: 160), Old Georgian (Fähnrich 1994: 83) and Laz (Holisky 1991: 436; Lacroix 2009: 402). In Georgian and Old Georgian the situation is somewhat complicated by the apparent overlap between the SUPRA function and the translative, as the preverb *ga(r)da-* is consistently translated as indicating both 'over' and 'across'. This primarily a translational issue, as the preverb *ga(r)da-* should optimally be analysed as supratranslative.

(417) Georgian (Bolkvadze & Kiziria 2023: 192) (my glossing)
čit'i *am* *mta-s* *gada-i-pren-s*
bird DEM.DAT mountain-DAT SUPRA-FUT-fly-PRS.3SG
'The bird will fly across (lit. over) this mountain'

The preverbs *čew-* and *čitːi-* appear to indicate the orientation SUPRA in Xaidaq (Temirbulatova 2004: 189), but no examples are given. There are potentially numerous SUPRA preverbs in the Nakh-Dagestanian languages if preverbs indicating movement described as 'down from above' were included, but I have encoded these preverbs as only delative, cf. section 6.11.2.

The preverb orientation HAND

The preverb orientation HAND indicates that an action is related to the hands, typically corresponding to constructions of the type 'in/into the hands' or 'from the hands'. The HAND orientation should likely be regarded as lexical affixes, thus comparable to incorporated nouns, while they often refer to the hand as a landmark rather than an object or an instrument. It has therefore been included in this study, particularly since it occurs in both the Northwest Caucasian and Nakh-Dagestanian language families. The orientation HAND is found in most Northwest Caucasian languages, e.g. Kabardian (Kumaxov 2006: 279), Adyghe (Rogava & Keraševa 1966: 128), Shapsug Adyghe (Smeets 1984: 260), Abkhaz (Hewitt 2010: 115; Aristava 1968: 157-158), Ubykh (Fenwick 2011: 113). The Northwest Caucasian languages typically also have a preverb orientation related to the feet (Aristava

[233] Glossed and transliterated from *Шъузыр чэумэ къа-шъхьэдэ-кIэ Урысбый дэжь къэкIуагъ*, and translated from Russian 'Женщина, переходя через плетни, пришла к Урысбию'.

1968: 153; Rogava & Keraševa 1966: 127; Kumaxov 2006: 204), but I have not been able to find this in any of the Nakh-Dagestanian languages.

(418) Ubykh (Fenwick 2011: 114) (reglossed)
sɨʁʷɜ *sɨ-w-q'ɜː-ʁɜ-tʷ'-ɐj-ɜw*
1SG 1SG.ABS-2SG.OBL-HAND-ABL-escape-ITER-FUT
'I will escape from you again' (lit. 'I will escape from your hands again')

HAND preverbs are also found in a handful of Nakh-Dagestanian languages, e.g. Ingush (Nichols 2011: 365), Aghul (Magometov 1970: 159), Kubachi (Magometov 1963: 179-180), Itsari Dargwa (Sumbatova & Mutalov 2003: 63) and Sanzhi Dargwa (Forker 2020a: 218).

(419) Itsari Dargwa (Sumbatova & Mutalov 2003: 65) (my glossing)
b=at-kʷi-r-ma-ka-b=uˤq-aq-i-tː
NH=leave-HAND-ABL-PROH-DE-NH=LV-CAUS-PROH-2SG
'Do not let (it) out (from your hands downwards)'

The preverb orientation CONT

The preverb orientation CONT indicates that an action is carried out in contact with something, as it is more or less identical to the local case orientation CONT (cf. section 5.1.2.1). The vertical element of the local case CONT appears to be less prominent for the preverb orientation, which has led me to treat them separately.

The preverb orientation CONT is found in numerous Lezgic languages, e.g. Tabasaran (Babaliyeva 2013: 159), Rutul (Ibragimov 1978: 98), Aghul (Magometov 1970: 158) and Budukh (Alekseev 1994b: 271). In Lezgian the orientations CONT and SUB seem to have merged into the preverb *(V)k-*, as it combines with typical CONT forms such as *k-äʁun* 'touch' and *ek-isun* 'hit' (Haspelmath 1993: 171-172), cf. Aghul CONT *kV-* and SUB *kːV-* (Magometov 1970: 158).

(420) Aghul (Magometov 1970: 160) (my glossing)
ke-q-as
CONT-put-INF
'to hang on a wall, on a tree, on a hook'[234]

In Rutul, the CONT orientation is obligatory for certain verbs that by definition involve contact, such as the verbs *ki-xis* 'write' and *ki-jq'as* 'touch' (Ibragimov 1978: 98).

[234] Translated from Russian 'повесить на стену, на дерево, на крюк'.

(421) Rutul (Maxmudova 2001: 130) (reglossed)
naida-ra za-s misga kaʁat-biʀ ki-r-xe-re
Naida-ERG 1SG-DAT always letter-PL.ABS CONT-IPFV-write-GNOM
'Naida always writes me letters'[235]

Preverbs expressing the orientation CONT are also potentially found in some Northwest Caucasian languages, but this needs further investigation. Possible examples are Kabardian *č'erə-* (Kumaxov 2006: 267), Adyghe *ʂʷ'axe-* (Rogava & Keraševa 1966: 132) and Abkhaz *šʷə-* (Hewitt 2010: 114). In Kartvelian, the Laz preverb *ok'o-* seems to at least partly indicate the orientation CONT (Holisky 1991: 436; Öztürk & Pöchtrager 2011: 108), which is potentially also true for the Megrelian preverb *ak'o-* (Reseck 2015: 74-75).

The preverb orientation INTER

The preverb orientation INTER indicates that an action takes place 'within a mass' or 'within a liquid', and is therefore identical to the spatial case function INTER (cf. section 5.1.2.1). The orientation INTER is found in all Northwest Caucasian languages (Genko 1955: 172; Rogava & Keraševa 1966: 115, 134; Smeets 1984: 260, 274; Spruit 1986: 29; Paris 1989: 196; Kumaxov 2006: 261; Hewitt 2010: 115; Fenwick 2011: 113; Konuk 2022: 332). Abkhaz also seems to have a separate preverb to indicate location 'within a liquid' (Spruit 1986: 24).

(422) Abzakh Adyghe (Konuk 2022: 332) (gloss adapted from source)
ordɜk-xɜ-r psə-m xɜ-s-əx
duck-PL-ABS water-OBL INTER-sit-3PL.ABS
'The ducks are on the river (lit. sitting in the water)'[236]

INTER preverbs are also found in a few Lezgic languages, e.g. Tabasaran (Babaliyeva 2013: 159) and Rutul (Alekseev 1994a: 227; Maxmudova 2001: 14).

(423) Rutul (Maxmudova 2001: 14) (my glossing)
gada xidi-k kaˁ-č'u-ri
boy.ABS water.OBL-INTER INTER-go-PST
'The boy went into the water'[237]

[235] Transliterated from *Наида-ра за-с мысга кагъат-быр кирхье-ре*, and translated from Russian 'Найда всегда пишет мне письма'.

[236] Translated from French 'Les canard sont dans la rivière'.

[237] Glossed and transliterated from *гада хьидик каI-чIури*, and translated from Russian '[М]альчик в воду вошел'.

The preverb orientations VERT, LATER and APUD

The preverb orientations VERT and LATER indicate that an action has a vertical or lateral relation to a landmark, typically indicating vertical landmarks or actions relating to the side of a landmark. It is important to point out that these preverbs do not imply vertical or lateral movement, as they refer to orientation and not direction. While I have tried to differentiate between the vertical and lateral orientations, these orientations overlap in e.g. Ubykh (Fenwick 2011: 112) and Megrelian (Reseck 2015: 84-85).

The orientations VERT and LATER appear to be found in some form in most Northwest Caucasian languages e.g. Kabardian *bʁʷərə-* (Kumaxov 2006: 267), Adyghe *gʷa-* (Rogava & Keraševa 1966: 125), Abkhaz *k'əd-* (Hewitt 2010: 120) and Ubykh *fɜ-* (Fenwick 2011: 112), but these preverbs need further investigation.

In the Kartvelian languages the VERT and LATER orientations have been described in both Megrelian (Harris 1991b: 359; Reseck 2015: 84-87) and Laz (Holisky 1991: 436; Öztürk & Pöchtrager 2011: 104). The various Megrelian complex preverbs formed with *la-* appear to mainly express a vertical orientation, but Reseck translates its orientation as related to 'Seite' (Reseck 2015: 95), which indicates the close affiliation between the vertical and lateral orientations. The orientation LATER is potentially also found in Chechen with possible cognates in Ingush, e.g. the preverb *ayuor-* 'on the side' (Čokaev 1970: 123, 138).

The preverb orientation APUD indicates that an action takes place next to something, typically corresponding to the preposition 'next to' or 'beside'. The preverb orientation APUD is semantically close to the orientation LATER, while APUD preverbs imply by definition a lack of contact. Preverbs that seem to expressing the orientation APUD are found in Kabardian (Kumaxov 2006: 267), Adyghe (Rogava & Keraševa 1966: 132), Abzakh Adyghe (Konuk 2022: 331), Shapsug Adyghe (Smeets 1984: 259), Abkhaz (Aristava 1968: 153; Hewitt 2010: 121), Abaza (Genko 1955: 172; O'Herin 2002: 216). Defining the exact relationship between these orientations is unfortunately beyond the scope of this thesis and would need further research and data.

The preverb orientation HOME

The preverb orientation HOME (or HOUSE) indicates that an action is specifically related to one's home or to a house in general. The orientation HOME is found in Abkhaz (Aristava 1968: 153; Hewitt 2010: 121), Shapsug Adyghe (Smeets 1984: 261), Chechen (Molochieva 2010: 21-22) and Ingush (Nichols 2011: 365).

(424) Chechen (Molochieva 2010: 21) (gloss adapted from source)

so	*c'a-v-eʔ-ča*	*naːn-na*
1SG.ABS(V)	HOME-V-come-TEMP.CVB	mother-DAT
xaza	*xiːti-ra*	
beautiful	seem.WPST	

'Mother was happy when I came home'

The Abaza preverb *ʕʷna-* possibly has the orientation HOME as well (Genko 1955: 172; O'Herin 2002; 17; O'Herin 2020: 463), but it is often difficult to distinguish the orientations IN and HOME as they both convey the meaning 'inside' or 'indoors'.

(425) Abaza (O'Herin 2002: 128) (gloss adapted from source)
də-ʕʷna-sə-r-čʷ'a-t'
3HSG.ABS-HOME-1SG.ERG-CAUS-sit-DYN
'I seated him/her in the house'

6.11.2. Directional spatial functions

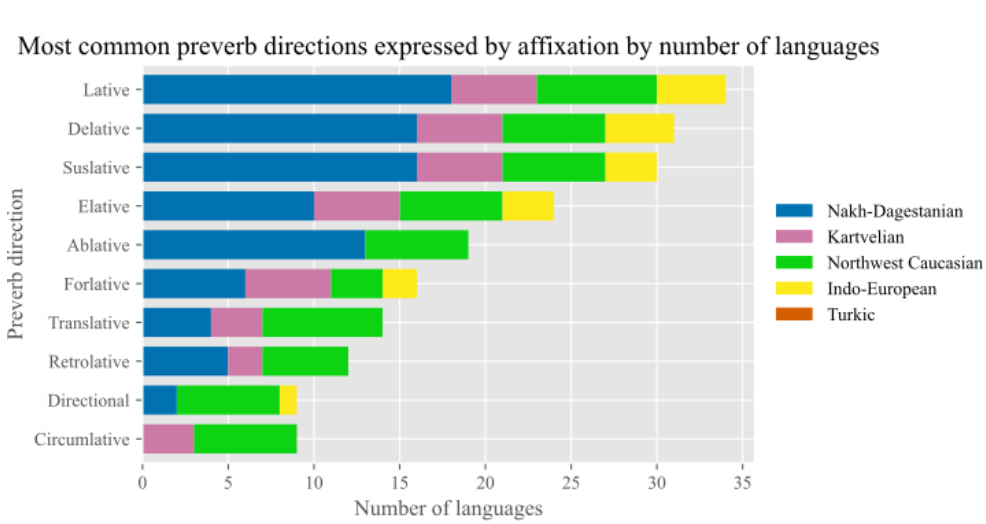


Figure 6.11: The most common preverb directions expressed through affixation in the data according to number of languages.

The Lative preverb direction

Lative preverbs indicate a motion to an object, typically corresponding to the preposition 'to', and a lative direction is implied in most orientational preverbs. The lative direction is the most general direction to or towards an object, which means that preverbs typically have a lative direction if another direction has not been specified, and it is identical to the lative spatial case direction (cf. section 5.1.2.2). The Northwest Caucasian languages demonstrate that the category *spatial preverbs* is not optimal from a functional perspective since most Northwest Caucasian languages also encode spatial information on the verb by means of suffixes. Lative verbal suffixes are found in e.g. Abzakh Adyghe (Paris 1989: 184), Shapsug Adyghe (Smeets 1984: 274), Abkhaz (Chirikba 2003a: 39) and Abaza (Lomtatidze &

Klychev 1989: 105). Since the lative direction is implied in most orientational preverbs, it is implicitly found in all languages with spatial preverbs in the Caucasus.

The Delative preverb direction

Delative preverbs indicate a motion that is directed downwards, typically corresponding to the adverbs 'down' or 'downwards'. The term delative (from the Latin prefix *de-* 'from, down, off') is not conventionally used for preverbs as it is primarily used for the delative case in Hungarian with the function of expressing motion 'from/off' a surface (Kenesei, Vago & Fenyvesi 1998: 244). The Hungarian delative case is from a Caucasian perspective technically a superelative, and the term delative for the Hungarian case might therefore be misleading. The Hungarian delative case can express the function 'down', but it also conveys all other motions from a surface, making any further analogies between the Hungarian delative case and the preverb direction 'down' problematic. I have however chosen to apply the term *delative* to preverbs indicating a downward motion, as the Latin verbal prefix *de-* specifically indicates a downward motion, and e.g. Rogava & Keraševa (1966: 291) also use the term delative in the same sense.

(426) Chechen (Molochieva 2010: 226) (reglossed)

so	*na:bar-ieħ*	*oha-j-oež-na*	*xilla*
1SG.ABS(J)	sleep-LOC	DE-J-fall.PFV-PTCP.PFV	be.UWPST

'I fell down (from the bed) in [my] sleep'

Delative preverbs are found in all Kartvelian languages, where the simple preverb *da-* 'down' in Georgian often combines with the orientational preverb *še-* 'into', which developed from **še-da-* into Old Georgian *šta-* and subsequently modern Georgian *ča-* 'down into' (Shanidze 1982: 27; Bolkvadze & Kiziria 2023: 160).

(427) Old Georgian (Tuite 1998: 66) (gloss adapted from source)

mraval	*gz-is*	*šta-vard-i-s*	*igi*	*cecxl-sa*
many	way-GEN	IN.DE-fall-IPF-3SG	3SG	fire-DAT

'He often falls into the fire'

(428) Georgian (Bolkvadze & Kiziria 2023: 160) (my glossing)

bavšv-i	*k'ibe-ze*	*ča-vid-a*
child-NOM	stair-SUPER	IN.DE-go-AOR.3SG

'A child went down the stairs'

The Northwest Caucasian languages once again challenge the validity of the concept of spatial *preverbs*, as Kabardian, Adyghe and Abzakh Adyghe all have delative verbal circumfixes where both the associated prefix and suffix convey the delative function (Rogava & Keraševa 1966: 291; Paris 1989: 197; Kumaxov 2006: 287) while Shapsug Adyghe appears to only have a delative suffix (Smeets 1984: 275).

(429) Adyghe (Rogava & Keraševa 1966: 291) (my glossing)
ħaneχʷ-ər *šə-m* *je-wəkʷ'areje-xə-ʁ*
Xanaxu-ERG horse-OBL DE-fall-DE-PST
'Xanaxu fell down from the horse'[238]

The Suslative preverb direction

Suslative preverbs indicate a motion that is directed upwards, typically corresponding to the adverbs 'up' or 'upwards'. There is no conventional term for the suslative preverb direction. Rogava & Keraševa (1966: 291) use the term *sublative* in this sense, which is derived from the Latin prefix *sub-* 'upwards from below', but the term sublative must be avoided due to the Nakh-Dagestanian local case sublative (cf. section 5.1.2). I therefore propose the term *suslative* (from Vulgar Latin *sūsum*[239] 'upwards' from *sūrsum* (Väänänen 1981: 62)) in order to introduce a specific term for preverbs indicating an upward motion.

(430) Old Georgian (Tuite 2008a: 161) (my glossing)
aɣ=dg-om-il *ar-s* *mk'wdr-et-it*
SUS=stand-ITR-PTCP.PFV be-3SG.PRS dead-PL-INS/ABL
'He has risen from the dead'

(431) Itsari Dargwa (Sumbatova & Mutalov 2003: 82) (reglossed)
u *učitil* *aɣ-u-tː-aj-maˁq'inna*
2SG teacher become.PFV-3.A-2.P-SUBJ-TERM.CVB[240]
amha *kːalkː-li-j* *haˌb=uˁq-an-ni*
ass tree-OBL-SUPER.LAT SUS.NH=climb.IPFV-NEC-FUT
'Before you become a teacher, an ass will climb up the tree!'

As with the delative affixes mentioned above, the Circassian languages stand out, as they all have suslative circumfixes (Rogava & Keraševa 1966: 291; Smeets 1984: 439; Paris 1989: 185; Kumaxov 2006: 80).

(432) Adyghe (Arkadiev & Lander 2020: 382) (reglossed)
de-kʷ'e-ja-ʁ
SUS-go-SUS-PST[241]
'S/he went up'

[238] Glossed and transliterated from *Хьанэхъур шым е-укIорэе-хыгъ* and translated from Russian 'Ханаху упал вниз (свалился) с лошади'.

[239] Cf. the Latin phrase *susque dēque*, literally 'both up and down'.

[240] Sumbatova and Mutalov's glosses it as 'become:PF-U-2-SUBJ-until'.

[241] Arkadiev and Lander glosses this as 'LOC-go-UP-PST' however.

The Elative preverb direction

Elative preverbs indicate that a motion has the direction out of a source, typically corresponding to the adverbs 'out', 'out of' or 'outwards' (Blake 2001: 153). The term *elative* (from Latin *ex-* 'out') is used both for Nakh-Dagestanian and Uralic spatial cases (De Hoop & Zwarts 2009: 179-180; Haspelmath 2009: 516). However, the functions of the Nakh-Dagestanian and the Uralic elative cases differ somewhat, as the Nakh-Dagestanian elative indicates all motions from a source (Haspelmath 2009: 516; Creissels 2009a: 617), whereas the Uralic elative specifically indicates movement outwards from inside (Spencer 2009: 196; Haspelmath 2009: 516), thus corresponding to the Nakh-Dagestanian *inelative* (Creissels 2009a: 617). I have chosen to follow the Uralic usage of the term elative, since it is derived from the Latin prefix *ex-*, which has an identical function.

(433) Ingush (Nichols 2011: 432) (gloss adapted from source)
zhwalii *aara-d-ealar*
dog ELA-D-go.WPST
'The dog went out'

(434) Laz (Öztürk & Pöchtrager 2011: 85) (gloss adapted from source)
bere *avla-şe* *gama-xt'-a-s*
child.NOM garden-ALL ELA-go-SUBJ-PRS.3SG
'The child may/can go out to the garden'

(435) Iron Ossetic (Bagaev 1965: 157) (my glossing)
sqoladzaw-tæ *ra-cid-isti* *sqola-jæ*
pupil-PL ELA-go.PST-PST.3PL school-ABL
'The pupils went out of the school'[242]

The Ablative preverb direction

Ablative preverbs indicate a motion from a source, typically corresponding to the preposition 'from' (De Hoop & Zwarts 2009: 179-180). The ablative preverb direction is therefore functionally identical to the ablative case (cf. section 5.1.2.2), and it is similar but less specific than the elative direction described above. As for the ablative/elative spatial case direction, the ablative preverb affixes can often not occur on their own, as they typically attach to or are intertwined with the preverb orientation. The ablative preverb direction is found in all Northwest Caucasian languages and in most Nakh-Dagestanian languages with preverbs, e.g. Chechen (Čokaev 1970: 123), Ingush (Nichols 2011: 652), Rutul (Maxmudova 2001: 122),

[242] Glossed and transliterated from *Скъоладзаутæ рацыдысты скъолайæ*, and translated from Russian 'Ученики вышли из школы'.

Aghul (Magometov 1970: 159), Khinalug (Khvtisiashvili 2013: 213) and most Dargic languages (Sumbatova 2020: 164).

(436) Ubykh (Fenwick 2011: 114) (gloss adapted from source)
ɐ-w-ʁʒ́-sɨ-wːtʷʼɨ-n
3SG.ABS-2SG.OBL-ABL-1SG.ERG-take.DYN-PRS
'I take it away from you'

(437) Rutul (Maxmudova 2001: 104) (reglossed)
zer *dahr-aː-la* *liː-b-xu-ri*
cow.ABS cliff-SUPER-ELA CONT/VERT.ABL-III.SG-fall-PST
'The cow fell off the cliff'[243]

The Northwest Caucasian languages stand out once more, as some of them have ablative verbal suffixes instead of prefixes, cf. the *delative* and *suslative* preverb directions, which is true for Abzakh Adyghe (Paris 1989: 184), Abkhaz (Chirikba 2003a: 39) and Abaza (Lomtatidze & Klychev 1989: 105). In Kabardian and Adyghe the ablative function is typically expressed by the verb root *čʼə-* 'go out/from' (Kumaxov 2006: 80, 241; Arkadiev & Lander 2020: 394).

The 'Forlative' preverb direction

Forlative preverbs indicate that a motion is directed away from the speaker or a source, typically corresponding to the adverb 'away'. Since this preverb direction lacks any analogous terms in previous literature, I have chosen to introduce the term *forlative* derived from Old French *for-*, as the Old French prefix *for-* has been used to calque the West Germanic prefix **fur-* indicating e.g. motion away from a source, cf. the cognates German *ver-* and English *for-*. The forlative direction is functionally related to the distal orientation, as distal constructions can often be used to express motion away from the speaker without a specific spatial deixis, while the forlative does not convey deictic information.

(438) Lezgian (Haspelmath 1993: 273) (reglossed)
čun *ada-w-aj* *qaq-at-na*
we.ABS 3SG-AD-ELA FORL-go-AOR
'We went away from him'

[243] Transliterated from *Зер дагьр-аа-ла пии-б-хьу-ри*, and translated from Russian 'Корова сорвалась со скалы'.

(439) Megrelian (Rostovtsev-Popiel 2020: 557) (reglossed)
k'in=i *mida-Ø-rt-es* *o.nadir.u-ša* *boš-ep-k*
back=EV[244] FORL-3.S-go-PST.3SG hunt.MSD[245]-ALL boy-PL-ERG
'The boys left for hunting again'

The Translative/Perlative preverb direction

A Translative or Perlative direction indicates a transitory 'motion through the domain' (Daniel & Ganenkov 2009: 674), typically corresponding to the prepositions 'through' or 'along' (Haspelmath 2009: 515-516). The term *translative* is conventionally used to describe the corresponding spatial case direction in Nakh-Dagestanian languages (Daniel & Ganenkov 2009: 674), although a plethora of other terms are used for the same function in other language families, e.g. *perlative*, *prolative* and *mediative* (Haspelmath 2009: 515).

(440) Ubykh (Fenwick 2011: 115) (reglossed)
ɐ-q^{wʕ}'ʒ́-n *ʃi-Ø-bɜtɛ'ɜ-ʁɜ-lɜ-χ^{w}ɜ-q'ɜ-n*
DEF-cavern-OBL 1PL.ABS-3SG.OBL-SUB-TRANS-LOC-pass-PST-PL
'we passed through [lit. 'through under'] the cavern'

(441) Ingush (Nichols 2011: 411) (gloss adapted from source)
cysjk *jaashjkaa* *jiq'ie-gholla* *chaq-iiqqar*
cat box.DAT in.middle-along TRANS[246]-dash.WPST
'The cat ran through the box'

(442) Laz (Lacroix 2009: 420) (reglossed)
ma *çai-s* *goşo-b-ul-u*
1SG tea-DAT TRANS-1SG-pass-TS
'I pass through the tea (field)'[247]

The 'Retrolative' preverb direction

Retrolative preverbs indicate a motion back to a previous position, typically corresponding to the adverb 'back'. I am proposing the term *retrolative* (from Latin *retrō* 'back') as this function appears to lack a term in the previous literature. Retrolative verbal affixes are found in all the three endemic language families of the

[244] Rostovtsev-Popiel's gloss for 'euphonic vowel'.

[245] Rostovtsev-Popiel glosses this as 'SUPINE-hunt-SUPINE-ALL', but it should rather be analysed as a masdar (Vamling & Tchantouria 1993: 72), cf. the Laz masdar *o-…-u* (Holisky 1991: 439).

[246] Nichols glosses this as 'through-'.

[247] Translated from French 'Je passe à travers le champ de thé'.

Caucasus, in e.g. Georgian (Hewitt 1995: 169; Boeder 2005: 22)[248], Old Georgian (Shanidze 1982: 82; Fähnrich 2012: 389), Ingush (Nichols 2011: 365), Bats (Holisky & Gagua 1994: 184), Tsakhur (Ibragimov 1990: 124-125), Rutul (Ibragimov 1978: 98), Abkhaz (Hewitt 2010: 119), Adyghe (Rogava & Keraševa 1966: 310) and Shapsug Adyghe (Smeets 1984: 275).

(443) Old Georgian (Shanidze 1982: 82) (my glossing)
a. *uk'un-i-kc-e-n*
RE-SV-turn-OPT-3PL
'They returned/turned back'[249]
b. *uk'un-s-c-a*
RE-3SG.O-give-AOR.3SG
'He gave it back to him'

(444) Rutul (Maxmudova 2001: 82)

wa-xde	*minnet*	*w-iʔi,*	*q-iq'-a!*
2SG-SUB	plea.ABS	III-be.PRS	RE-come-IMP

'I beg you (lit. there is grace under you), come back!'[250]

In Adyghe the retrolative suffix *-ʒə* is identical to the repetitive aspect (Rogava & Keraševa 1966: 310-311), indicating a semantic connection between these grammatical functions in Adyghe, which is likely also true for Kabardian and Ubykh, cf. section 6.2.

(445) Adyghe (Rogava & Keraševa 1966: 279) (my glossing)
se *sə-kʷ'a-ʒə-ʂʷə-ɕt-ep*
1SG 1SG-go-RE-POT-FUT-NEG
'I will not be able to return/go back'[251]

The 'Circumlative' preverb direction

A Circumlative direction indicates a motion around something, typically corresponding to the preposition 'around'. The term *circumlative* (from Latin *circum-* 'around') is not conventionally used with reference to preverbs, but the term has been applied to an identical case function in older descriptions of the Classical

[248] The Old Georgian preverb *uk'u(n)-* 'back' has largely been lost in modern Georgian, except in constructions such as *uk'u-a-gd-eb* 'you throw X/them back' (Hewitt 1995: 169).

[249] Translated from German '[S]ie kehrten zurück'.

[250] Transliterated from *Ва-хьде миннет в-иъи, хъ-икь-а*, and translated from Russian 'Прошу тебя, вернись'. The light verb construction *minnet wiʔin* is however translated as *молить* 'pray, beg' by Ismailova (2011), most likely borrowed from Azerbaijani *minnət* 'favour, grace'.

[251] Glossed and transliterated from *Сэ сыкIо-жъы-шъу-щтэн* and translated from Russian '[Я] не смогу вернуться'.

Armenian *casus circumlativus* (Lauer 1869: 89). Circumlative verbal affixes are found in most Kartvelian languages, e.g. Georgian (Bolkvadze & Kiziria 2023: 159, 165), Old Georgian (Fähnrich 1994: 84; Tuite 2008a: 156), Megrelian (Rostovtsev-Popiel 2020: 554), Laz (Öztürk & Pöchtrager 2011: 102). In Georgian and Old Georgian the circumlative preverb is composed of the preverbs *mi-* 'thither' and *mo-* 'hither', literally meaning 'hither and thither' or 'back and forth' (Tuite 2008a: 156; Bolkvadze & Kiziria 2023: 159), while the preverb *da-* alone can also convey the meaning 'around' in modern Georgian (Hewitt 1995: 164; Bolkvadze & Kiziria 2023: 165).

(446) Old Georgian (Tuite 2008a: 156) (reglossed and translated into English)
mi.mo-da=x-xed-v-id-a
CIRC-DE=3SG.O-look-SM-IPF-3SG
'circumspectavit' = 'he/she looked around (onto)'[252]

(447) Georgian (Bolkvadze & Kiziria 2023: 357) (my glossing)
sakartvelo-s q'vela raioni da-m-i-vli-a,
Georgia-GEN every district CIRC-1SG.O-SV-walk-PRF.3SG
magram aseti lamazi adgili ar m-i-nax-av-s
but such beautiful place NEG 1SG.O-SV-see-SM-PRF.3SG
'I have been (walked over, around) in every region of Georgia, but have not seen such a beautiful place'

Circumlative preverbs are also found in the Northwest Caucasian languages, as preverb constructions with a circumlative function have been described in e.g. Kabardian (Kumaxov 2006: 287), Adyghe (Rogava & Keraševa 1966: 289), Abkhaz (Hewitt 2010: 121) and Abaza (Klyčev 1995: 272). In Adyghe, the proximal prefix *qə-* and the lexical suffixes *-tɛ'ə* (*-кIы*) or *–ħe* (*-хьэ*) form circumlative circumfixes (Rogava & Keraševa 1966: 289), and similar constructions are found in the other Circassian languages (Smeets 1984: 439; Paris 1989: 185; Kumaxov 2006: 283).

(448) Adyghe (Rogava & Keraševa 1966: 289) (my glossing)
sjə-č'əg^{w} ʁ^{w}əne-xe-r qə-s-k^{w}'ə-ħa-ʁe-x
1SG.POSS-land edge-PL-ABS CIRC-1SG.ERG-go-LV-PFV-3PL.ABS
'I have walked around the edges of my lands'[253]

[252] Translation provided by Manana Kock Kobaidze (p.c.), while the preverb *da-* primarily indicates downard motions (Fähnrich 1994: 83).

[253] Glossed and transliterated from *СичIыгу гъунэхэр къэ-с-кIу-хьа-гъэх* and translated from Russian 'Я свои земли обошёл'.

Other preverb directions

Finally, there are certain preverb directions that are either highly unusual, poorly described, or simply lacking in previous descriptions. The Terminative direction indicates a motion which ends at something, typically corresponding to the adverb 'until', and it appears to be found in Shapsug Adyghe (Smeets 1984: 275). The terminative direction is identical to the terminative spatial case, but Smeets does not give any examples which makes it unclear if it is truly a terminative verbal affix. I have not found terminative verbal constructions in the other Northwest Caucasian languages, which might suggest that they do not occur or that I have failed to identify them.

The preverb *ble-* in Kabardian, Adyghe and Shapsug Adyghe expresses motion 'past' or 'further, beyond'[254] (Rogava & Keraševa 1966: 128; Smeets 1984: 259; Kumaxov 2006: 265), which I analyse as a 'ultralative' direction indicating movement past or beyond a landmark.

The Chechen prefix *ħalxa-* indicates motion forward (Čokaev 1970: 122). Since this would constitute a prolative direction with the *sensu stricto* of the Latin prefix *pro-* 'forward', I would suggest only using the term prolative for forward motions.

A directional preverb orientation, indicating movement to or toward a landmark, is potentially found in some Northwest Caucasian languages. These directional suffixes, typically with the form *-ɬ'e*, are sometimes described as suffixes (Smeets 1984: 275; Paris 1989: 197), derivational suffixes (Kumaxov 2006: 239) or as secondary verb roots (Rogava & Keraševa 1966: 290), which would suggest that they could be serial verb constructions instead.

[254] Translated into Russian as 'мимо, дальше'.

6.12. Summary of the verbal affixation systems in the Caucasus

As with the nominal functions in section 5.8, I will summarise the observed verbal affixation systems for each grammatical category, which are then summarised into a typological comparison between the verbal affixation systems of all 56 languages. All systems will again be categorised on an alphabetical scale where A is the smallest observed system and Z would be the theoretically largest system, and zero will indicate that the category is either absent or not expressed by affixation.

The tense systems can be group into three categories, where category A has two tense affixes, of which one is a past tense affix, as these are found in all languages of the Caucasus. The B category is specific to some Kartvelian languages, as the future tense does not have a dedicated affix. The last category contains tense systems with explicit past, present and future tense affixes. If the general tense had been considered a tense, cf. section 6.1, it would be included in a fourth category.

Table 6.4: Tentative summary of all tense systems expressed by affixation in the Caucasus

Tense affixation systems by size, where each system type contains any of the listed functions.	
A	2: PST, PRS, FUT
B	3: PST, PRS, PV+PRS=FUT
C	3: PST, PRS, FUT

The aspectual systems range from a two-way distinction between perfective and imperfective, where only one of these aspects is obligatorily marked with an affix, to systems of up to ten aspects. The B category lacks a perfective/imperfective distinction, although these languages instead have a specific generic aspect and/or an iterative. The remaining systems are all built on a fundamental perfective/imperfective distinction, while also adding habitual, progressive and aorist aspects. The larger systems also typically include resultative, repetitive, iterative and inchoative aspects, while the largest Northwest Caucasian systems also express excessive, dynamic and exhaustive aspects by means of affixation.

Table 6.5: Tentative summary of all aspectual systems expressed by affixation in the Caucasus

Aspectual affixation systems by size, where each system type contains any of the listed functions.	
A	1-2: PFV, IPFV
B	1-2: GNOM, ITER
C	2-4: PFV, IPFV, HAB, PROG, AOR, GNOM
D	3-5: PFV, IPFV, HAB, PROG, AOR, GNOM, RES, ITER
E	5-7: PFV, IPFV, HAB, PROG, AOR, RES, EXC, REP, DYN
F	7-9: PFV, IPFV, HAB, PROG, AOR, GNOM, RES, ITER, INCH, PROSP
G	7-10: PFV, IPFV, HAB, PROG, AOR, RES, EXC, ITER, REP, DYN, INCH, EXH, MOM

The modal systems exhibit a larger variation, as they range from the smallest systems with a three-way distinction between indicative, imperative and subjunctive/irrealis, up to the largest systems that express up to eleven moods. The larger systems are found in both the Nakh-Dagestanian and Northwest Caucasian language families, where they typically include the (often zero-marked) indicative, optative, interrogative, imperative, subjunctive/irrealis, prohibitive, conditional, potential, hortative, necessitative/debitive and dubitative affixes.

Table 6.6: Tentative summary of all modal systems expressed by affixation in the Caucasus

Modal affixation systems by size, where each system type contains any of the listed functions.	
A	3: (IND), IMP, SUBJ
B	3: (IND), Q, PERM
C	4: (IND), OPT, Q, IMP, SUBJ
D	4-6: (IND), OPT, Q, IMP, SUBJ, PROH, COND
E	4-6: (IND), OPT, Q, IMP, SUBJ, PROH, COND, POT, HORT
F	6: (IND), OPT, Q, IMP, PROH, DESID
G	6-8: (IND), OPT, Q, IMP, SUBJ, PROH, COND, POT, HORT, NEC/DEB
H	7-8: (IND), OPT, Q, IMP, SUBJ, PROH, COND, POT, HORT, NEC/DEB, MIR, DUBIT, APPR
I	9-10: (IND), OPT, Q, IMP, SUBJ, PROH, COND, POT, NEC/DEB, PERMIS
J	9-11: (IND), OPT, Q, IMP, SUBJ, PROH, COND, POT, HORT, NEC/DEB, DUBIT, APPROB

The evidential systems are primarily based on a distinction between non-witnessed and witnessed events, where many languages only explicitly mark one of these evidential forms. The larger systems are based on this non-witnessed/witnessed distinction while adding quotative, hearsay and assumed evidential-marking affixes. Category A is peculiar as it includes either a quotative or a hearsay evidential, which could be problematised, as the distinction between the hearsay and non-witnessed evidential is not always clear-cut.

Table 6.7: Tentative summary of all evidential systems expressed by affixation in the Caucasus

Evidential affixation systems by size, where each system type contains any of the listed functions.	
A	1: Quotative, Hearsay
B	1-2: Non-witnessed, Witnessed
C	1-2: Non-witnessed, Assumed
D	2-3: Non-witnessed, Witnessed, Quotative
E	3-4: Non-witnessed, Witnessed, Quotative, Hearsay
F	4: Non-witnessed, Witnessed, Hearsay, Assumed

The valency systems almost all include a causative, while passive, reflexive and transitive/intransitive and reciprocal affixes are widespread but not evenly distributed. Category E includes the Kartvelian versions, which could potentially be reanalysed as other grammatical functions. The Northwest Caucasian valency systems all express various applicatives and they are summarised as category F.

Table 6.8: Tentative summary of all valency systems expressed by affixation in the Caucasus

Valency-changing affixation systems by size, where each system type contains any of the listed functions.	
A	1: CAUS
B	2: CAUS, PASS, TR, ITR, MP
C	2-3: CAUS, REFL, PASS, TR, ITR, ANTIP
D	4-5: CAUS, REFL, PASS, TR, ITR, REC, ANTIC
E	8-9: CAUS, PASS, TR, ITR, LOC, OV, IOV, SV, NV, ANTIC
F	8-9: CAUS, REFL, REC, BEN, COMIT, LOC, MAL, JOINT, INS, FACT

The preverb orientations can be divided into three wider categories, where categories A to C are small systems with up to four distinctions, while categories D and E are mid-range systems that typically express specific orientations where the smaller systems have clear parallels with the Nakh-Dagestanian local case systems. The last category includes all the largest systems in category F, which can be distinguished from all other systems by the presence of a proximal/distal distinction and wide range of highly specific orientations. These specific orientations should possibly rather be analysed as noun incorporation.

Table 6.9: Tentative summary of all orientational preverb systems expressed by affixation in the Caucasus

Orientational preverb systems by size, where each system type contains any of the listed functions.	
A	1: IN
B	2: IN, SUPER
C	3-4: IN, SUPER, SUB, PROX, DIST, SUPRA
D	5-7: IN, SUPER, SUB, POST, INTER, AD
E	6-10: IN, SUPER, SUB, PROX, DIST, POST, ANTE, SUPRA, HAND, LATER, etc.
F	14+: IN, SUPER, SUB, PROX, DIST, POST, ANTE, SUPRA, HAND, INTER, LATER, CONT, VERT, APUD, etc.

The directional preverb systems are less complex than the orientational preverb systems, and there seems to be a hierarchical tendency where the smaller systems typically have a distinction between lative/ablative but with the addition of delative, suslative and elative directions. The largest systems are almost all built upon these basic distinctions, while adding translative, forlative, directional and retrolative directions.

Table 6.10: Tentative summary of all directional preverb systems expressed by affixation in the Caucasus

Directional preverb systems by size, where each system type contains any of the listed functions.	
A	2: Lative, Ablative
B	3-4: Lative, Delative, Suslative, Elative, Ablative
C	5: Lative, Delative, Suslative, Elative, Ablative, Forlative, Translative
D	5-7: Lative, Delative, Suslative, Elative, Ablative, Forlative, Translative, Directional, Retrolative
E	7-9: Lative, Delative, Suslative, Elative, Ablative, Forlative, Translative, Directional, Retrolative
F	7-10: Lative, Delative, Suslative, Elative, Ablative, Forlative, Translative, Directional, Retrolative, Circumlative, Terminative

The converb systems present large inventories of specialised converbs in many Nakh-Dagestanian and Northwest Caucasian languages, while the smallest systems are all based on a small group of temporal converbs, the conditional converb and/or a general converb. The larger systems typically include simultaneous, temporal, immediate anterior, anterior, causal and concessive converbs, while more specialised converbs are added in a seemingly hierarchical manner, as certain converb functions are only found in the largest systems, e.g. a perfective/imperfective converb distinction and the comparative, substitutive and immediate posterior converbs.

The summarised converb systems should be further investigated however, as some languages lack a fine-grained description of their respective converbs, which means that the summary presented here is rather a summary of the converbs described in previous literature. The extra-ordinarily rich converb systems of category E deserve even more attention, since they are both found in some Nakh-Dagestanian and Northwest Caucasian languages, which is remarkable.

Table 6.11: Tentative summary of all converb systems expressed by affixation in the Caucasus

Converb systems by size, where each system type contains any of the listed functions.			
A	1-3: Conditional, Simultaneous, Temporal, Immediate Anterior, General		
B	6-8: Conditional	Posterior	
	Simultaneous	Purposive	
	Temporal	Inceptive	
	Immediate Anterior	General	
	Anterior		
	Causal		
	Concessive		
	Terminative		
	Irrealis Conditional		
C	8-12: Conditional	Posterior	
	Simultaneous	Purposive	
	Temporal	Inceptive	
	Immediate Anterior	General	
	Anterior	Manner	
	Causal	Negative	
	Concessive	Equitemporal	
	Terminative	Similative	
	Irrealis Conditional		
D	10-14: Conditional	Posterior	Perfective
	Simultaneous	Purposive	Imperfective
	Temporal	Inceptive	Gradual
	Immediate Anterior	General	Consecutive
	Anterior	Manner	
	Causal	Negative	
	Concessive	Equitemporal	
	Terminative	Similative	
	Irrealis Conditional	Locational	
E	14-19: Conditional	Posterior	Perfective
	Simultaneous	Purposive	Imperfective
	Temporal	Inceptive	Gradual
	Immediate Anterior	General	Consecutive
	Anterior	Manner	Directional
	Causal	Negative	Ablative
	Concessive	Equitemporal	Comparative
	Terminative	Similative	Substitutive
	Irrealis Conditional	Locational	Immediate Posterior

Table 6.12: Tentative summary of the Kartvelian, Nakh-Dagestanian and Northwest Caucasian verbal affixation systems (excl. gender/noun class) in the Caucasus.

T	*A*	*M*	*E*	*S*	*O*	*NEG*	*Val*	*PV OR*	*PV DIR*	*CVB*	*Languages*
B	C	E	D	5	4	No	E	C	F	0	Georgian
B	C	E	D	5	4	Yes	E	E	C	0	Megrelian
C	C	D	D	5	4	Yes	E	F	F	C	Laz
C	C	D	A	6	6	Yes	E	C	F	0	Old Georgian
C	D	C	B	6	6	No	E	C	C	0	Svan
C	C	J	B	2	0	Yes	0/A	E	D/E	E	Chechen, Ingush
C	A	G	B	4	0	No	D	C	D	B	Bats
C	D	G	E	0	0	Yes	0	0	0	E	Avar
C	B	D	B	0	0	Yes	A	0	0	B	Andi
C	A	D	B	0	0	Yes	C	0	0	D	Tindi
B	C	G	A	0	0	Yes	A	0	0	B	Bagvalal
C	B	G	D	0	0	Yes	C	0	0	B	Chamalal
C	B	D	E	0	0	Yes	A	0	0	C	Karata
C	C	H	B	2	0	Yes	A	0	0	E	Akhvakh
C	D	G	B	0	0	Yes	A	0	0	D	Ghodoberi
C	D	E	E	1	0	Yes	C	0	0	E	Tsez
C	C	G	E	0	0	Yes	A	0	0	E	Khwarshi
C	D	H	E	0	0	Yes	D	0	0	D	Hinuq
A	F	F	0	0	0	Yes	C	0	0	E	Beztha
C	C	F	0	2	0	Yes	C	0	0	D	Hunzib
C	C	E	A	0	0	Yes	A	D	D	D	Lezgian
C	D	G	C	0	0	Yes	0	E	A	C	Aghul
C	C	G	0	4	4	Yes	0	D	A	E	Tabasaran
C	C	E	0	6	0	Yes	C	A	D	B	Udi
A	C	D	0	0	0	Yes	0	D	C	B	Rutul
C	B	G	F	3	0	Yes	A	C	E	C	Tsakhur
C	D	G	0	0	0	Yes	0	C	C	C	Kryts
C	C	D	0	0	0	Yes	B	C	C	B	Budukh
C	E	J	D	0	0	Yes	0	0	0	D	Archi
C	C	H	0	0	0	Yes	0	C	A	B	Khinalug
C	C	H	0	6	0	Yes	A	E	B	E	Standard Dargwa
C	C	H	0	0	0	Yes	A	C	C	E	Mehweb
C	D	G	0	6	4	Yes	A	E	B	B	Kubachi
C	C	I	0	6	0	Yes	A	E	C	E	Xaidaq
C	C	I	0	6	4	Yes	C	E	C	C	Itsari Dargwa
C	D	G	B	4	0	Yes	A	E	C	D	Sanzhi Dargwa
C	D	J	C	3	0	Yes	B	0	0	D	Lak
C	G	E	C	6	6	Yes	F	F	F	D	Kabardian
C	E	H	C	6	6	Yes	F	F	F	D	Adyghe
C	G	H	0	6	5	Yes	F	F	F	B	Shapsug Adyghe
C	G	E	0	6	6	Yes	F	E	F	E	Abzakh Adyghe
C	G	H	0	6	6	Yes	F	F	C	D	Ubykh
C	G	H	B	6	6	Yes	F	F	F	E	Abkhaz, Abaza

Table 6.13: Tentative summary of the Indo-European and Turkic verbal affixation systems (excl. gender/noun class) in the Caucasus.

T	*A*	*M*	*E*	*S*	*O*	*NEG*	*Val*	*PV OR*	*PV DIR*	*CVB*	*Languages*
C	F	G	B	6	0	Yes	D	0	0	0/A	Eastern Armenian
C	C	A	0	6	0	Yes	B	0	0	0	Classical Armenian
C	C	A	0	6	0	No	B	A	C	A	Iron Ossetic
C	A	G	0	6	0	Yes	B	A	D	0	Talysh
A	C	D	B	6	0	Yes	A	A	B	A	Tat
C	C	G	B	6	0	Yes	C	B	B	B	Juhuri
C	C	G	0	6	0	Yes	D	0	0	C	N Azerbaijani
C	C	C	0	6	0	Yes	D	0	0	C	S Azerbaijani
C	C	G	B	6	0	Yes	D	0	0	C	Kumyk
C	C	E	0	6	0	Yes	D	0	0	C	Karachay-Balkar
C	C	B	B	6	0	Yes	D	0	0	C	Nogai

The summaries in tables 6.12 and 6.13 of the verbal affixation systems in the Caucasus present a remarkably complex picture in terms of the depth and breadth of these systems, while once more revealing how much interfamilial variation these systems exhibit. Although most closely related languages belong to the same categories, only a few languages have identical systems, which either indicates that there are inconsistencies in the data or that these languages actually do vary on a micro-level. As we could see for the nominal affixation systems, related languages often differ in only one category, which could reflect how morphology changes over time and space, function by function.

The preverb systems and converb systems are worth mentioning, as they form complex affixation systems that share a surprising number of grammatical functions across the various endemic language families of the Caucasus. Even though preverbs and converbs are found to a varying degree in the Indo-European and Turkic languages, they do not reach the same heights as the systems in the three endemic language families. It is also fascinating to see how complex preverb systems are found in all three endemic language families, while converbs are primarily shared between Nakh-Dagestanian and Northwest Caucasian. The only potential Kartvelian converb system is found in Laz and possibly Megrelian, while the Laz converb system aligns with most Turkic systems, it has some unexpected similarities with the smaller Nakh-Dagestanian systems. Both preverbs and converbs should be problematised, as they form categories of grammar which occupy a grey area between morphology and syntax and they might potentially be better analysed as lexical rather than grammatical.

7. Results

In this chapter I will present the results of the analyses of the two data sets. The order will follow the research questions presented in section 1.1, as I start with the morphological results and then present the phonological results. Finally, the interaction between morphology and phonology will be examined in the morphophonological results.

7.1. Morphological results

If the grammatical functions described in chapters 5 and 6 are summarised as binary variables to generate a Principal Component Analysis (PCA) plot, we can see that the Northwest Caucasian and Nakh-Dagestanian languages almost perfectly align according to language family (cf. figure 7.1), while the Kartvelian, Indo-European and Turkic languages form one shared cluster. The Northwest Caucasian languages are positioned furthest apart from the other languages in cluster A. All Nakh-Dagestanian languages except Udi cluster together in the rather tight-knit cluster B. The Nakh languages are found on separate sides of Dagestanian languages, with Chechen and Ingush on one side and Bats is positioned almost in-between cluster B and cluster C. Udi is furthest from the other Nakh-Dagestanian languages, as it is positioned halfway to the Turkic and Indo-European languages in cluster C, which is expected since Udi has historically had close ties with Armenian and it is primarily surrounded by Azerbaijani-speaking communities.

Cluster C comprises all Kartvelian, Turkic and Indo-European languages, with the addition of Udi. This is not surprising from a historical perspective, since all Turkic peoples of the region have been in extensive contact with various Indo-European peoples and vice versa which is particularly true for Azerbaijani as it has supplanted most of the historical Iranian languages of Azerbaijan. Although the Kartvelian languages are somewhat set apart from the Turkic and Indo-European languages, they still belong to the same cluster.

Figure 7.1: Principal Component Analysis (PCA) of the grammatical functions expressed by affixation in all 56 languages. Letters A to C indicate clusters calculated by *k*-means clustering.

A second PCA plot including only the nominal grammatical functions (figure 7.2.), thereby plotting a binary version of table 5.12 in section 5.8, presents rather different results. The languages still mainly cluster according to language family, while the Kartvelian, Turkic and Indo-European languages seem to form a continuum. Cluster A, containing all Northwest Caucasian languages, is set furthest apart. Cluster B again comprises all Nakh-Dagestanian languages except Udi. Chechen and Ingush are still set apart from the other Nakh-Dagestanian languages, while Bats is positioned in the middle of cluster B, which suggests that the nominal morphology in Bats rather aligns with the Dagestanian languages. The position of Khinalug together with Chechen and Ingush is possibly due to their shared lack of proper local case systems, while the fact that Lezgian is also found halfway towards cluster C could potentially be explained by language contact, as these four languages all have been in extensive contact with Turkic languages.

Figure 7.2: Principal Component Analysis (PCA) of the nominal grammatical functions expressed by affixation in all 56 languages. Letters A to C indicate clusters calculated by *k*-means clustering.

Cluster C comprises all Kartvelian, Indo-European and Turkic languages and Udi, which is possibly due to historical factors as these languages have been in contact for millennia. Udi is surprisingly closest to the Kartvelian languages, which could either be due to historical language contact with Georgian or that it is simply positioned between clusters B and C. The position of Classical Armenian is remarkably just as close to Old Georgian as it is to Eastern Armenian, which could be explained by the prolonged language contact between Armenian and Georgian.

A third morphological PCA plot for only the verbal grammatical functions (figure 7.3) demonstrates that the verbal data clearly differ from the nominal data.

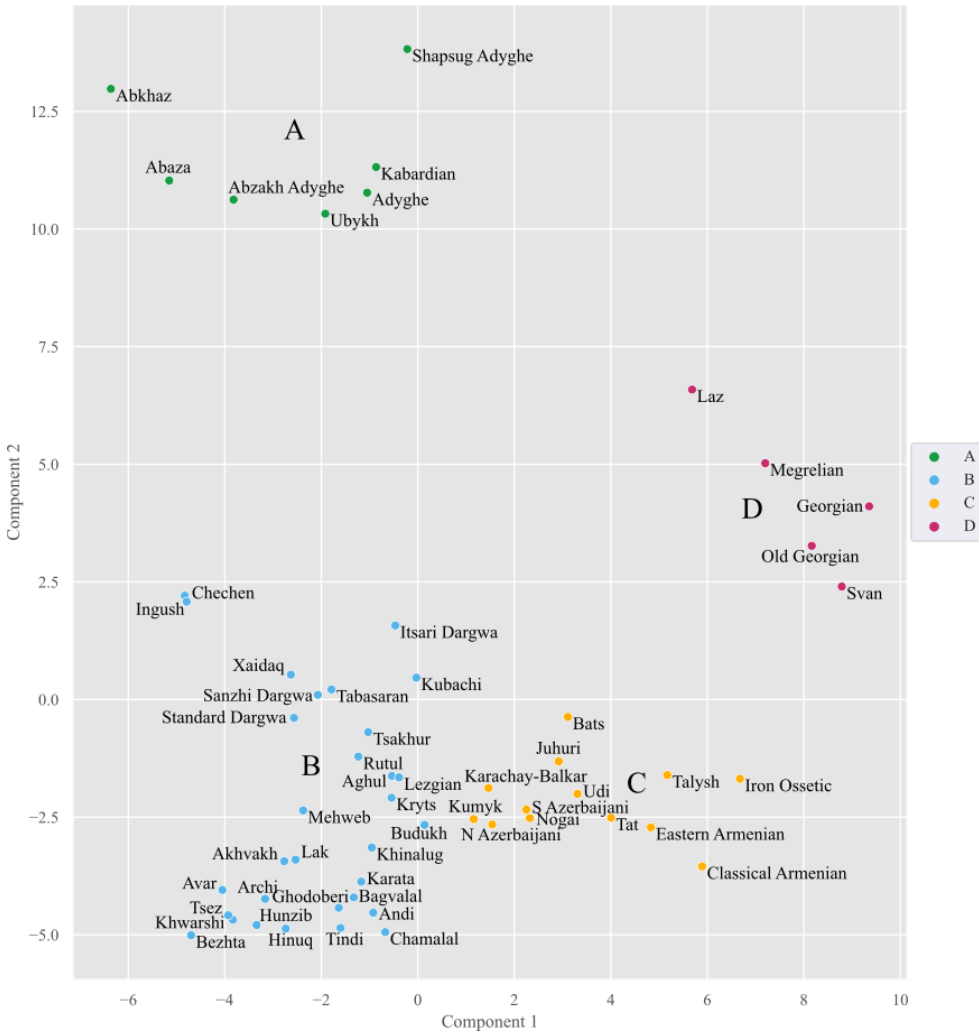


Figure 7.3: Principal Component Analysis (PCA) of the verbal grammatical functions expressed by affixation in all 56 languages. Letters A to D indicate clusters calculated by *k*-means clustering.

The results from the plotted verbal functions (figure 7.3) aligns even more according to language family, as more clusters can be distinguished. The Northwest Caucasian languages still form their own cluster A. The Nakh-Dagestanian languages are primarily found in cluster B, whereas both Bats and Udi cluster with the Turkic languages in cluster C. The Nakh-Dagestanian languages seem to form a continuum within cluster B, from the Avar-Andic-Tsezic languages in the bottom left corner to the Dargic and Lezgic languages closer to the centre. The position of most Lezgic languages close to the Turkic languages further demonstrates the effect of language contact on morphological variation. Chechen and Ingush are again set apart from the rest of cluster B, suggesting that they potentially form a separate cluster.

Cluster C contains all Indo-European and Turkic languages, plus the additions of Bats and Udi. It is unsurprising from a geographical and historical perspective to find Udi clustering with the Turkic languages, but it is somewhat unexpected to find Bats in the same cluster. The final cluster D contains all Kartvelian languages, which clearly sets them apart from the Turkic and Indo-European languages. Figures 7.2 and 7.3 demonstrate how nominal and verbal morphology does not have to change in parallel, as languages such as Bats and Udi end up clustering with different language families. It is remarkable to see how certain clusters are stable both with regard to nominal and verbal morphology, e.g. the Northwest Caucasian and Nakh-Dagestanian clusters.

The first research question asked whether the affix inventories of the languages of the three endemic Caucasian language families display sufficient morphological similarities to support the notion of a Caucasian Sprachbund. Based on the conclusion above and the clear results of the functional PCAs (7.1, 7.2 and 7.3), it seems difficult to claim that the three endemic language families of the Caucasus form a morphological Sprachbund. Although there are evident similarities between the three families, most of them are restricted to similarities between only two of the three language families. Chirikba's claim that the Northwest Caucasian and Nakh-Dagestanian languages form an entity is even more questionable from a morphological perspective, as they share surprisingly few functions apart from common TAM functions, spatial preverbs and converbs. This claim is further weakened by the presence of numerous shared morphological functions between the Nakh-Dagestanian languages, the Northwest Caucasian languages and the Turkic languages, which suggests that these similarities are likely not specific to the Nakh-Dagestanian and Northwest Caucasian languages. However, it is difficult to explain the presence of Nakh-Dagestanian-like noun class systems in Abkhaz and Abaza, as the difficulties of reconstructing morphology in small families such as Northwest Caucasian become apparent.

Although the Northwest Caucasian languages and the Kartvelian languages seem to share numerous grammatical functions, they mainly share the instrumental and adverbial cases, some TAM functions, person-marking of subjects and objects, causatives, applicatives and spatial preverbs. Some of these shared features are somewhat misleading, as the person-marking in Northwest Caucasian is clearly

more complex than in Kartvelian and the specific applicative functions are generally not shared between the language families, which is indicated by the functional PCAs above. The grammatical functions of the affix inventories of the three endemic language families can therefore not support the notion of a Caucasian Sprachbund.

The second morphological research question asked whether the affixation patterns of the Turkic and Indo-European languages spoken in the Caucasus exhibit morphological similarities with any of the endemic language families of the Caucasus. The results suggest that the grammatical functions in the Turkic and Indo-European languages are most similar to each other according to the summarised affixation systems in sections 5.8 and 6.12 and the functional PCAs presented above. The Iranian and Turkic languages are the most similar. The verbal affixation of the Turkic languages exhibit functional similarities with numerous languages of southern Dagestan and, which might be due to language contact. The nominal affixation in Classical Armenian also appears to have some functional similarities with the Kartvelian languages, which might be worth further investigation.

The third morphological research question concerned hierarchical patterns in the distribution of grammatical functions expressed by affixation across the five language families of the Caucasus, but it is difficult to draw any definite conclusions. A potential pattern could be observed in sections 5.8 and 6.12 as the summarised nominal and verbal affixation systems suggest that larger affix inventories are built upon grammatical functions also found in the smaller affix inventories. The functional composition of the larger affix inventories does therefore not appear to be random, which seems to be a tendency for almost all grammatical categories included in chapters 5 and 6. However, the results do not generally demonstrate grammatical functions that are found in all language families, while certain functions are only present in the Nakh-Dagestanian, Northwest Caucasian and Kartvelian languages. This warrants further investigation, along with comparison with large affixation inventories outside the Caucasus, as these tendencies could either be superficial or restricted to the Caucasus.

7.2. Phonotactic results

The phonotactic results show a clear preference for affixes in the Caucasus to contain only one consonant and/or one vowel, i.e. the syllables structures C, CV, VC and V, while the only exception is that bi-consonantal affixes are slightly more frequent than mono-consonantal affixes in the Turkic languages (cf. figure 7.4).

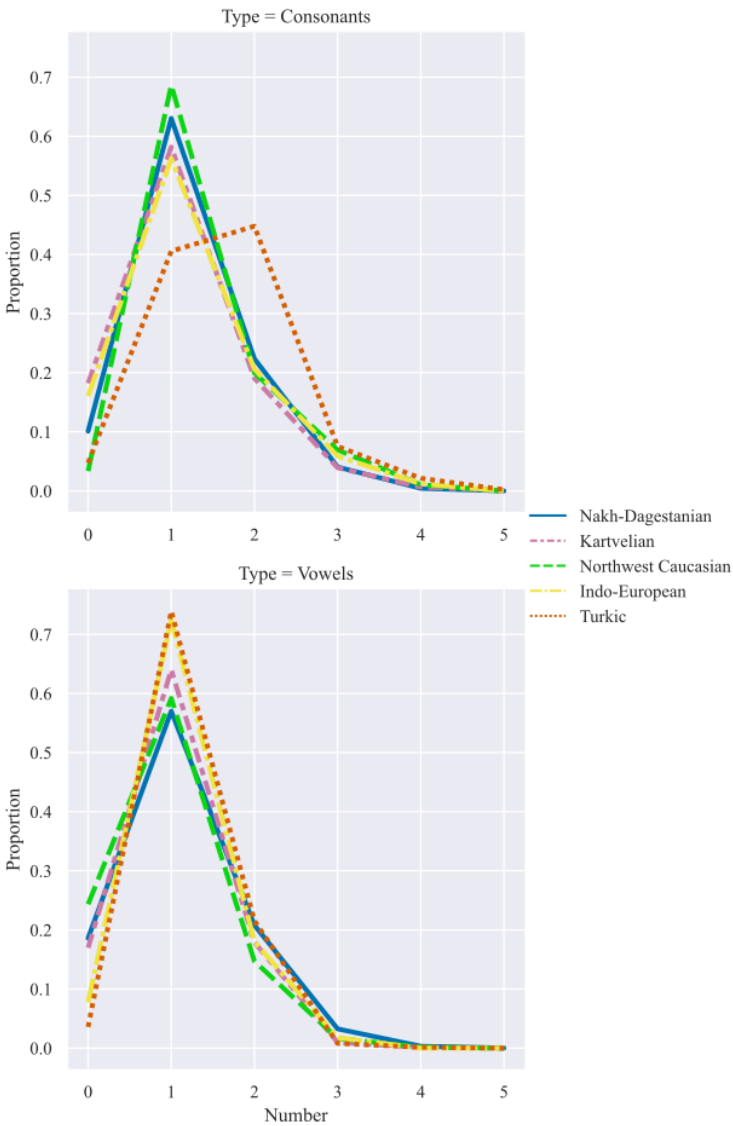


Figure 7.4: Mean distribution of the number of consonants and vowels per affix by language family.

These results could be seen as supporting the assumption that affixes are optimally mono-syllabic, while they only partially support the assumption that affixes are optimally mono-consonantal. This could suggest a tendency where languages with smaller consonant inventories, e.g. Turkic languages, employ CVC affixes to a larger extent as mono-consonantal affixes alone would not be a viable option to maintain complex morphology for these languages, even more so for Turkic languages as they are strictly suffixing.

Table 7.1: Mean proportions of the number of syllables per affix in the Caucasus.

Number of syllables	Mean proportion	*SD*
0	16.67%	9.13%
1	61.07%	10.48%
2	19.55%	7.45%
3	2.47%	2.95%
4	0.23%	0.62%
5	-	-

The results in table 7.1 suggest that 77.7% of the affixes in the Caucasus can be accounted for by the equation of all potential mono-syllabic and mono-consonantal morphemes postulated in section 2.7, while it cannot account for all zero-syllabic affixes, as morphemes of the CC type pose a theoretical issue as they are either ωCL or κCL, cf. section 2.7. This issue is negligible, as they only comprise a mean proportion of 0.6% of all affixes in the data. The results for the specific syllable structures indicate a preference for CV structure in the three endemic language families, but not in Indo-European and Turkic, see figure 7.5.

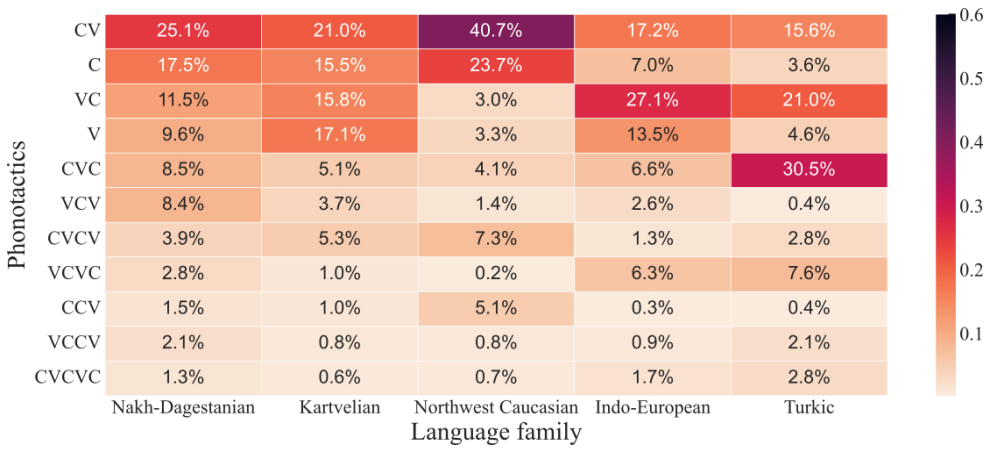


Figure 7.5: Mean proportions of the most common syllable structures of the affix data by language family.

7.3. Place and manner of articulation

7.3.1. Place of articulation

The phonological results for the distribution of places of articulation for all consonants in the affix data show that alveolar/dental consonants constitute the most common place of articulation for all 56 languages. The mean proportion of alveolar/dental consonants is 54.4%, but there are apparent differences between the individual languages as the alveolar/dental consonants range from the maximum 72.7% in Avar to the minimum 32.2% in Adyghe. The Circassian languages and Ubykh stand out as they have the lowest proportions of alveolar/dental consonants of all languages of the data, ranging from 32.2% in Adyghe to 36.9% in Ubykh, whereas all other languages have proportions above 40%. The mean distribution of places of articulation across all languages is presented in table 7.2.

Table 7.2: Mean proportions of the places of articulation in the affixal data.

Place of articulation	**Mean proportion**	***SD***
Alveolar/Dental	54.44%	8.80%
Bilabial	12.64%	6.01%
Palatal	7.28%	4.25%
Velar	7.03%	5.34%
Postalveolar	6.07%	3.79%
Uvular	4.59%	3.80%
Labio-velar	2.88%	2.88%
Labio-dental	1.41%	2.15%
Pharyngeal	1.37%	3.14%
Glottal	1.13%	1.47%
Alveolo-palatal	0.84%	2.31%
Retroflex	0.26%	1.34%
Epiglottal	0.06%	0.30%

Although the bilabial place of articulation is the second most common in 31 of the 56 languages, palatals and velars are the second most common places of articulation in 14 languages. This cannot be explained by genealogical factors, as e.g. Udi (Nakh-Dagestanian) and Megrelian (Kartvelian) stand out with velars comprising 22.9% and 22.3% respectively of all consonants in the affix data. Other outliers are the Nakh languages Bats and Chechen, where the pharyngeals are the second most common place of articulation at 14.9% and 12.8% respectively, and the Tsezic language Bezhta, where the second most common place of articulation is glottals at 7.4%.

The alveolar/dental consonants are similarly the most common place of articulation in all languages in the lexical data, but the mean proportion of alveolars/dentals is noticeably lower at 44.2%. The language with the lowest

proportion of alveolar/dental consonants is still Adyghe for the lexical data at 26.2%. The highest lexical proportion of alveolars/dentals is found in Classical Armenian at 58.7%, which is possibly affected by the relatively small amount of lexical data for Classical Armenian. The four most frequent places of articulation are different though, as they are in descending order alveolars/dentals, velars, uvulars and bilabials.

Table 7.3: Mean proportions of the places of articulation in the lexical data with the mean proportions of the affixal data for comparison.

Place of articulation	Mean proportion of the lexical data	Mean proportion of the affixal data
Alveolar/Dental	44.13%	54.44%
Velar	12.05%	7.03%
Uvular	11.70%	4.59%
Bilabial	11.07%	12.64%
Postalveolar	9.18%	6.07%
Palatal	3.33%	7.28%
Glottal	2.23%	1.13%
Pharyngeal	1.79%	1.37%
Labio-dental	1.67%	1.41%
Labio-velar	1.38%	2.88%
Alveolo-palatal	0.95%	0.84%
Retroflex	0.28%	0.26%
Epiglottal	0.22%	0.06%

Comparing the results from the affixal data (figure 7.6) to the results for place of articulation for to the lexical data (figure 7.7), it becomes apparent that the distribution of places of articulation in the lexical data is largely similar, albeit far from identical to the affix data. The higher proportions of velars and uvulars are easily observed in figure 7.7, which is particularly true for the Nakh-Dagestanian languages. Alveolars/dentals can therefore not be described as overrepresented in the affixal data, since it is simply the most common place of articulation for all these language families, as demonstrated in figures 7.6 and 7.7.

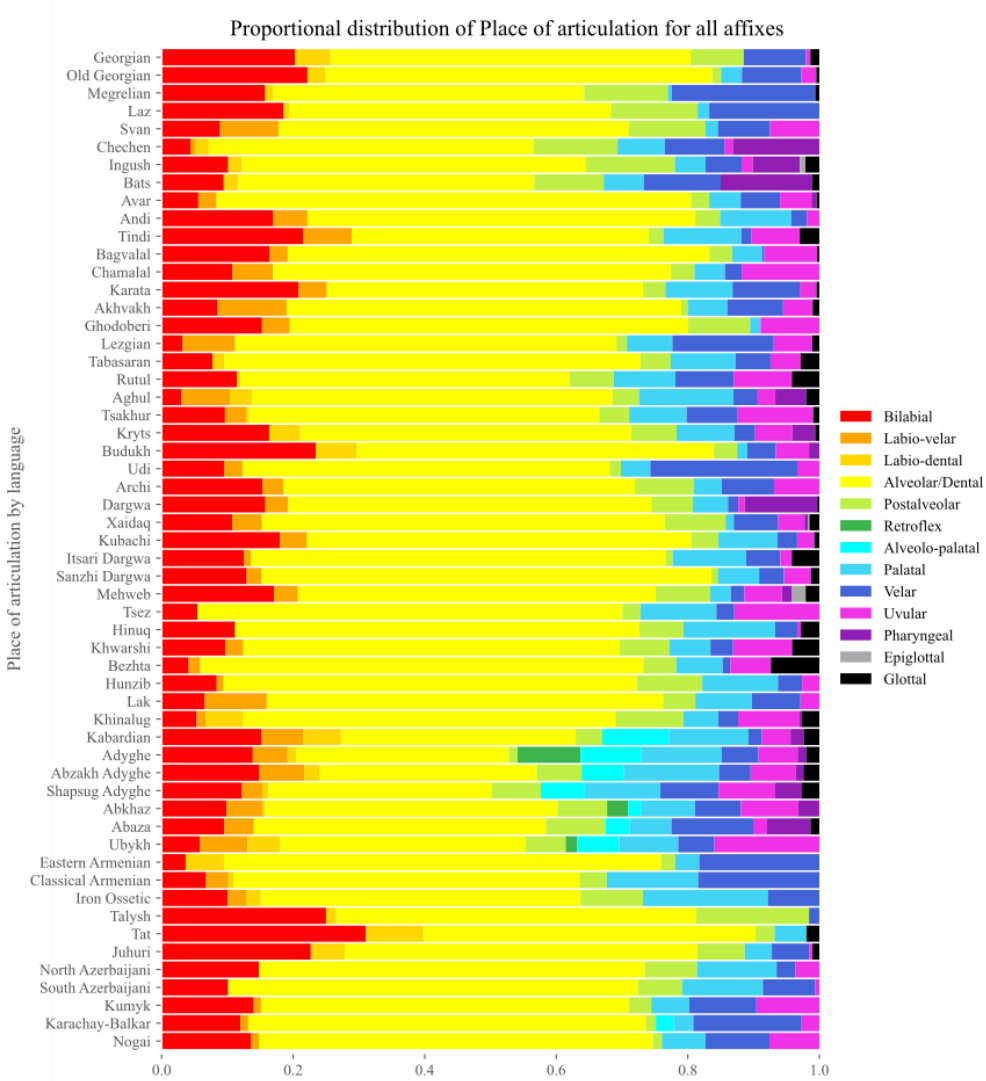


Figure 7.6: Bar plot of the distribution of places of articulation for the affix data by language, ordered by language family. The places of articulation are ordered from the lips (left) backwards to the glottis (right).

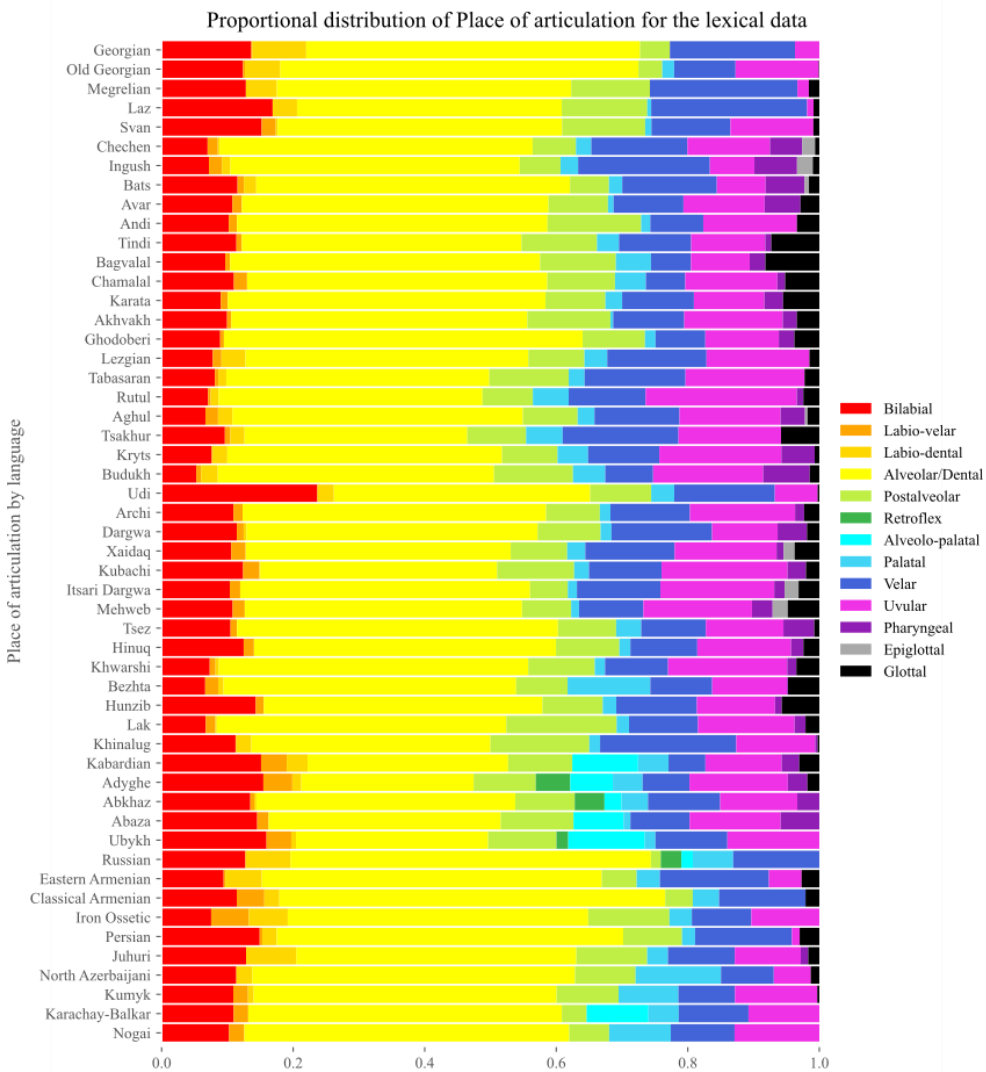


Figure 7.7: Bar plot of the distribution of places of articulation for the lexical data by language, ordered by language family. The places of articulation are ordered from the lips (left) backwards to the glottis (right).

Figures 7.8 and 7.9 below demonstrate that, although there are differences in the mean proportions of the places of articulation between the language families, the most common place of articulation is alveolar/dental, with higher mean proportions of alveolars/dentals in the affixal data for all language families.

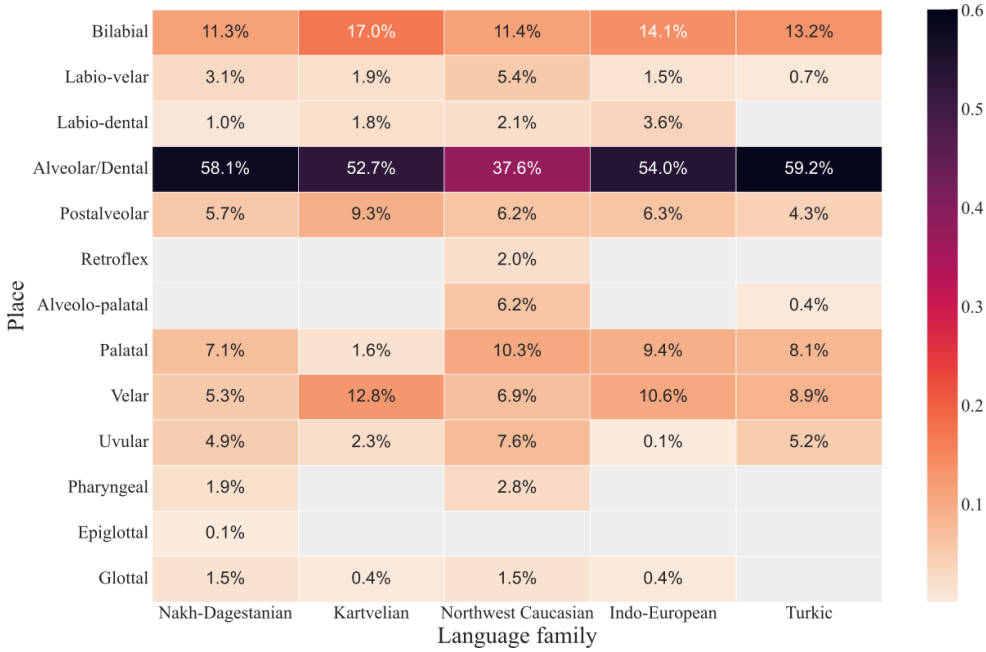


Figure 7.8: Heat map of the mean proportions of the places of articulation in the affix data by language family.

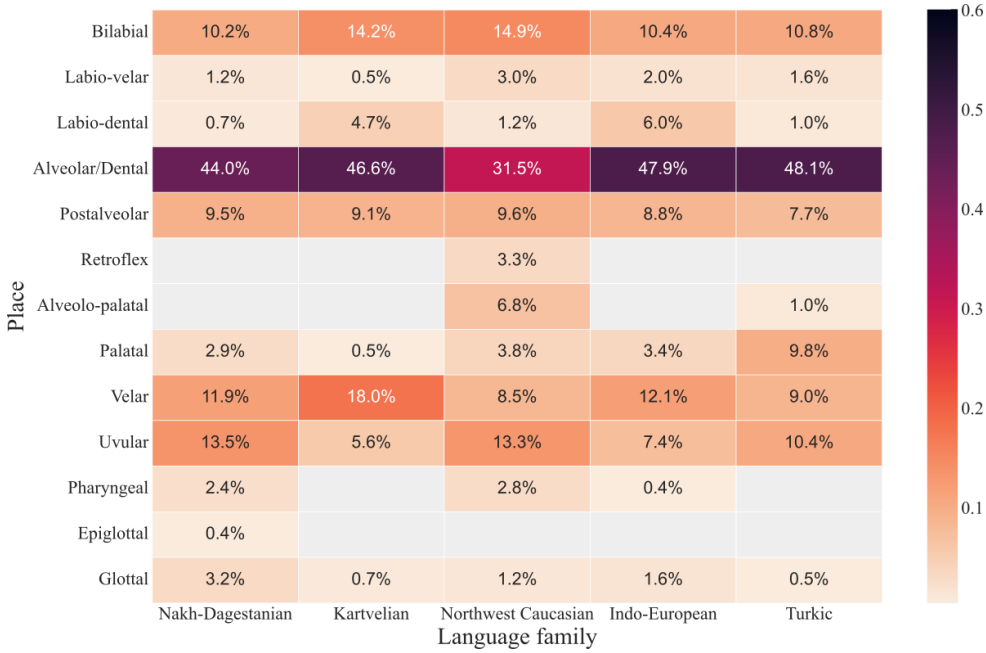


Figure 7.9: Heat map of the mean proportions of the places of articulation in the lexical data by language family.

The most common places of articulation, i.e. alveolars/dentals, bilabials, palatals and velars comprise more than 80% (90.2% in Turkic) in all families except the Northwest Caucasian languages, where these four places only comprise 66.0%. The distribution in the Northwest Caucasian languages differs from the other language families, which cannot be explained by the number of places of articulation alone, as Nakh-Dagestanian has only one place of articulation less than Northwest Caucasian, i.e. 11 vs. 12.

Treating the affixal data and the lexical data for the places of articulation as two paired dependent variables, the differences between these two data sets can be tested statistically. A Shapiro-Wilk test rejects the null hypothesis that the affix and lexical data for place of articulation are drawn from a normal distribution ($W = .622$, $p < .001$ for the affixal data and $W = .720$, $p < .001$ for the lexical data). Therefore a non-parametric two-sided Wilcoxon signed-rank test can be used to compare the mean proportions of each language between the affixal and lexical data. The results indicate that the median difference between the affix data ($Md = 0.047$) and the lexical data ($Md = 0.071$) is not significant, as $z = -1.69$, $p = .092$, $r < -0.1$. Since the null hypothesis cannot be rejected, there is not a significant difference in the distributions of places of articulation between the affixal and lexical data sets.

7.3.2. Manner of articulation

The phonological results for the distribution of manners of articulation for all consonants in the affix data do not present any manner that is generally the most common. The distribution of manners in the affixal data evidently varies to a much larger degree between the languages than for the distribution of places, cf. figure 7.6. The stops/plosives have the highest mean proportion at 25.1%, but stops are only the most common manner of articulation in 28 of the 56 languages. The distribution of stops/plosives varies considerably from the maximum in Kubachi at 51.7% to the minimum 3.9% in Udi.

The second most common manner based on the mean proportion is fricatives, and they are only the most common manner of articulation in 12 languages, ranging from the maximum 36.5% in Shapsug Adyghe to 4.8% in Andi. The fricatives are the most common manner of articulation in the affix data for all Northwest Caucasian languages. The third most common manner of articulation based on mean proportion is nasals, which is the most common manner of articulation in 13 languages. The distribution of nasals ranges from the maximum 46.9% in Talysh to 4.5% in Bezhta.

Table 7.4: Mean proportions of the manners of articulation in the affixal data.

Manner of articulation	Mean proportion	*SD*
Stop/Plosive	25.21%	8.20%
Nasal	18.28%	9.64%
Fricative	18.23%	7.66%
Approximant	10.14%	5.73%
Tap/Trill	9.82%	5.31%
Lateral approximant	7.92%	5.56%
Affricate	3.77%	4.21%
Ejective	2.19%	3.05%
Lateral fricative	1.51%	2.54%
Ejective affricate	1.27%	1.65%
Lateral affricate	0.77%	1.97%
Lateral ejective affricate	0.60%	1.26%
Ejective fricative	0.22%	0.87%
Lateral ejective fricative	0.07%	0.27%

A comparison between the mean proportions of manners of articulations per language family shows even greater differences between the manners than for the places of articulation. The stops/plosives, fricatives and nasals are generally the most common except for the Northwest Caucasian languages, as they constitute more than 50% regardless of language family. This is particularly true for the Kartvelian, Indo-European and Turkic languages, where the stops/plosives, fricatives and nasals comprise more than 70% of the affixal data. The different patterns for Nakh-Dagestan and Northwest Caucasian could possibly be explained by the extra manners of articulation. Ejectives are completely absent in the Indo-European and Turkic affixes of the Caucasus, while all five ejective manners compromise 11.7% of the Northwest Caucasian affix data, 5.8% of the Kartvelian affix data, and 4.1% of the Nakh-Dagestanian affix data.

Comparing the results from the affixal data to the results for manner of articulation for to the lexical data (figure 7.10), the lexical data appear to demonstrate less variation while also presenting noticeable differences between the various languages and language families.

The stops/plosives, fricatives and nasals have the highest mean proportion for the lexical data, whereas the mean proportion is higher for fricatives and lower for nasals in the lexical data. The lexical data differ from the affixal data, as stops/plosives and fricatives are also the two most common places of articulations in all individual languages except Andi and Classical Armenian. The higher proportion of ejectives, affricates and ejective affricates and the lower proportion of approximants are other noticeable differences, cf. table 7.5.

Table 7.5: Mean proportions of the places of articulation in the lexical data with the mean proportions of the affixal data for comparison.

Manner of articulation	Mean proportion of the lexical data	Mean proportion of the affixal data
Stop/Plosive	24.85%	25.21%
Fricative	24.10%	18.23%
Nasal	10.30%	18.28%
Tap/Trill	9.68%	9.82%
Ejective	6.96%	2.19%
Lateral approximant	6.80%	7.92%
Affricate	6.32%	3.77%
Approximant	4.77%	10.14%
Ejective affricate	3.89%	1.27%
Lateral fricative	0.93%	1.51%
Lateral ejective affricate	0.58%	0.60%
Lateral affricate	0.58%	0.77%
Ejective fricative	0.21%	0.22%
Lateral ejective fricative	0.02%	0.07%

A comparison between the mean proportions of manners of articulations per language family, cf. figures 7.10 and 7.11, shows that there are not shared tendencies similar to the ones found for the places of articulation, as the most common manners of articulation in the affixal data differ between the five language families. However, Kartvelian, Indo-European and Turkic do share a pattern, since stops and nasals are the most common manners of articulation in the affixal data as these manners constitute more than 50% of all manners. Stops, fricatives and nasals are the most common manners in the affixal data for all language families except Northwest Caucasian, which again stands out from the other language families.

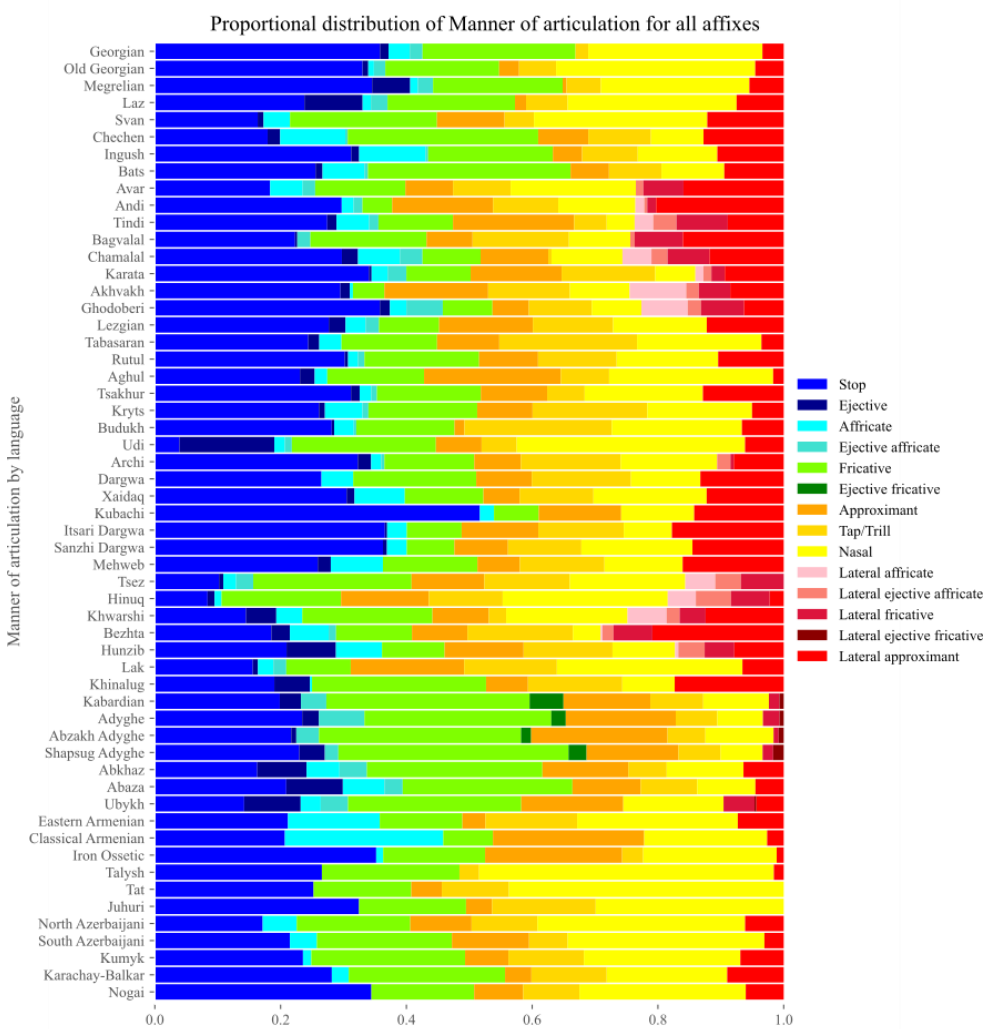


Figure 7.10: Bar plot of the distribution of the manners of articulation for the affix data by language, ordered by language family.

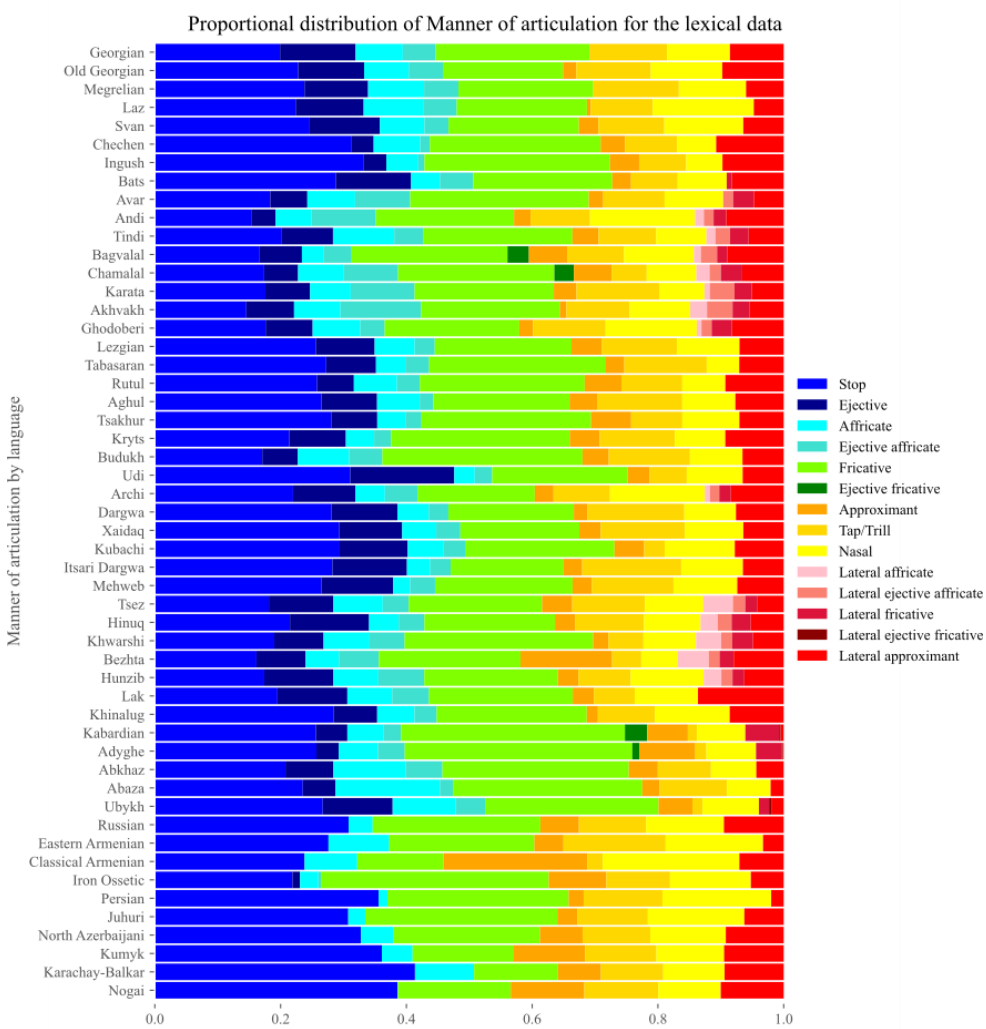


Figure 7.11: Bar plot of the distribution of the manners of articulation for the lexical data by language, ordered by language family.

The lexical data in figure 7.11 show a clearer tendency, as stops and fricatives are the most common manners of articulation in all five language families. The most apparent differences between the lexical and affixal data are connected to nasals, as nasals are more frequent in the affixes of all language families. There also appears to be a generally higher frequency of approximants in the affixal data, but a higher frequency of ejectives and affricates in the lexical data. It is interesting to observe that, although ejectives are absent in the Indo-European affixes, there is a small proportion of ejectives in the Iron Ossetic lexical data.

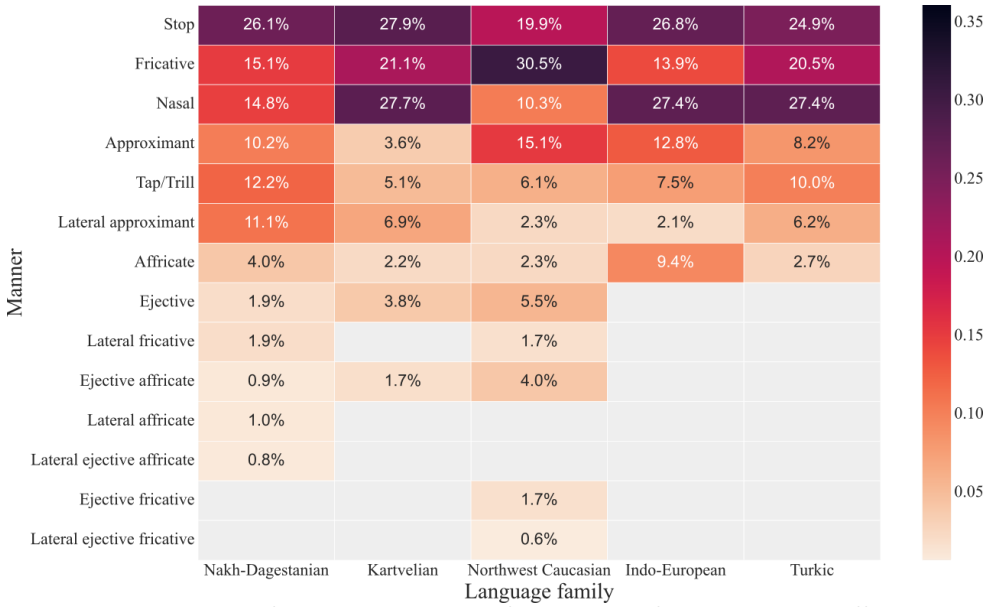


Figure 7.12: Heat map of the mean proportions of the manners of articulation in the affix data by language family.

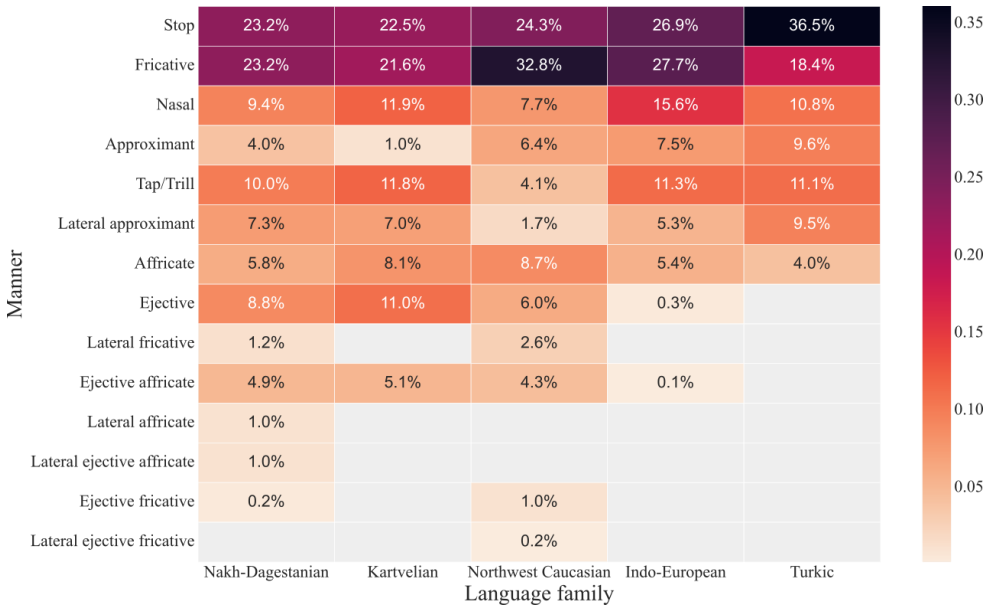


Figure 7.13: Heat map of the mean proportions of the manners of articulation in the lexical data by language family.

Testing the difference between the distributions of manner of articulation in the affixal and lexical data, I again treat these as two paired dependent variables. A Shapiro-Wilk test rejects the null hypothesis that the affix and lexical data for manner of articulation are drawn from a normal distribution ($W = .792$, $p < .001$ for the affix data and $W = .813$, $p < .001$ for the lexical data), as we could see for the places of articulation. A two-sided Wilcoxon signed-rank test indicates that the median difference between the affixal data ($Md = 0.078$) and the lexical data ($Md = 0.078$) is clearly not significant, as $z = -1.08$, $p = .276$, $r < -0.1$. Since the null hypothesis cannot be rejected, there is no significant difference in the distributions of manners of articulation between the affixal and lexical data sets.

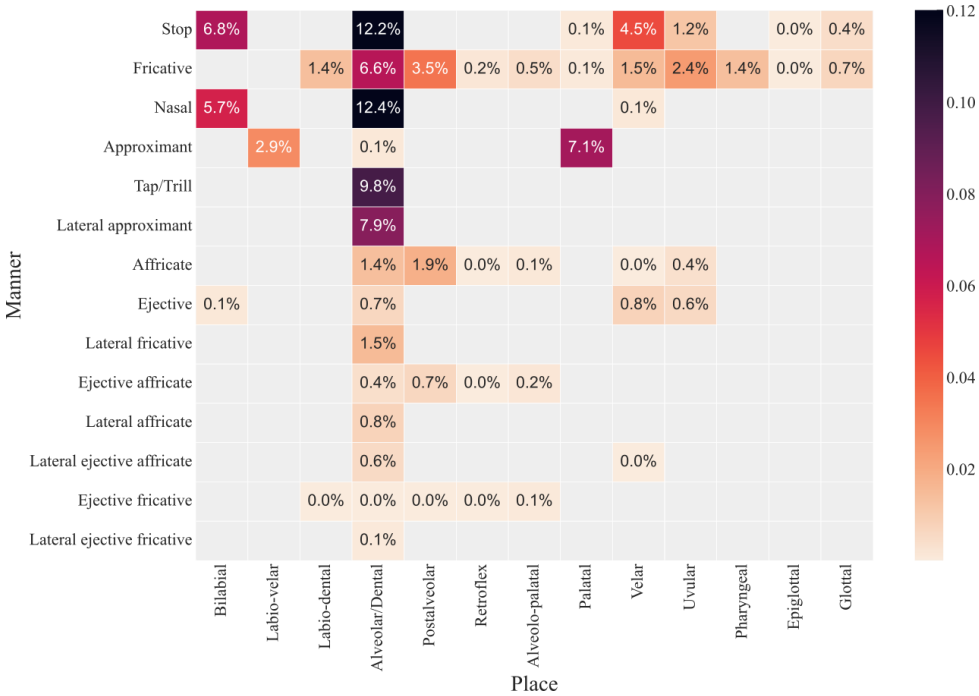
7.3.3. Combining place and manner of articulation

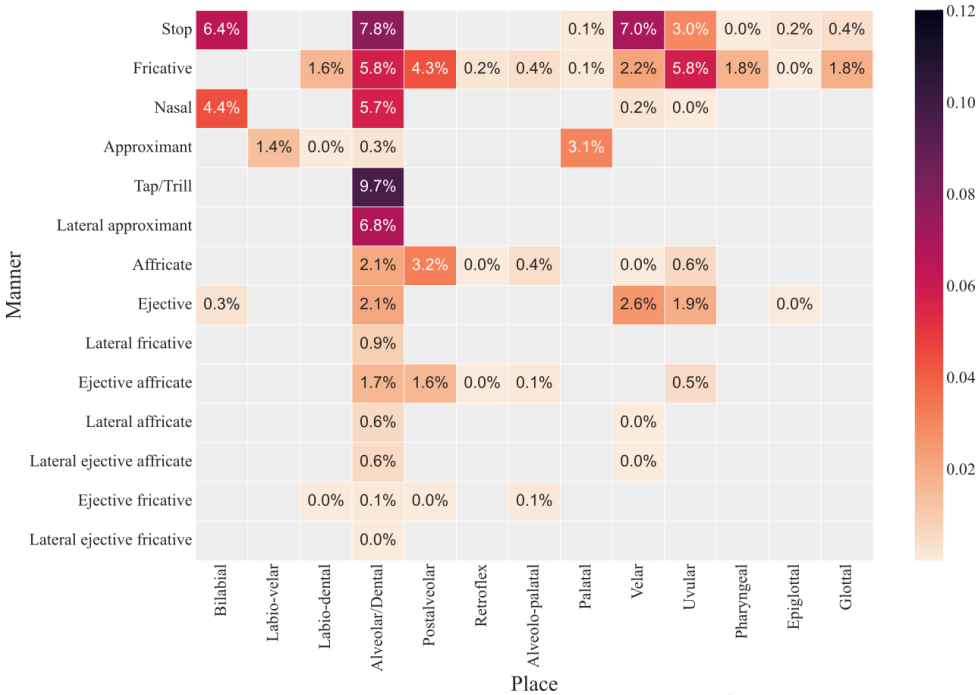
The distribution of the combinations of place and manner of articulation for the affix data in figure 7.12 reiterates the previous results. The stops/plosives, fricatives and nasals are the most common manners in the affix data, while it also becomes apparent that the places of articulation are not evenly distributed across the manners.

Figure 7.12 demonstrates that the alveolar/dental place of articulation is distributed across all manners, constituting the four most common combinations of place and manner, i.e. alveolar/dental nasals, stops, tap/trills and lateral approximants. These four most common combinations of place and manner of articulation comprise almost half of the consonants in the affix data at 42.3%.

Table 7.6: Mean proportions of the ten most frequent combinations of place and manner of articulation in the affixal data.

Combination of Place and Manner	**Mean proportion**	***SD***
Alveolar/Dental Nasal	12.44%	7.08%
Alveolar/Dental Stop	12.15%	6.60%
Alveolar/Dental Tap/Trill	9.82%	5.31%
Alveolar/Dental Lateral approximant	7.92%	5.56%
Palatal Approximant	7.13%	4.16%
Bilabial Stop	6.81%	4.80%
Alveolar/Dental Fricative	6.59%	3.89%
Bilabial Nasal	5.69%	4.50%
Velar Stop	4.51%	4.20%
Postalveolar Fricative	3.50%	3.67%





Figures 7.14 (top) and 7.15 (bottom): Heat maps of the distribution of combined places and manners of articulation for the affixal and lexical data.

The distribution of the combinations of place and manner of articulation for the lexical data in figure 7.15 differs from the affixal data in figure 7.14, while these differences are primarily observable for certain combinations. The four most common combinations differ somewhat, as alveolar/dental taps/trills, alveolar/dental stops, velar stops and alveolar/dental lateral approximants are most frequent in the lexical data. The lower proportion of both alveolar/dental and bilabial nasals in the lexical data is worth mentioning, while the proportion of bilabial stops is almost identical. Velars, uvulars, affricates and all ejective manners are more frequent in the lexical data.

Table 7.7: Mean proportions of the ten most frequent combinations of place and manner of articulation in the lexical data.

Combination of Place and Manner	Mean proportion of the lexical data	Mean proportion of the affixal data
Alveolar/Dental Tap/Trill	9.68%	9.82%
Alveolar/Dental Stop	7.77%	12.15%
Velar Stop	7.00%	4.51%
Alveolar/Dental Lateral approximant	6.80%	7.92%
Bilabial Stop	6.41%	6.81%
Alveolar/Dental Fricative	5.79%	6.59%
Uvular Fricative	5.77%	2.45%
Alveolar/Dental Nasal	5.71%	12.44%
Bilabial Nasal	4.36%	5.69%
Postalveolar Fricative	4.33%	3.50%

By using the same approach as for the places and manners of articulation, we can test the difference between the combinations of place and manner statistically, treating the combinations as two paired dependent variables. A Shapiro-Wilk test rejects the null hypothesis that the affixal and lexical data for the combinations of place and manner of articulation are drawn from a normal distribution ($W = .782$, $p < .001$ for the affix data and $W = .913$, $p < .001$ for the lexical data). A subsequent two-sided Wilcoxon signed-rank test indicates that the median difference between the affix data ($Md = 0.019$) and the lexical data ($Md = 0.032$) is significant, as $z = -4.94$, $p < .001$, $r = -0.1$. The results demonstrate a significant difference in the distribution of combinations of place and manner, while the effect size is small ($r = -0.1$).

These results can therefore answer the fourth research question, as there are significant phonological differences between affixes and lexical stems in the languages of the Caucasus, but the effect size is small. This could suggest that the interaction of place and manner is significant for differentiating affixes from lexical stems, while place or manner alone are not significant factors. It is therefore relevant to investigate each place and manner separately, to see whether the effect size and significance differ between the various combinations of place and manner. Table 7.8 compares the proportions of affixal and lexical combinations of place and

manner for each language, thereby testing parts of the whole data which yielded the significant results above.

Table 7.8: Wilcoxon signed-rank tests comparing subgroups of the combinations of place and manner of articulation for the affixal and lexical data.

Place	***z*-value**	***p*-value**	**Corrected *p*-value**	**Effect size**
Alveolar/Dental	z = -3.13	p = .002	p = .035	r = -0.1
Bilabial	z = -0.44	p = .663	p = 1.000	r < -0.1
Glottal	z = -3.80	p < .001	p = .003	r = -0.3
Labio-dental	z = -1.03	p = .301	p = 1.000	r = -0.1
Labio-velar	z = -3.38	p < .001	p = .014	r = -0.4
Palatal	z = -4.81	p < .001	p < .00025 (**)	r = -0.5
Pharyngeal	z = -2.62	p = .009	p = .180	r = -0.3
Postalveolar	z = -5.90	p < .001	p < .00005 (***)	r = -0.3
Velar	z = -6.69	p < .001	p < .00005 (***)	r = -0.4
Uvular	z = -8.30	p < .001	p < .00005 (***)	r = -0.5
Manner	***z*-value**	***p*-value**	**Corrected *p*-value**	**Effect size**
Stop	z = -1.52	p = .128	p = 1.000	r < -0.1
Fricative	z = -6.08	p < .001	p < .00005 (***)	r = -0.2
Nasal	z = -4.50	p < .001	p < .00025 (**)	r = -0.3
Approximant	z = -5.87	p < .001	p < .00005 (***)	r = -0.4
Tap/Trill	z = -0.36	p = .721	p = 1.000	r < -0.1
Lateral approximant	z = -1.92	p = .055	p = 1.000	r = -0.2
All lateral manners	z = -2.88	p = .004	p = .080	r = -0.2
Affricate	z = -5.00	p < .001	p < .00005 (***)	r = -0.3
Ejective	z = -8.69	p < .001	p < .00005 (***)	r = -0.5
All ejective manners	z = -11.08	p < .001	p < .00005 (***)	r = -0.5

The results in table 7.8 indicate that only certain combinations of place and manner of articulation differ significantly between the affixal and lexical data. The differences for the postalveolar, velar and uvular combinations are strongly significant with medium to large effect sizes (r = -0.3 to r = -0.5), indicating that these places likely play a significant part in the differentiation between affixes and lexicon in the Caucasus. The palatal combinations also differ significantly, although to a lesser degree but with a large effect size (r = -0.5), suggesting that palatals are generally more common in affixes than in lexical stems. The non-significant difference for the bilabial combinations is in line with the results from figures 7.14 and 7.15.

The strongly significant results for combinations of fricatives, approximants and affricates are interesting, while the effect sizes vary from below medium (r = -0.2) for the fricative combinations to above medium (r = -0.4) for the approximants. The less significant results for nasal combinations are also relevant, while the effect size is only medium (r = -0.3). The results for the taps/trills and stops indicate that there is no distributional difference between the affixal and lexical data for these combinations, which is supported by the results from figures 7.14 and 7.15. Finally,

the most interesting results are the strongly significant result and large effect size (r = -0.5) for ejectives, as it suggests that the widespread presence of ejectives in the Caucasus potentially facilitates the differentiation between affixes and lexicon, which is further supported by the even stronger significant results for all ejective manners.

Testing the ten most common combinations of place and manner of articulation, we can observe a similar tendency, while it reveals differences between certain combinations. The non-significant result for bilabial nasals is interesting, as bilabial nasals are widely used in affix formation, but the result is supported by the overall non-significant results for bilabial combinations. In contrast, the strongly significant result and large effect size for alveolar/dental nasals indicate that the proportions of [n] differ significantly between affixes and lexicon in the Caucasus.

Table 7.9: Wilcoxon signed-rank tests comparing subgroups of the ten most common combinations of place and manner of articulation for the affix and lexical data.

Place and Manner	*z*-value	*p*-value	Corrected *p*-value	Effect size
Alveolar/Dental Nasal	z = -5.38	p < .001	p < .0001 (***)	r = -0.5
Alveolar/Dental Stop	z = -3.84	p < .001	p < .005 (*)	r = -0.4
Alveolar/Dental Tap/Trill	z = -0.36	p = .721	p = 1.000	r < -0.1
Alveolar/Dental Lateral appr.	z = -1.92	p = .055	p = .550	r = -0.2
Alveolar/Dental Fricative	z = -1.84	p = .065	p = .650	r = -0.2
Bilabial Stop	z = -0.29	p = .772	p = 1.000	r < -0.1
Bilabial Nasal	z = -0.88	p = .380	p = 1.000	r = -0.1
Palatal Approximant	z = -5.00	p < .001	p < .0001 (***)	r = -0.5
Postalveolar Fricative	z = -3.06	p = .002	p = .022	r = -0.3
Velar Stop	z = -3.58	p < .001	p < .005 (*)	r = -0.4

The weakly significant results for alveolar/dental and velar stops indicate that the combinations must be considered separately, as the previous tests indicate that place or manner alone cannot predict whether the combinations differ significantly between the affixal and lexical data. This is particularly true for the alveolar/dental fricatives, as the results for all combinations of fricatives are strongly significant while the alveolar/dental and postalveolar fricatives do not differ significantly, which is rather remarkable since particularly [s] is frequently used in affix formation. The strongly significant results and high effect size of the palatal approximants are remarkable, but they reiterate the results from table 7.8.

The fifth research question asked whether there are significant differences in the distribution of certain places and manners of articulation between affixes and lexicon in the Caucasus, and the results indicate that there are significant differences. The interaction of place and manner must be taken into account, as sections 7.3.1 and 7.3.2 demonstrate that neither place nor manner alone differ significantly. The results instead suggest that certain combinations of place and manner differ significantly, which would mean that only certain combinations are more common in the affixes and lexical stems respectively. These differences

cannot be explained by the general frequency of these consonants, which could suggest that certain combinations of place and manner play a more important role in affixation and morphology, potentially helping speakers to differentiate between morphology and lexicon in these languages. The results are somewhat inconclusive for certain combinations of place and manner, which makes it relevant to add the variable of voicing to better understand these phenomena, cf. section 7.4.1.

7.4. Voicing

The last variable of the trivariate consonant model is voicing, so it is therefore relevant to investigate how voicing is distributed between the affixal and lexical data. The results show that voiced consonants are more frequent than voiceless consonants in the affixes of all five language families of the Caucasus. Although voiced consonants are also more common in the lexical data, the language families differ as voiceless consonants are more common in the Northwest Caucasian lexical data, while the difference in mean proportion between voiced and voiceless consonants in the Nakh-Dagestanian lexical data is negligible. The Turkic languages stand out as they have the largest difference in voicing in their affixes, which is likely explained by the lack of prefixes in the Turkic data, cf. section 7.7. This is further supported by the positive correlation between mean proportion of voiceless consonants in affixes and the mean proportion of prefixes, as the Pearson correlation coefficient for these variables is $r(55) = .45$ $p < .001$, which explains the relatively high proportions of voiceless consonants in both the Northwest Caucasian and Kartvelian affixes.

Table 7.10: Mean proportions of voiced and voiceless consonants in the affixal and lexical data.

Voicing	μ	ND	Kartvelian	NWC	IE	Turkic
Voiced, affixes	69.47%	71.38%	63.15%	56.43%	70.85%	79.76%
Voiceless, affixes	30.53%	28.62%	36.85%	43.57%	29.15%	20.24%
Voiced, lexicon	52.63%	50.03%	55.81%	46.80%	64.10%	59.62%
Voiceless, lexicon	47.37%	49.97%	44.19%	53.20%	35.90%	40.38%

Combining the voicing variable with the places and manners of articulation above to form trivariate consonants makes it possible to test whether there is a significant difference in voicing between the affixal and lexical data, as suggested by the results in table 7.11. The results indicate that there is a moderately significant difference between the mean proportions of voiced consonants in affixes and lexicon, but a highly significant difference for the voiceless consonants. To test whether the results

for the voiceless consonants were primarily driven by the significant results for all ejective manners mentioned in section 7.3.3, an additional test was carried out on the proportions of all non-ejective voiceless consonants. The third test shows a highly significant difference for the proportions of voiceless consonants between affixes and lexicon in the Caucasus, while the effect size is only moderate.

Table 7.11: Wilcoxon signed-rank tests comparing the proportions of voiced and voiceless consonants for the affixal and lexical data.

Voicing	*z*-value	*p*-value	Corrected *p*-value	Effect size
Voiced consonants	$z = -3.28$	$p = .001$	$p < .003$ (**)	$r < -0.1$
Voiceless consonants	$z = -13.72$	$p < .001$	$p < .0003$ (***)	$r = -0.3$
Voiceless consonants (excl. all ejective manners)	$z = -9.29$	$p < .001$	$p < .0003$ (***)	$r = -0.3$

These results could indicate, bearing in mind the distributions of voiced and voiceless consonants presented in table 7.10, that voiceless consonants help to distinguish lexical stems from morphological affixes in these languages, which is particularly true for the ejective consonants. While this would be less of an issue in the Turkic languages as they do not have prefixes, the presence of voiceless consonants potentially facilitates to indicate boundaries between long strings of affixes and new lexical stems. Although the results for the voiced consonants are also significant, it is likely due to the opposite tendency that voiced consonants tend to be more common in affixes, possibly helping speakers to identify affixes as well.

7.4.1. Combining place, manner and voicing

In order to test the interaction between place, manner and voicing, the most common trivariate consonants in the lexical data were compared to the same combinations in the affixal data by means of a two-sided Wilcoxon signed-rank test for each trivariate consonant, cf. table 7.12. This reveals that only certain combinations differ significantly, while the effect size of these significant differences is large ($r > -0.5$), which should be compared to the generally smaller effect sizes of the results for combinations of only place and manner. This indicates that place, manner and voicing should all be considered, as the inclusion of voicing demonstrates that there are differences between voiced and voiceless variants of the same segments. This difference cannot be explained by voicing alone however, as [d] is significantly more common in the affixes while [t] is not and [p] is significantly more common in the lexical stems whereas [b] does not differ significantly, thereby demonstrating that interaction between place, manner and voicing is surprisingly intricate.

Table 7.12: Wilcoxon signed-rank tests comparing the 20 most common trivariate consonants in the lexical data with the affixal data, combining place, manner and voicing.

Place, Manner and Voicing	IPA	*z*-value	*p*-value	Corrected *p*-value	Effect size
Voiced Alv./Dent. Tap/Trill	[ɾ]/[r]	$z = -0.32$	$p = .750$	$p = 1.000$	$r < -0.1$
Voiced Alv./Dent. Lat. Appr.	[l]	$z = -1.86$	$p = .062$	$p = 1.000$	$r = -0.2$
Voiced Alv./Dent. Nasal	[n]	$z = -5.41$	$p < .001$	$p < .00005$ (***)	$r = -0.5$
Voiced Bilabial Stop	[b]	$z = -1.20$	$p = .231$	$p = 1.000$	$r = -0.1$
Voiced Bilabial Nasal	[m]	$z = -0.97$	$p = .334$	$p = 1.000$	$r = -0.1$
Voiceless Velar Stop	[k]	$z = -3.76$	$p < .001$	$p = .003$	$r = -0.4$
Voiceless Alv./Dent. Stop	[t]	$z = -1.29$	$p = .196$	$p = 1.000$	$r = -0.1$
Voiced Alv./Dent. Stop	[d]	$z = -5.34$	$p < .001$	$p < .00005$ (***)	$r = -0.5$
Voiceless Uvular Fricative	[χ]	$z = -5.82$	$p < .001$	$p < .00005$ (***)	$r = -0.6$
Voiceless Alv./Dent. Fricative	[s]	$z = -1.20$	$p = .229$	$p = 1.000$	$r = -0.1$
Voiceless Postalv. Fricative	[ʃ]	$z = -2.96$	$p = .003$	$p = .061$	$r = -0.3$
Voiced Palatal Approximant	[j]	$z = -5.02$	$p < .001$	$p < .00005$ (***)	$r = -0.5$
Voiced Velar Stop	[g]	$z = -1.80$	$p = .073$	$p = 1.000$	$r = -0.2$
Voiceless Velar Ejective	[k’]	$z = -6.04$	$p < .001$	$p < .00005$ (***)	$r = -0.7$
Voiceless Postalv. Affricate	[tʃ]	$z = -3.29$	$p < .001$	$p = .020$	$r = -0.3$
Voiceless Uvular Stop	[q]	$z = -4.02$	$p < .001$	$p = .001$ (*)	$r = -0.5$
Voiced Alv./Dent. Fricative	[z]	$z = -0.12$	$p = .908$	$p = 1.000$	$r < -0.1$
Voiceless Alv./Dent. Ejective	[t’]	$z = -5.42$	$p < .001$	$p < .00005$ (***)	$r = -0.6$
Voiced Uvular Fricative	[ʁ]	$z = -4.24$	$p < .001$	$p < .0005$ (**)	$r = -0.5$
Voiceless Bilabial Stop	[p]	$z = -4.86$	$p < .001$	$p < .00005$ (***)	$r = -0.5$

The non-significant results for voiceless velar stops are somewhat unexpected, as the results from section 7.3.3 and the general results for voicing in table 7.12 could predict that [k] would be significantly less common in affixes, which it is not. This should be compared to the voiceless uvular stop [q], where the difference is only weakly significant, possibly explained by stops being the most common manner in both data sets. These results should be tested against data from languages outside of the Caucasus, as these differences either represent general phonological tendencies or patterns that are specific to the Caucasus.

If the affixal proportions of the combined place, manner and voicing are plotted out per language by means of a Principal Component Analysis (PCA) as seen in figure 7.16, we can see that the phonological results differ from the morphological PCAs.

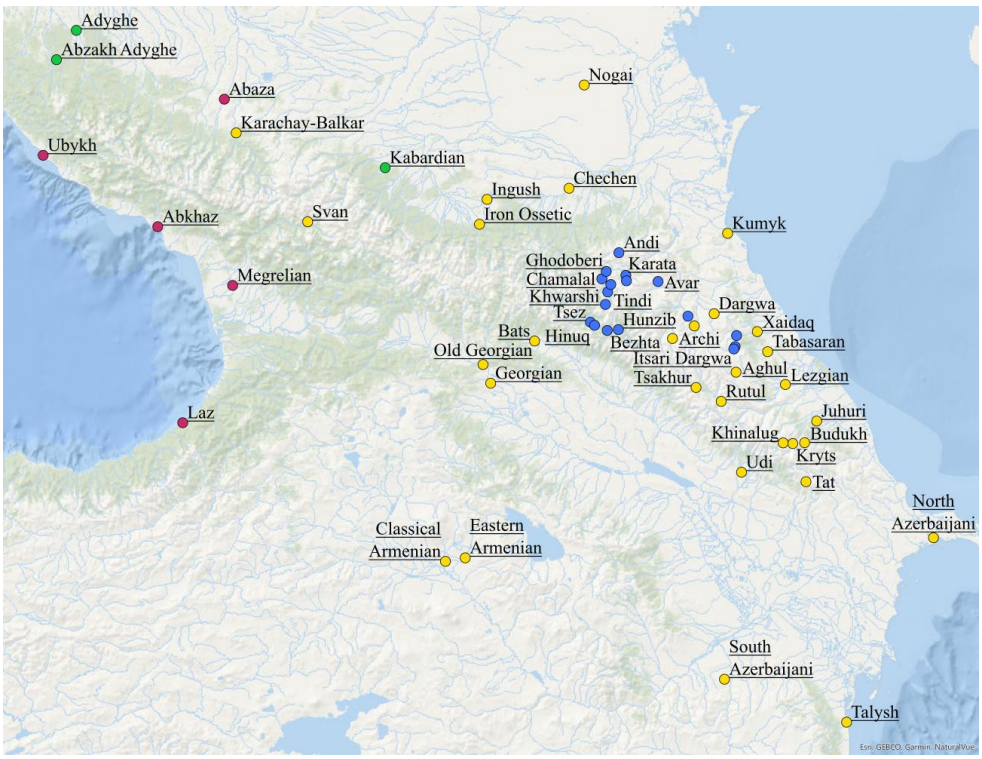
Figure 7.16: Principal Component Analysis (PCA) of the distribution of combined place, manner and voicing per language for the affixal data. Letters A to D indicate clusters calculated by *k*-means clustering.

The Circassian languages are still set furthest apart in cluster A, while the remaining Northwest Caucasian languages belong to the widespread cluster D together with Megrelian and Laz. Neighbouring cluster C contains the remaining Kartvelian and all Turkic, Indo-European, Nakh and Lezgic languages with the addition of Lak, Khinalug and some Dargic languages. Cluster C appears to be largely geographical, as it forms a continuum from Georgia and Armenia in the west through the central Caucasus to southern Dagestan and Azerbaijan. Geographical factors could also explain why Megrelian and Laz clusters with Abkhaz and Abaza, while it cannot explain the position of Udi. The close clustering of Turkic, Nakh, Lezgic and the Iranian languages reiterates previous morphological results, particularly the surprising similarities between Nakh and Turkic, which could potentially be linked to the historical use of Kumyk and Nogai as *linguae francae* among these groups (Forsyth 2013: 211).

Cluster B contains all remaining Nakh-Dagestanian languages, i.e. all Avar-Andic-Tsezic languages and the Dargic languages Kubachi, Itsari Dargwa, Sanzhi

Dargwa and Sanzhi Dargwa. Cluster B therefore mirrors the Avar-Andic-Tsezic clusterings in the morphological PCAs (figure 7.1 and 7.2), which suggest that these languages share numerous morphophonological similarities. The Avar-Andic-Tsezic cluster is an established taxon within the Nakh-Dagestanian family (Ganenkov & Maisak 2020), and these languages are all spoken in the western highlands of Dagestan between the Andi Koysu and Avar Koysu rivers.

These clusterings can generally be explained by geographical and historical factors, as the Avar-Andic-Tsezic cluster largely coincides with the most inaccessible parts of the Sulak river basin, while the remaining languages in cluster B primarily coincide with languages spoken in central Dagestan north of the Lezgic languages included in cluster C. The Nakh-Dagestanian languages of cluster C have been in extensive contact with Turkic and Iranian languages during the last thousand years, which is further enforced by the intense contact between these Turkic and Iranian languages in the region. The longstanding contact and geographical vicinity between the Kartvelian languages and Armenian can also explain cluster C, although there are important phonological differences such as the absence of ejectives in most varieties of Armenian. The position of Udi is surprising however, especially when the close historical ties between Udis and Armenians is considered.



Map 7.1: The mapped out results of the affixal PCA in figure 7.16. Cluster colours: A (green), B (blue), C (gold) and D (cerise).

If the proportions of the combined place, manner and voicing of the lexical data are instead plotted out per language by means of a PCA (figure 7.17), we find that the languages cluster in a similar way as the affixal PCA plot (figure 7.16), although there are noticeable differences.

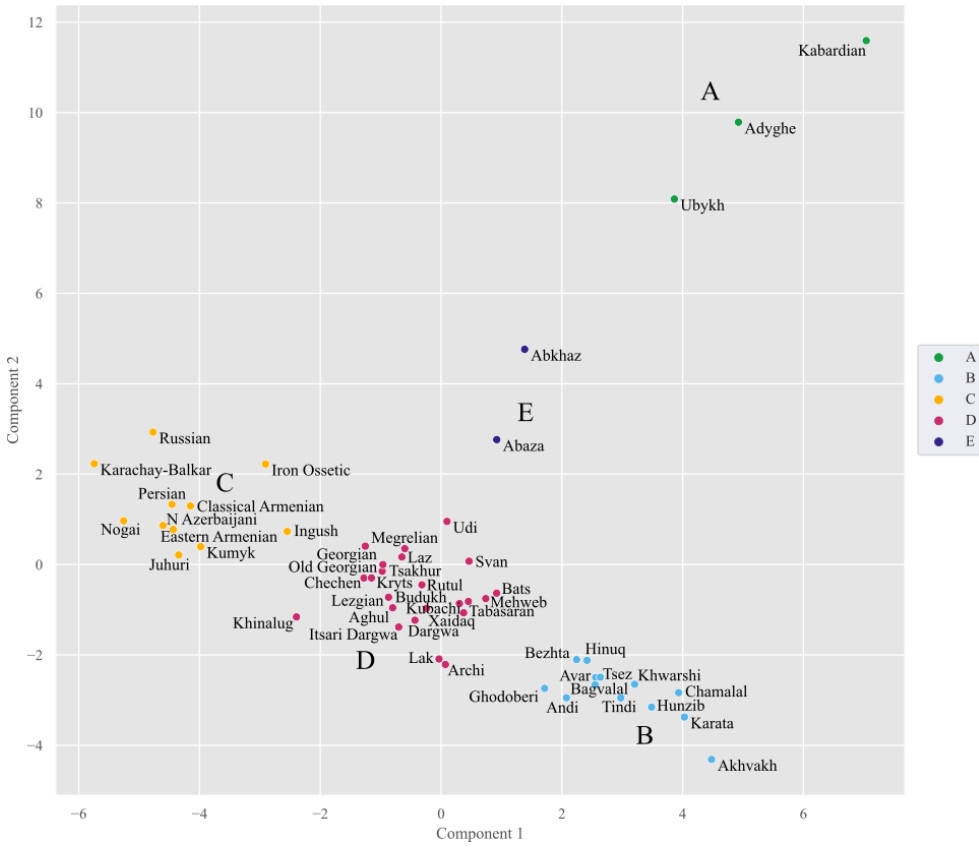
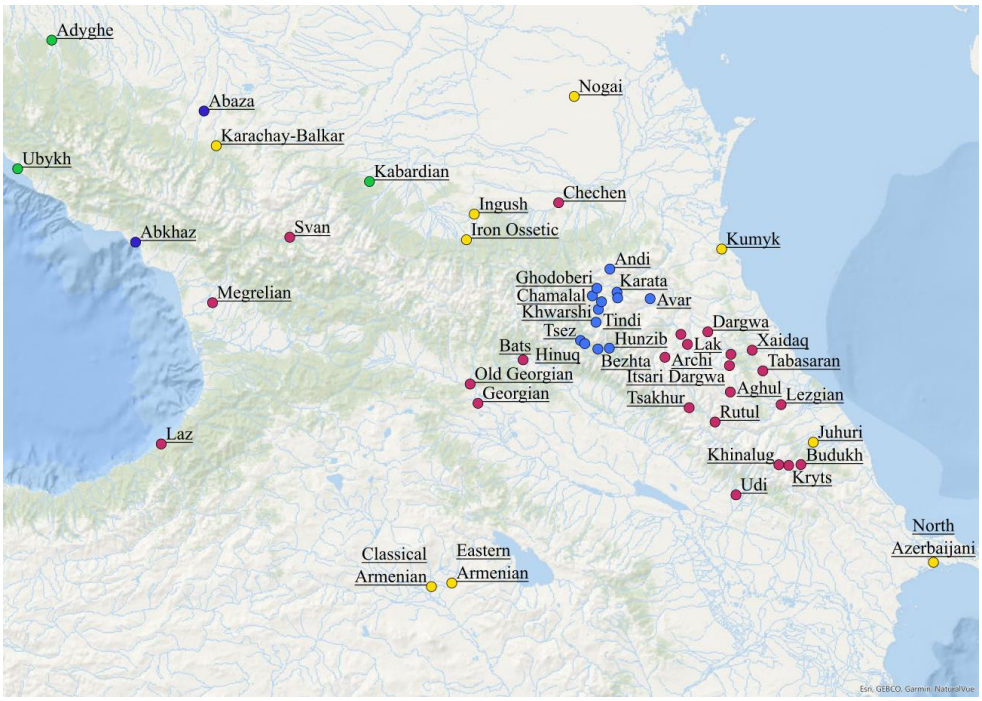


Figure 7.17: Principal Component Analysis (PCA) of the distribution of combined place. manner and voicing per language for the lexical data. Letters A to E indicate clusters calculated by *k*-means clustering.

The Circassian languages and Ubykh still cluster separately into cluster A, while Abkhaz-Abaza form its own intermediate cluster between clusters A and D. Cluster B contains all Avar-Andic-Tsezic languages, reiterating that they form a morphophonological unit. Cluster C groups all Turkic and Indo-European languages, with the addition of Ingush. The distance between Chechen and Ingush is difficult to explain, while its position halfway between Chechen and Iron Ossetic does reflect its geographical location. The last cluster D includes all Kartvelian and all remaining Nakh-Dagestanian languages except Ingush. The position of Udi is highly remarkable for the lexical PCA, as it again clusters closest to the Kartvelian languages.

The lexical PCA (figure 7.17) is not as easily explained as the affixal PCA (figure 7.16), but considering the strong areal effect observed in morphological change in Western Europe (Larsson 2013), it becomes plausible that the phonology of affixation is more likely to be affected by areal factors such as language contact. There is no reason to assume the opposite, that the phonological patterns of the lexicon are less easily affected by language contact, while it could reflect earlier language contact, as particularly the Nakh and Lezgic languages have been in geographical contact with the Kartvelian languages since before the advent of Turkic and Iranian peoples in the Caucasus. This would fail to explain why the Tsezic languages are not clustering with the Kartvelian languages, as they have also been in contact with Georgian for millennia (Comrie & Khalilov 2009: 418). The two PCAs above suggest that the phonological distribution of consonant phonemes are explained by an intricate combination of genealogical and areal factors.



Map 7.2: The mapped out results of the lexical PCA in figure 7.17 (excl. Russian and Persian). Cluster colours: A (green), B (blue), C (gold), D (cerise) and E (indigo).

The phonological PCAs above could indicate that there is indeed a phonological Caucasian Sprachbund, while such a sprachbund would potentially only include Kartvelian, Lezgic, Dargic and possibly Nakh. The Northwest Caucasian and Avar-Andic-Tsezic languages consistently form clusters clearly separated from the other languages, which makes it problematic to claim that they should be included in a

phonological Caucasian Sprachbund. Figure 7.16 potentially reflects phonological tendencies that apply to affixation and morphology in general however, as cluster C groups languages from all language families of the Caucasus except the Northwest Caucasian languages. It would therefore be highly relevant to add languages from outside the Caucasus to see how they would affect the results.

7.5. Consonant segments

Table 7.13 presents the ten most common consonant segments in the affix data by mean proportion in descending order. This reveals a tendency not shown by the results for place and manner of articulation.

Table 7.13: Mean proportions of the ten most common consonant segments in the affixal data.

Segment	Mean proportion	*SD*
[n]	12.36%	7.13%
[r]/[ɾ]	9.94%	5.16%
[d]	8.27%	4.51%
[l]	7.82%	5.48%
[j]	7.11%	4.16%
[b]	5.95%	4.99%
[m]	5.69%	4.50%
[s]	3.36%	2.65%
[w]	2.87%	2.88%
[z]	2.64%	3.20%

The four most common segments are all voiced alveolars and one of these four consonants is the most common consonant segment in 46 of the 56 languages. The alveolar nasal [n] is only the most common consonant in 21 languages, ranging from the maximum 35.9% in Talysh to the minimum 0.7% in Tindi.[255] The four most common consonant segments, i.e. [n], [r]/[ɾ], [d] and [l], form >25% of all consonants in the affixal data (μ = 38.4%) in 47 of the 56 languages. The remaining nine languages are the Northwest Caucasian languages (except Abkhaz), Iron Ossetic, Classical Armenian and Tindi. The situation in Armenian is complicated by the phonemic differentiation between [r] and [ɾ]/[ɹ], which is alleviated by the apparent absence of the segment [r] in both Eastern Armenian and Classical Armenian affixes. The Northwest Caucasian languages stand out as they are highly prefixing, which affects the ratio of these segments (cf. section 7.7), where the

[255] The extremely low percentage in Tindi is likely due to the process of reducing syllable-final alveolar nasals to nasalised vowels.

Circassian languages also have the lowest ratios of alveolar nasals [n] (μ = 5.6%) of all branches regardless of language family.

Table 7.14: The ten most common consonant segments in the affixal and the lexical data per language family.

ND Aff.	ND Lex.	Kart. Aff.	Kart. Lex.	NWC Aff.	NWC Lex.	IE Aff.	IE Lex.	Turk. Aff.	Turk. Lex.
[r]	[r]	[n]	[r]	[j]	[b]	[n]	[r]/[ɾ]	[n]	[r]
[n]	[l]	[m]	[l]	[r]	[r]	[m]	[n]	[r]	[t]
[l]	[n]	[d]	[n]	[z]	[p]/[pʰ]	[r]/[ɾ]	[k]/[kʰ]	[d]	[l]
[d]	[b]	[s]	[m]	[n]	[m]	[j]	[m]	[m]	[q]
[b]	[m]	[ʃ]	[b]	[w]	[n]	[k]/[kʰ]	[t]/[tʰ]	[z]	[j]
[j]	[d]	[l]	[v]	[m]	[χ]	[d]	[s]	[j]	[n]
[m]	[χ]	[tʰ]	[k’]	[d]	[t]/[tʰ]	[b]	[l]	[l]	[b]
[w]	[k]	[g]	[x]	[t]	[s]	[ts]/[tsʰ]	[ʃ]	[s]	[k]
[s]	[j]	[r]	[tʰ]	[p]	[w]	[t]	[d]	[t]	[s]
[tʃ]	[t’]	[kʰ]	[d]	[ʁ]	[d]	[ʃ]	[g]	[g]	[ʃ]

The voiced palatal approximant [j] is the fifth most common consonant segment in the affixal data. It is the most common consonant segment in West Circassian affixes while the maximum is found in Iron Ossetic at 19.0%. The voiced bilabial stop [b] and the voiced bilabial nasal [m] are almost equally distributed in the affixal data, while the distribution for these segments varies considerably between the languages. The maximum of [b] is found in Karata at 19.1% where the ratio of [m] is at 1.7%, while the maximum of [m] is found in Old Georgian at 19.9% where the ratio of [b] is 1.8%. The ratios in Karata and Old Georgian seem to suggest a complementary relationship, where languages either have high ratios of [b] or [m] but not both, which is only contradicted by Tat where the ratio of [b] is 12.6% and [m] is 18.4%.

Lezgian and Aghul are also worth mentioning in this context as they have the lowest combined ratios of [b] and [m] of all 56 languages at 3.2% and 3.0% respectively, where Lezgian completely lacks the segment [b] in nominal and verbal affixes.[256] The segment [b] is also absent in Svan and Eastern Armenian affixes, while [m] is only absent in Avar and Bezhta affixes. The remaining three segments are the voiceless and voiced alveolar fricatives [s]/[z] and the labio-velar approximant [w], where [s] stands out as it is the only voiceless consonant among

[256] This is relevant as the phoneme [b] is a central component in most Nakh-Dagestanian noun class or gender systems, which Lezgian and Aghul also lack, cf. section 5.8.

the ten most common consonant of the affix data. The voiceless alveolar fricative has the highest ratios in the Kartvelian languages, Tsakhur, Eastern and both varieties of Armenian, while the voiced counterpart [z] is much less frequent in these languages as [z] is absent in Kartvelian affixes except for in Georgian and Laz. The lack of voiced alveolar fricatives in affixes is not only a Kartvelian phenomenon, as 21 of the 56 languages completely lack the segment [z] in affixes. The labio-velar approximant is also lacking in 18 languages, which is primarily due to the often complementary distribution of the segments [v] and [w] as some languages lack the segment [w] altogether.

Assuming that the phonemic affixal data and the phonemic lexical data are two paired dependent variables that are not normally distributed, a two-sided Wilcoxon signed-rank test indicates that the median difference between the affix data (Md = 0) and the lexical data (Md = 0.014) is highly significant, as z = -12.61, $p < .001$. A Shapiro-Wilk test clearly rejects the null hypothesis that the phonemic affixal and phonemic lexical data are drawn from a normal distribution, as W = .611, $p = 0$ for the affix data and W = .780, $p = 0$ for the lexical data. This is likely due to the systematic absence of certain segments in the affix data, which also explains why the median is zero for the affix data. If only the segments that occur in both the affix data and lexical data are included, a two-sided Wilcoxon signed-rank test still indicates a significant difference, as z = -5.28, p = < .001.

The general absence of certain consonant segments in the affixal data is relevant, as Bybee (2005) previously investigated whether 'highly marked segments' and 'highly complex segments' are excluded from affix inventories, which was weakly supported by her data. If the full inventories of consonant segments with secondary articulation are compared between the affixal and lexical data, almost 80 segments are only found in the lexical data. This number should be read with caution however, as the affixal data are primarily based on phonemic descriptions while parts of the lexical data are based on a more detailed transcription (particularly the IDS data). Nevertheless, there appears to be a large number of segments that are not present in the affixes of these languages, which further supports Bybee's previously weak results. The most common consonant segments that only occur in the lexical data are quite rare, and many of them could be categorised as highly complex, since most of them are lengthened ejectives, but they also include various labialised, pharyngealised and palatalised segments, which should not be categorised as highly complex per se. The differences are more pronounced if separate languages and/or language families are compared, which warrants further investigation but not in this thesis.

Table 7.15: Mean proportions of the ten most common consonant segments only found in the lexical data.

Segment	Mean proportion	*SD*
[qχ’]	0.24%	0.60%
[qχː’]	0.24%	0.82%
[tsː’]	0.17%	0.60%
[sː’]	0.11%	0.56%
[hʷ]	0.06%	0.16%
[tsʷ’]	0.05%	0.14%
[qː’]	0.05%	0.27%
[ʔ’]	0.04%	0.29%
[kː’]	0.04%	0.13%
[tʃː’]	0.04%	0.12%

7.6. Vowel segments

Figures 7.18 and 7.19 below present the distribution of unrounded and rounded vowels in the affixal data. Unrounded vowels account for almost 80% of all vowels, which is interesting as this can neither be explained by the three most common vowels worldwide, i.e. [i], [u] and [a], or the five most common vowels that also include [e] and [o] (Moran & McCloy 2019), as an even distribution of these vowels would suggest at least 33% rounded vowels. This is even more interesting, as both Turkic and some Nakh-Dagestanian and Indo-European languages make extensive use of front rounded vowels in their morphology. The Northwest Caucasian vowel systems will naturally affect the results however, as most vowel segments in these languages are described as unrounded, although most of these segments have rounded allophones depending on the surrounding consonants (Kuipers 1960: 22-23). To account for this phenomenon in Northwest Caucasian, the vowel distributions for each language family are be presented in chapter 7.8.

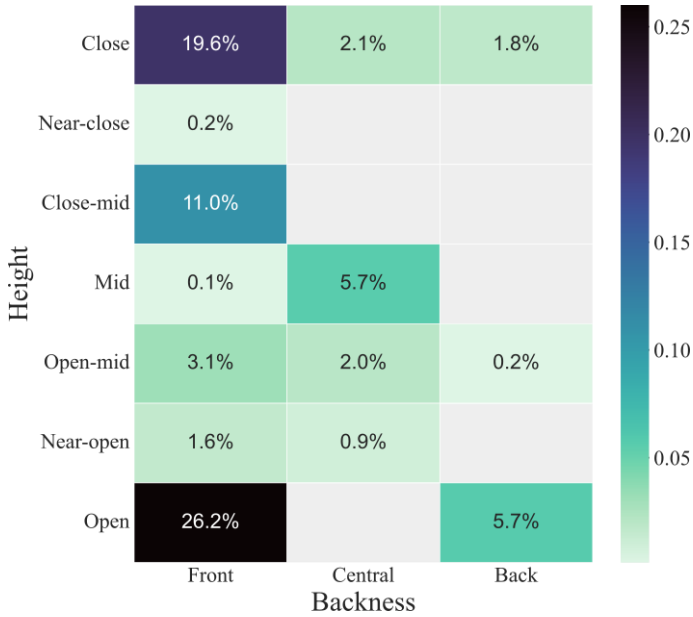


Figure 7.18: Heat map of the proportions of unrounded vowels of all vowels in the affixal data.

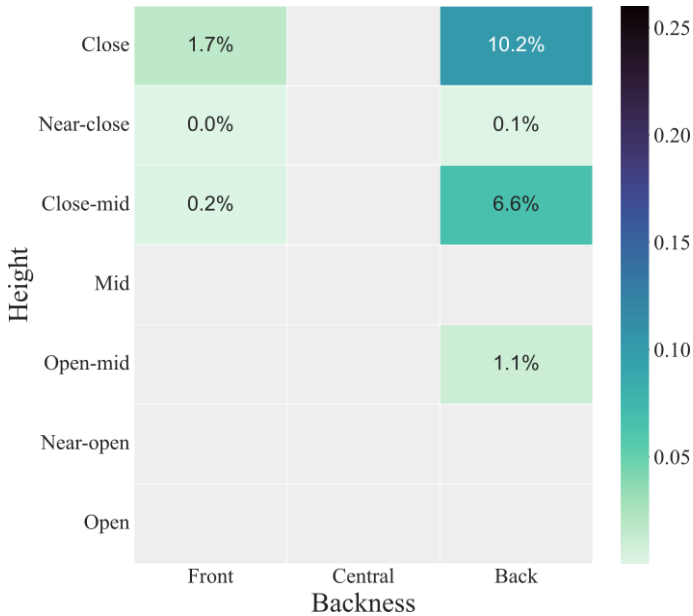


Figure 7.19: Heat map of the proportions of rounded vowels of all vowels in the affixal data.

The distribution of vowel segments aligns with some global tendencies, as the most common vowel segments are in descending order [a], [i], [u], [e] and [o]. The exact descriptions of the segments [a], [e] and [o] vary between languages however, as more detailed descriptions differentiate between [a]/[ɑ], [e]/[ɛ] and [o]/[ɔ], and the segments [ɑ], [ɛ] and [ɔ] are likely underreported in the data. This supports Maddieson's approach of only distinguishing between close, mid and open vowels in cross-linguistic comparisons (Maddieson 1984: 167-168). The vowel distribution also clearly demonstrates the preference for front vowels being unrounded and back vowels being rounded (Moran & McCloy 2019). The notable exception is the open unrounded back vowel [ɑ], which is even more interesting since it is potentially underreported in the data. The two other exceptions to this tendency are the close rounded front vowel [y] and the close unrounded back vowel [ɯ], which have almost equal proportions (1.7% and 1.8%). These equal proportions are likely explained by the Turkic languages, where these vowels are in complementary distribution due to vowel harmony.

The low proportion of close-mid front rounded vowels [ø] and complete absence of open-mid front rounded vowels [œ] and rounded central vowels should be mentioned. Although these vowel segments are typologically rare (Moran & McCloy 2019), particularly [ø] is not uncommon in Turkic and Chechen (Komen, Molochieva & Nichols 2020; Ragagnin 2022; Berta & Csató 2022). Rounded central vowels do not appear to occur as phonemes in any of the languages in the data, while rounded central allophones have been reported in Kabardian (Kuipers 1960: 22-23).

7.7. Affix type

The results show a clear preference for suffixes, which is a well-known cross-linguistic tendency (Himmelmann 2014). However, there are considerable differences between the five language families, as the Turkic languages are strictly suffixing whereas the Northwest Caucasian languages are primarily prefixing, cf. table 7.17.

Table 7.17: Mean proportions of the most common affix types by language family.

Affix type	*μ*	*SD*	ND	Kartvelian	NWC	IE	Turkic
Suffix	77.10%	18.39%	83.25%	60.41%	39.58%	81.82%	100%
Prefix	19.87%	18.24%	12.81%	34.60%	59.11%	17.23%	-
Infix	2.00%	3.68%	3.25%	0.12%	-	0.73%	-
Circumfix	0.89%	2.00%	0.47%	4.87%	1.23%	0.23%	-
Transfix	0.14%	0.55%	0.22%	-	0.07%	-	-

These results falsify the claim that the three Caucasian language families demonstrate a 'predominance of prefixal conjugation', as although these results in table 7.17. include nominal affixes as well, the mean proportion of only verbal prefixes in the Nakh-Dagestanian languages is still 19.91%, which includes preverbs. Thus, conjugation in the Nakh-Dagestanian languages is not predominantly prefixal, and the Iranian languages in the Caucasus are generally more prefixing than the Nakh-Dagestanian languages, which indicates that Chirikba's diagnostic feature of 'predominance of prefixal conjugation' only applies to the Northwest Caucasian and Kartvelian languages.

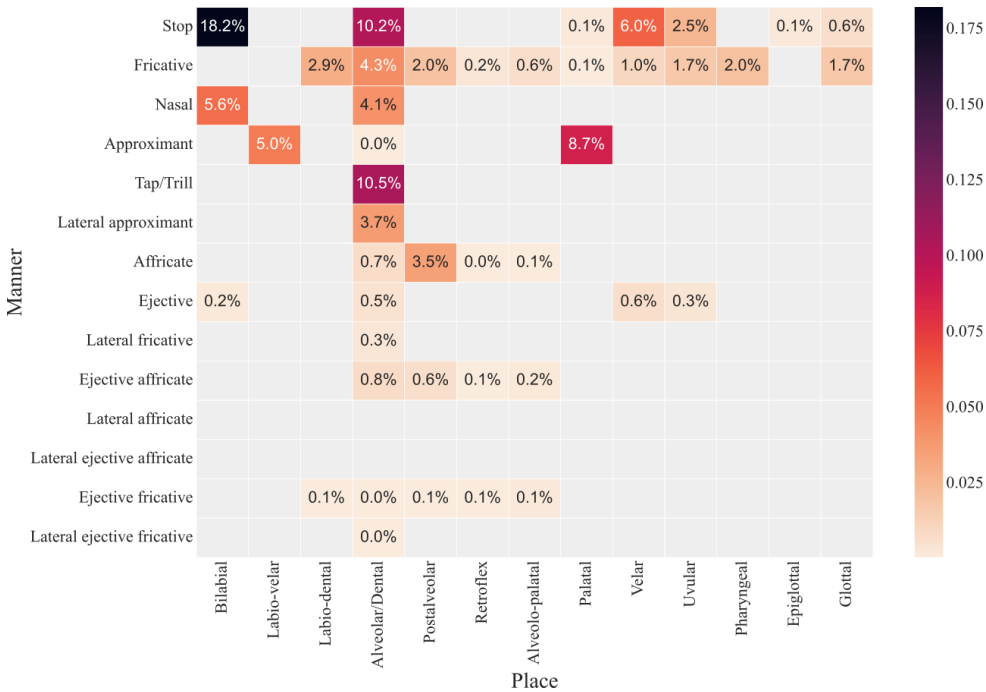
The Northwest Caucasian results should be problematised, as the large inventories of preverbs in these languages affect the outcomes, since preverbs constitute a mean proportion of 23.8% of the Northwest Caucasian affix data. If the preverbs are excluded, the Northwest Caucasian prefixes only form a mean proportion of 50.2%, which is then just slightly higher than the mean proportion of suffixes at 48.6%. If the preverbs instead are excluded from the prefixes for each Northwest Caucasian language, the results show that prefixes are the most common affix type for Abkhaz, Abaza, Kabardian and Adyghe. Abkhaz is the language with the maximal proportion of prefixes at 70.8%, which remains true even if all preverbs are excluded, as the proportion of non-preverb prefixes in Abkhaz is higher than all other languages at 62.0%.

This is relevant for two reasons. Firstly, as the preverb system in Abkhaz is notoriously rich, suggesting that Abkhaz has a general preference for prefixes, which likely will affect its morphology and phonology. Secondly, this shows that Abkhaz is an outlier among the languages of the Caucasus, which is relevant since it was the only Caucasian language included in Bybee's previous study on affixes (Bybee 2005). This shows that the widely applied method of random linguistic sampling is not unproblematic, as it risks including outliers such as Abkhaz that

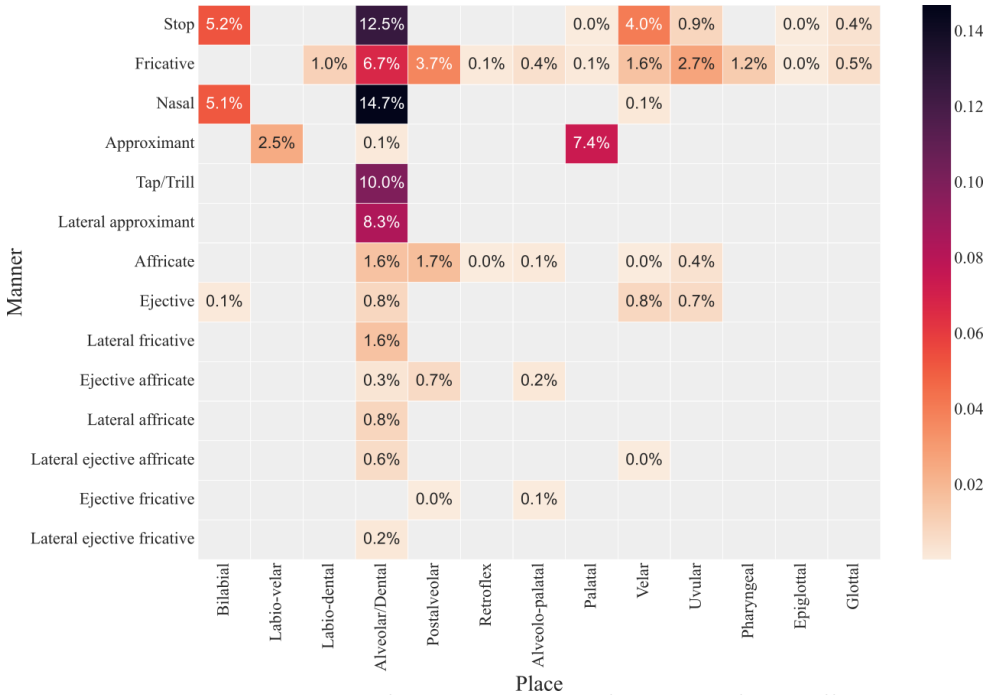
might skew the results. The preverbs naturally affect the Kartvelian results as well, because if the preverbs are excluded for the Kartvelian languages their mean affix type proportions are close to the mean proportions for all languages at 76.0% for suffixes and 17.9% for prefixes, while 5.9% for circumfixes. To exclude preverbs from the Kartvelian affixal data is however questionable, as they form an integral part of the TAM systems.

A comparison between the distributions of places and manners of articulation of prefixes (figure 7.20) and suffixes (figure 7.21) shows that there are demonstrable differences between the two affix types.

The prefixes demonstrate a distribution that is different from both affixes in general and from the lexical data, as stops/plosives are more frequent while fricatives and particularly nasals are less frequent. The stops/plosives are also the most frequent manner of articulation for the suffixes, while the nasals and particularly the alveolar stops are more common in the suffixes, as they are the most common combinations of place and manner of articulation.



Figures 7.20: Distribution of places and manners of articulation for all prefixes



Figures 7.21: Distribution of places and manners of articulation for all suffixes.

The fricatives are only slightly more frequent in the suffix data, while the lateral approximants are noticeably more common. The suffix data also differs from the prefix data as consonants from all 14 manners of articulation are found in the suffix data, while lateral affricates, lateral ejective affricates and lateral ejective fricatives are the only manners of articulation that are completely absent in the prefixes.

7.8. Phonological results by language family

7.8.1. Nakh-Dagestanian affixation

The phonological results for the Nakh-Dagestanian affixes are almost identical to the distribution for all languages, which is unsurprising since they constitute almost 60% of all languages in the data. The Nakh-Dagestanian affix data contain all manners of articulation except ejective fricatives and lateral ejective fricatives and all places of articulation except retroflex and alveolo-palatal consonants.

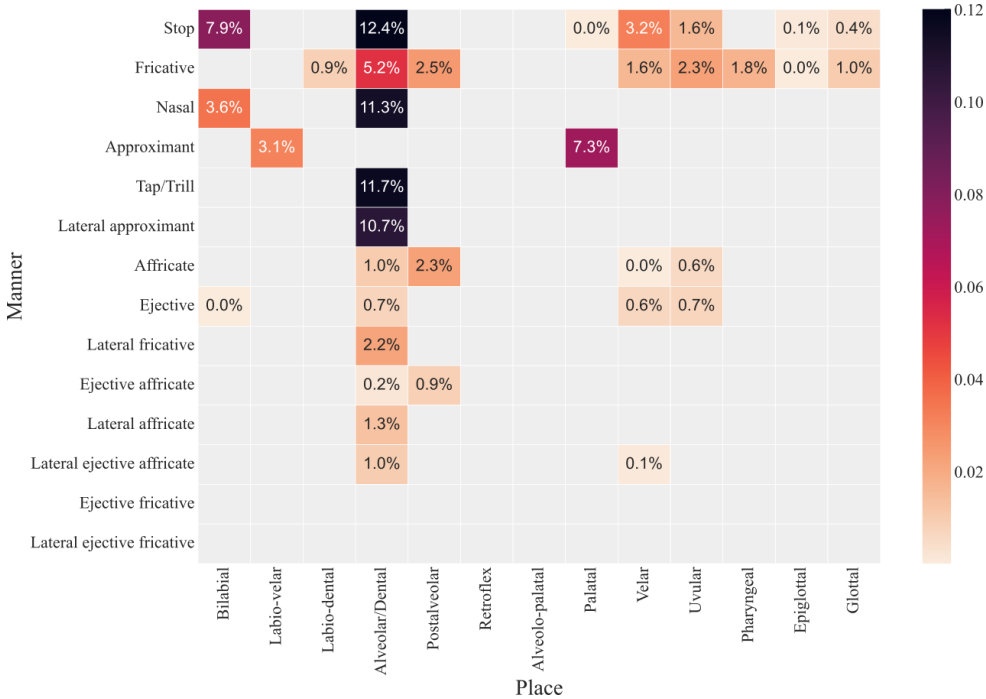


Figure 7.22: Distribution of places and manners of articulation for all consonants in the Nakh-Dagestanian affixes.

The most common combinations of place and manner are similar to the overall results, while alveolar/dental nasals are not the most common combination in the affixal data, which suggests that the generally high frequency of [n] is not primarily caused by the Nakh-Dagestanian languages. The high frequency of both taps/trills and lateral approximants in affixes seem to be a Nakh-Dagestanian tendency, as the other four families typically have either higher proportions of taps/trills or lateral approximants but not both. The lexical results in figure 7.23 are almost identical to the overall lexical results, while the Nakh-Dagestanian lexical data differ noticeably from the affixal data. The alveolar/dental stops and nasals both have lower mean proportions, while the alveolar/dental fricatives have identical proportions. The most apparent differences relate to velars, uvulars, ejectives, and ejective affricates, which mean that the Nakh-Dagestanian languages potentially skew the overall results.

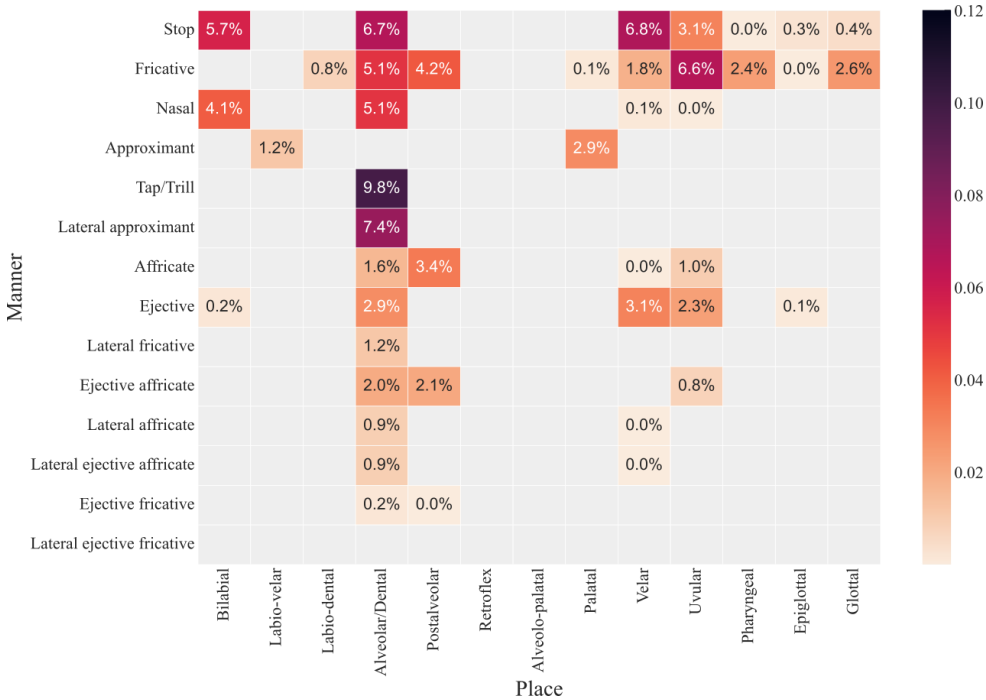


Figure 7.23: Distribution of places and manners of articulation for all consonants in the Nakh-Dagestanian lexical data.

The distribution of vowels in the Nakh-Dagestanian affixes reveals mean proportions that are almost identical to the overall results but with higher means, while the opposite is true for the typologically rare unrounded back vowels and the rounded front vowels. This does not apply for the rounded close-mid front vowel [ø], as these are only found in the Nakh-Dagestanian and Indo-European affixal data.

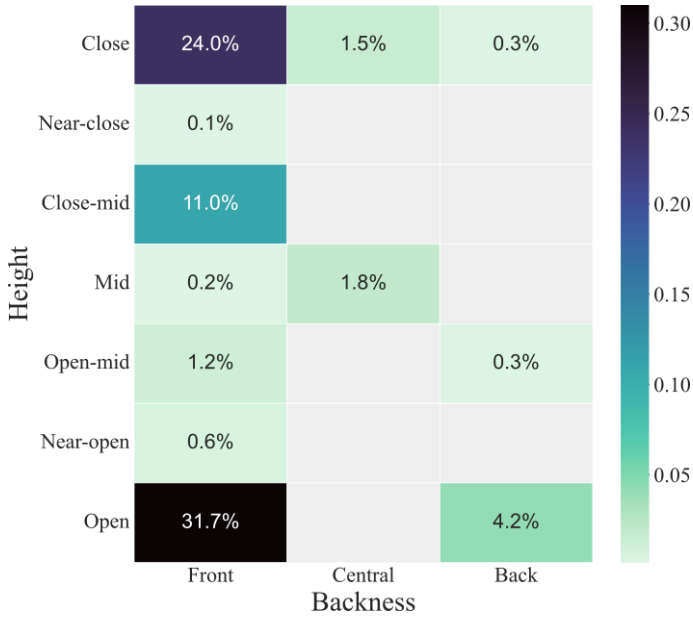


Figure 7.24: Distribution of vowel height and vowel backness for all unrounded vowels in the Nakh-Dagestanian affixes.

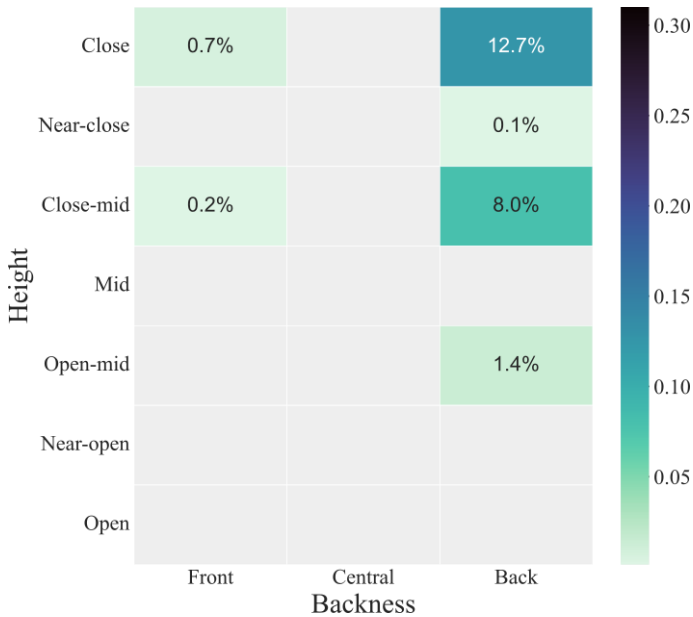


Figure 7.25: Distribution of vowel height and vowel backness for all rounded vowels in the Nakh-Dagestanian affixes.

7.8.2. Kartvelian affixation

The phonological results for the Kartvelian affixal data differ considerably from the distribution for all languages (cf. section 7.3.3). The Kartvelian languages have a smaller range of both places and manners of articulation in comparison to the other endemic language families of the Caucasus, as the Kartvelian languages generally have smaller consonant inventories.

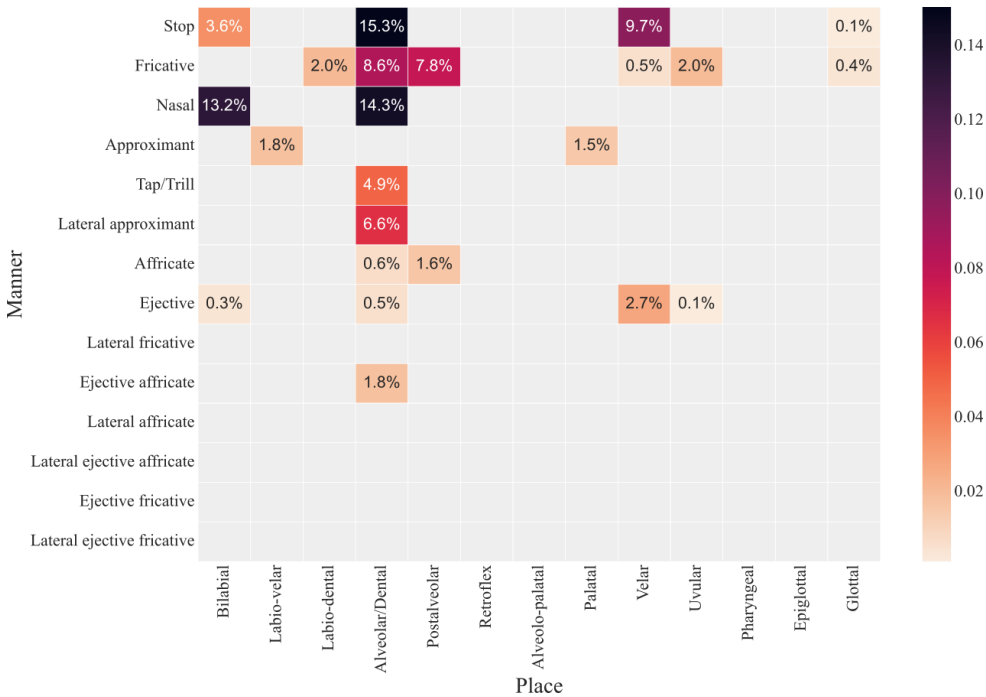


Figure 7.26: Distribution of places and manners of articulation for all consonants in the Kartvelian affixes.

The alveolar/dental stops also have the highest mean proportion in the Kartvelian affixes, while the alveolar/dental nasals and bilabial nasals are almost as common. The high proportion of velar stops set the Kartvelian affixes apart from the overall results, while the alveolar/dental and postalveolar fricatives are also more frequent. The Kartvelian affixes also have the lowest proportion of taps/trills of all five language families, while it shares the highest proportions of taps/trills in the lexical data with Turkic.

The lexical data reveal that the Kartvelian lexicon generally has higher mean proportions for all combinations except for the most common combinations in the affixal data, i.e. alveolar/dental stops, nasals and fricatives, postalveolar fricatives and velar stops. The higher mean proportions of ejectives, ejective affricates and affricates are similar to the tendency found for the Nakh-Dagestanian languages,

which could potentially explain why these languages cluster in the phonological PCA plots.

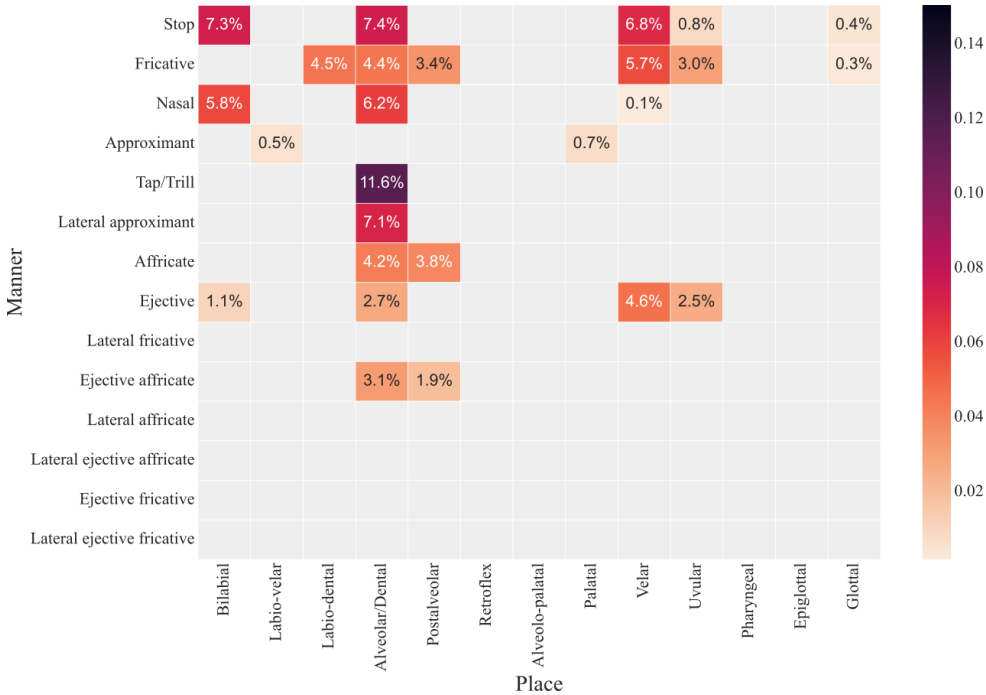


Figure 7.27: Distribution of places and manners of articulation for all consonants in the Kartvelian lexical data.

The distribution of vowels in the Kartvelian affixes clearly aligns with the division between front unrounded and back rounded vowels, while the open-mid front unrounded vowels are more frequent in comparison to the overall results. This is most potentially due to a more precise description of Kartvelian vowel segments, but there is nonetheless a remarkable difference as if the proportions of [e] and [ɛ] are merged as a single front mid vowel, they become the second most common vowel in the Kartvelian affixes. The results for the round vowels in Kartvelian affixes are also interesting, as they demonstrate an inverted relationship between [u] and [o] in comparison to the overall results. The proportions for the rounded mid-open back vowels [ɔ] are possibly to low, as they are only present in the Georgian affixal data, and the unrounded open back vowel [ɑ] is potentially present in Kartvelian as well, while the phonemes are typically only described as /a/.

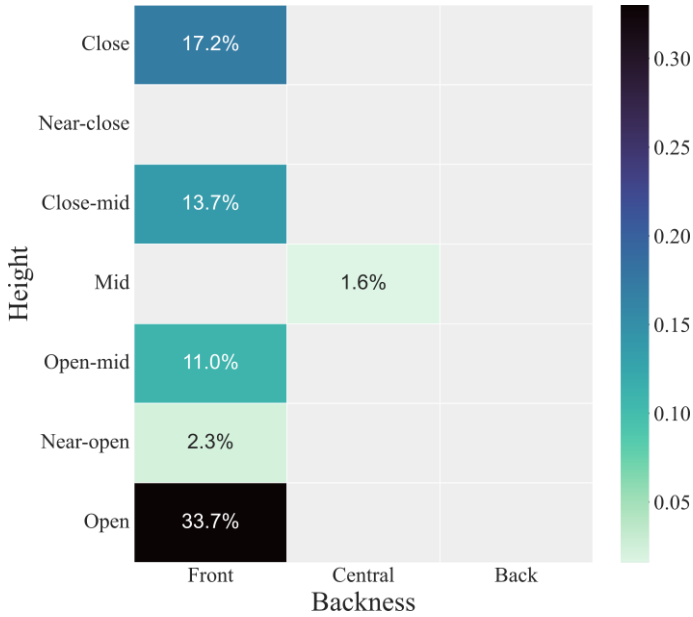


Figure 7.28: Distribution of vowel height and vowel backness for all unrounded vowels in the Kartvelian affixes.

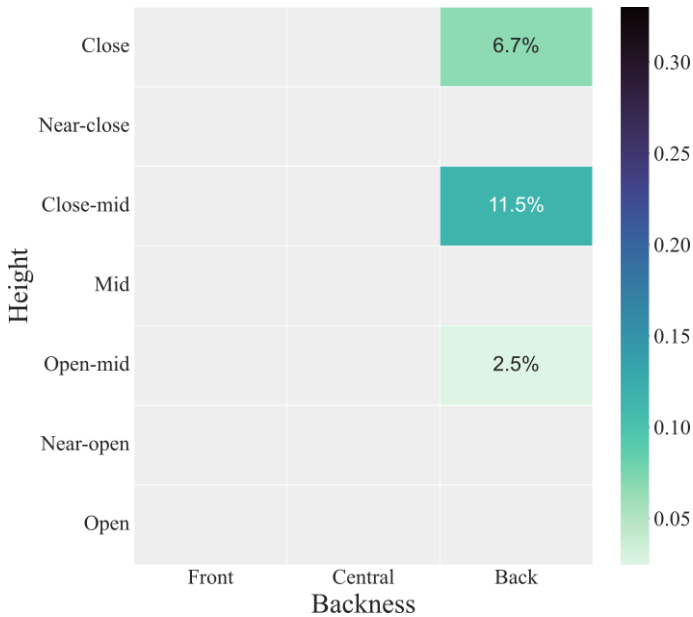


Figure 7.29: Distribution of vowel height and vowel backness for all rounded vowels in the Kartvelian affixes.

7.8.3. Northwest Caucasian affixation

The phonological results for the Northwest Caucasian affixes in section 7.3 appeared to present a starkly different pattern compared to the overall results, while figure 7.30 suggests that these differences are not as great as they might appear at first glance. The most obvious difference is that palatal approximants constitute the most common combination place and manner, while it is rather the lower frequencies of the most common alveolar/dental combinations that set the Northwest Caucasian affixal data apart.

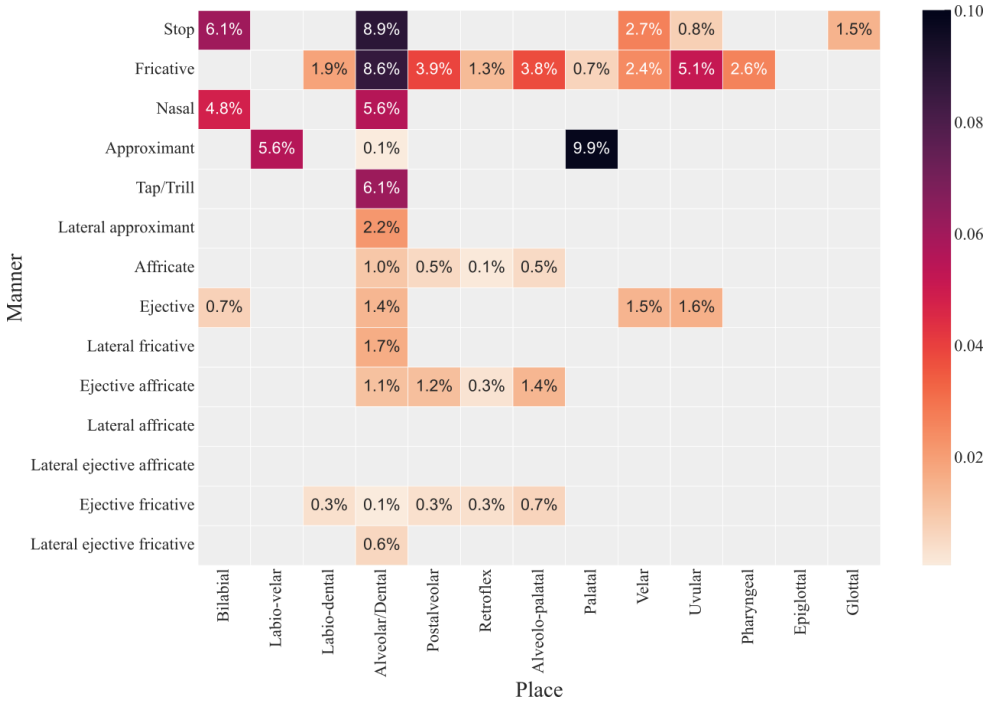


Figure 7.30: Distribution of places and manners of articulation for all consonants in the Northwest Caucasian affixes.

This does not apply to fricatives, as these are the most common place of articulation at >30%, and the mean fricative proportions are more similar to the overall results of the lexical data. The relatively high proportions of alveolo-palatal and retroflex fricatives present a particular pattern for the Northwest Caucasian languages. The low proportion of lateral approximants is not surprising since these are completely absent in most Northwest Caucasian languages, while the proportion of all lateral manners (4.4%) is still lower than for the lateral approximants in the other language families.

The stops present a distribution that is similar to the pattern found for affixes in all languages, while velar stops are less common (as they have become affricates in some languages) and the glottal stops are more common. The presence of ejective fricatives and lateral ejectives fricatives and simultaneous absence of lateral affricates and lateral ejective affricates almost present a mirror image of the results for the Nakh-Dagestanian affixes. The alveolar/dental nasals are also noticeably less common, while this tendency is not present for the bilabial nasals, which can likely be explained by the high proportion of prefixes in the Northwest Caucasian languages, cf. section 7.7.

The distribution of places and manners of articulation for the lexical data is largely different from both the overall lexical results and the Northwest Caucasian affixal data. The results of the Northwest Caucasian lexical data should be problematised, as they generally have a lower amount of lexical data per language which will affect the results.

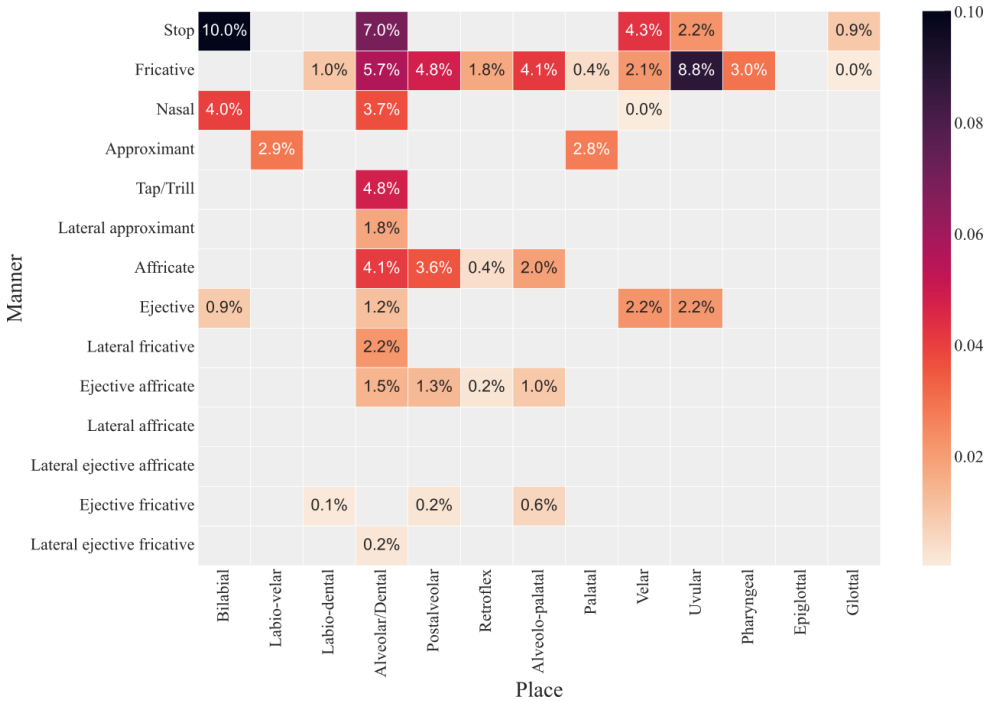
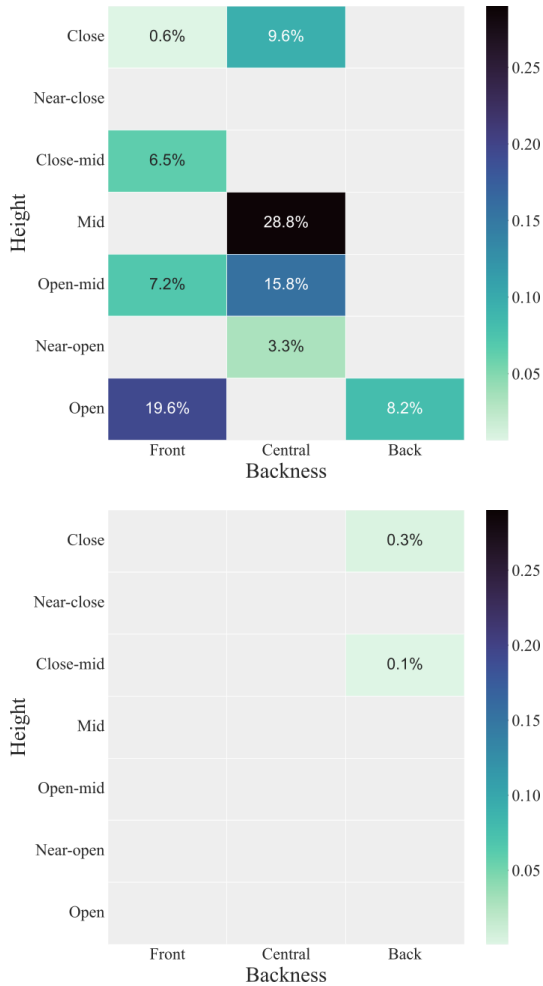


Figure 7.31: Distribution of places and manners of articulation for all consonants in the Northwest Caucasian lexical data.

The Northwest Caucasian lexical data have the highest proportions of numerous combinations, e.g. bilabial stops, uvular fricatives and pharyngeal fricatives, while they also have the lowest proportions of multiple combinations. The most interesting of these is the low mean proportion of alveolar/dental nasals, which is

mirrored by the affixal data, and it is the lowest proportion of all language families, both for the affixal and lexical data. The absence of certain combinations of ejective fricatives in the lexical data that are present in the affixal data is an indicator that the Northwest Caucasian lexical data appear to be insufficient, which might also explain some of the proportional highs and lows.



Figures 7.32 and 7.33: Distribution of vowel height and vowel backness for all unrounded (top) and rounded (bottom) vowel segments in the Northwest Caucasian affixes.

The distribution of vowel segments in the Northwest Caucasian affixal data present a truly different pattern, which we would expect from the typologically unusual vowel systems of these languages. The almost complete lack of rounded vowels might be misleading, as rounded allomorphs do occur depending on the surrounding

consonants. The high proportions of central unrounded vowels stand out from the other language families, while the remaining front unrounded vowels almost have the same distribution as the overall results, with the extremely low proportion of close front unrounded vowels being a noticeable exception.

7.8.4. Caucasian Turkic affixation

The distribution of places and manners in the Turkic affixes differs from the overall results, which is unsurprising as the Turkic languages have markedly fewer combinations in their phoneme inventories.

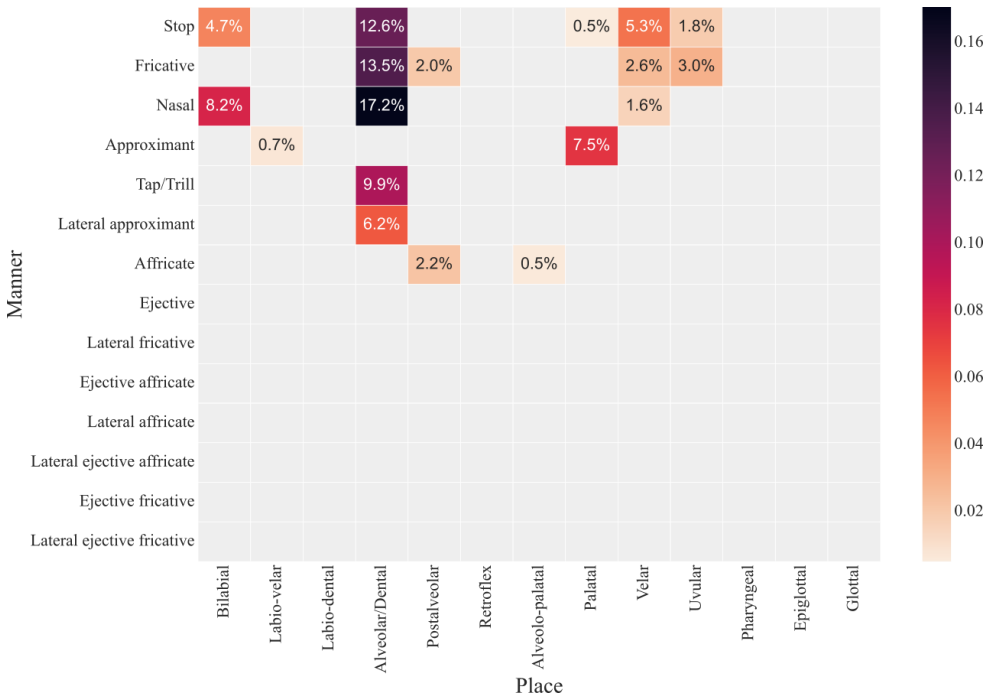


Figure 7.34: Distribution of places and manners of articulation for all consonants in the Turkic affixes.

The four most common combinations, i.e. alveolar/dental nasals, fricatives, stops and taps/trills, have remarkably high proportions as they form 53.2% of all combinations. The high mean proportion of alveolar/dental fricatives is worth mentioning as it is by far the highest of all language families for both affixal and lexical data. This is potentially connected to the fact that there are only four fricative combinations in the Turkic affixes, which could suggest that the proportions of manners are not dependent on the number of places of articulation in a language. This could also hold true for nasals, as the total proportions of nasals are almost identical for the Turkic, Kartvelian and Indo-European data, while another factor

could be that most of these languages have consonant inventories of roughly the same size, ranging from 22 to 38 with a mean inventory of 28 consonants. The high mean proportion of bilabial nasals differs from the overall results, while the same tendency is found in the Kartvelian and Indo-European affixes.

Comparing the Turkic affixal results with the lexical results in figure 7.35, the number of combinations is still small in comparison to the endemic language families and only three combinations are only found in the lexical data, i.e. labio-dental fricatives/approximants and glottal fricatives.

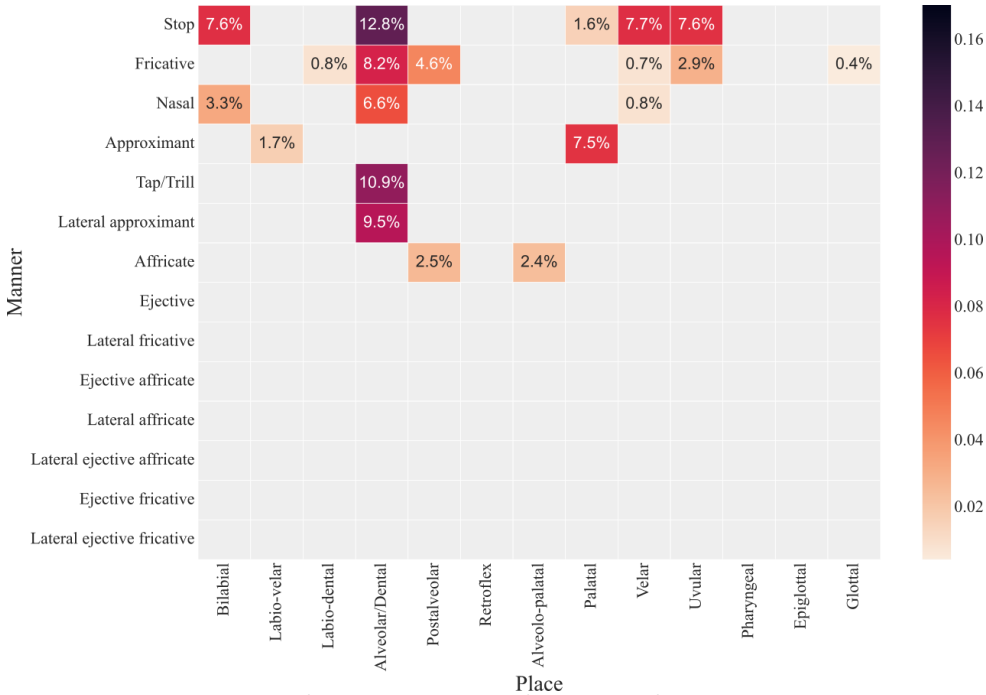
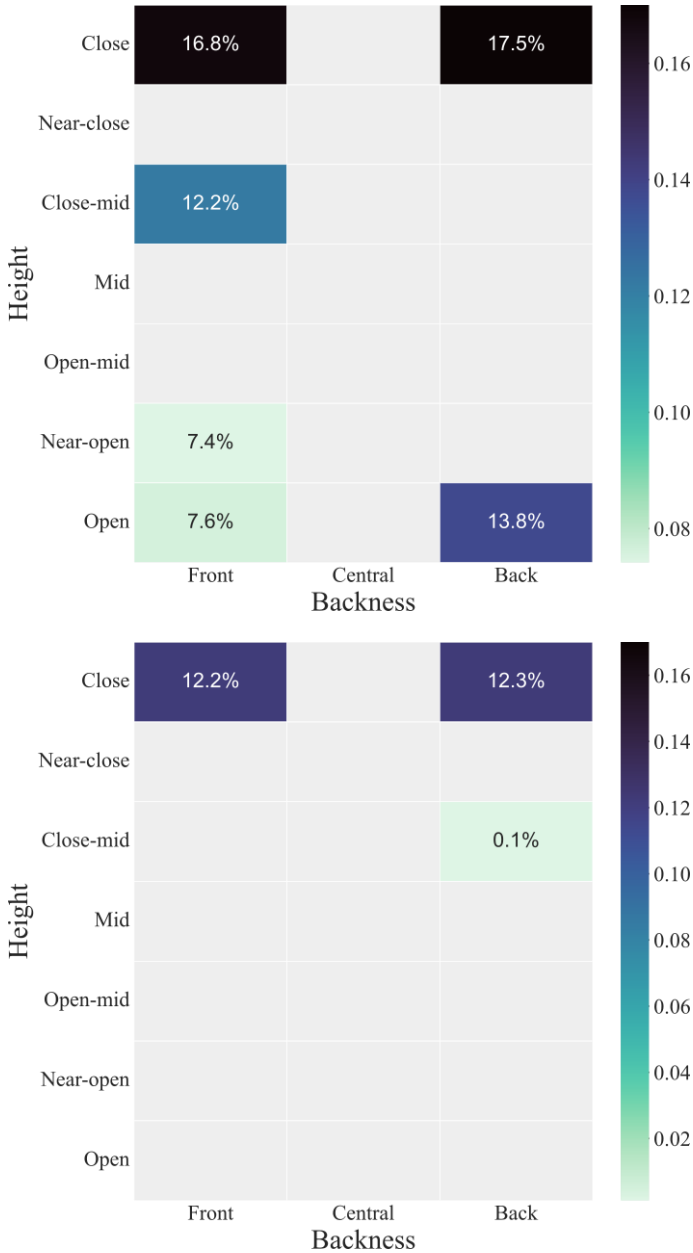


Figure 7.35: Distribution of places and manners of articulation for all consonants in the Turkic lexical data.

The Turkic lexical data differ quite considerably from the overall lexical results, while the five most common combinations overall are also the most common in Turkic, i.e. alveolar/dental stops, taps/trills and lateral approximants, velar stops and bilabial stops. It therefore becomes apparent that the most common combinations for both the Turkic affixal and lexical data are largely the same as for the overall results. This potentially suggests that larger phoneme inventories simply build upon a core set of combinations, where the typologically rare segments will also have lower proportions in the language. The low proportions of velar and uvular fricatives are surprising, while these fricatives might be underreported as certain Turkic languages tend to fricativise final velar and uvular stops.



Figures 7.36 and 7.37: Distribution of vowel height and vowel backness for all unrounded (top) and rounded (bottom) vowels in the Turkic affixes.

The vowel distributions for the Turkic affixal data demonstrate that the Turkic vowel harmony leads to a noticeably different pattern, which is primarily built around the distinctions of [ɯ], [i], [ɑ]/[a], [u], [y] and [e]. The near-open front

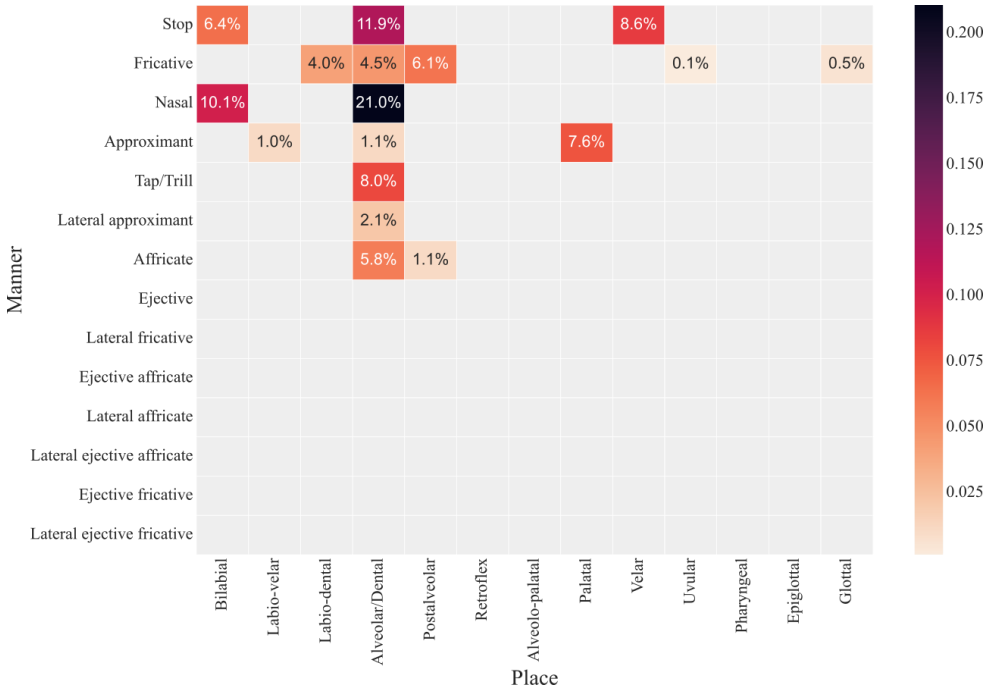
unrounded vowel [æ] is only found in North and South Azerbaijani affixes, which could suggest that it is an Iranian influence. The most interesting result is the complete absence of the mid front rounded vowels [ø]/[œ] and the near-complete absence of mid back rounded vowels [o]/[ɔ] in the Turkic affixal data, which is a well-known phenomenon in these languages, and it indicates that also vowels can be systematically excluded from affixes, cf. Bybee's notion of pattern exclusion (Bybee 2005).

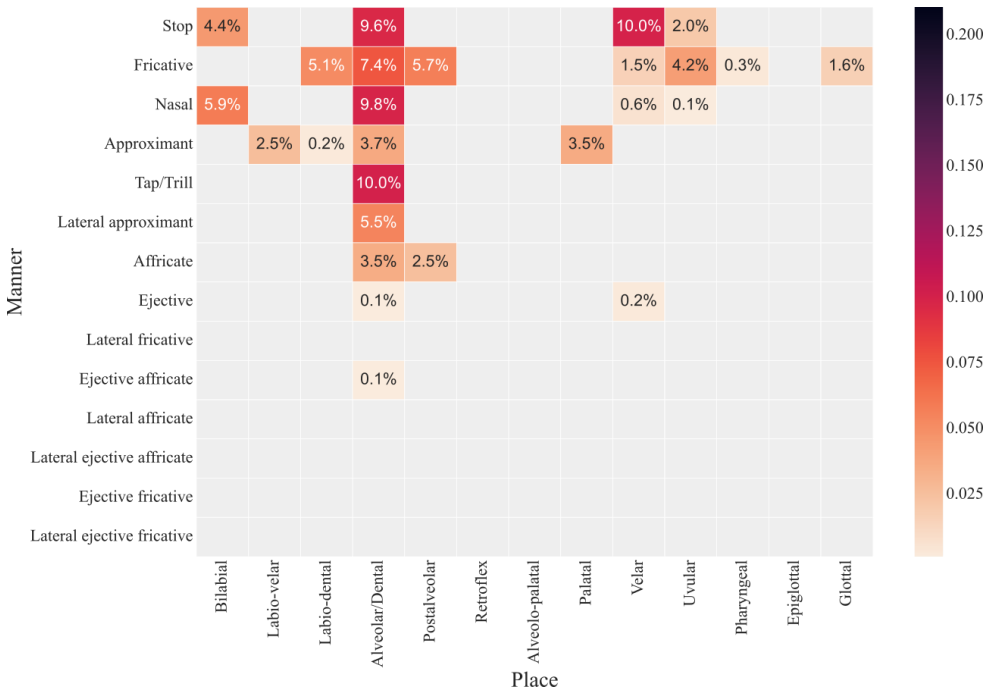
There are multiple possible explanations to why the Turkic languages would exclude mid rounded vowels from affixes, if the equation for potential morphemes from section 2.7 is applied. Turkic languages are well-known for their strict syllable structure, as initial consonant clusters are almost universally avoided while final consonant clusters with nasals, liquids and sibilants are allowed (Johanson 2022a: 27). Since Turkic languages are also strictly suffixing with a preference for CVC suffixes while having relatively small consonant inventories, the relevance of vowel pattern exclusion becomes apparent. This will likely help speakers to both predict word boundaries and detect suffixes in a surprisingly efficient manner, which is a potential impetus to the development of vowel harmony, as it primarily concerns affixation.

7.8.5. Caucasian Indo-European affixation

The distribution of places and manners of articulation for the Indo-European affixal data shows similarities with both the Turkic and Kartvelian languages, while they also have the fewest combinations of place and manner of the five language families. The Indo-European affixes differ from the other language families in certain regards, as they exhibit the highest proportion of alveolar/dental nasals and affricates, as well as labio-dental fricatives, while almost completely lacking velar and uvular fricatives.

The alveolar/dental approximant [ɹ] is only described in Classical Armenian, which could be criticised as inconsistent coding. However, it does suggest that the taps/trills category perhaps should be merged with the approximants, as this could potentially explain the high proportions of taps/trills as instances of a wider category of alveolar/dental approximant-like consonants in the analyses of the segmental distributions. The low mean proportion of lateral approximants is the lowest of all the affixal results, which is particularly interesting since many Northwest Caucasian languages lack lateral approximants altogether.





Figures 7.38 (top) and 7.39 (bottom): Distribution of places and manners of articulation for all consonants in the Indo-European affixal (top) and lexical data (bottom) (excl. Persian and Russian).

The lexical results for the Indo-European data reveal a distribution that is again similar to the Kartvelian and Turkic languages, while generally lacking ejectives, setting them apart from the Kartvelian languages. The small proportions of ejectives present are all found in the Iron Ossetic data. Similar to the affixal data, the proportion of alveolar/dental nasals is the highest for all lexical results. The difference in number of place and manner combinations between the lexical and affixal data sets distinguishes the Indo-European languages from the Turkic languages, as 26 combinations occur in the lexical data while only 16 combinations are found in the affixal data, cf. the Turkic difference of 21 vs. 18 combinations.

The distribution of Indo-European vowels generally align with the overall results. Close front unrounded vowels [i] are more frequent while the open front unrounded vowels [a] are less frequent, even if they are combined with the open back unrounded vowel [ɑ]. The high proportion of near-open central unrounded vowels [ɐ] is completely connected to Erschler's description of the phoneme /æ/ in Iron Ossetic as being realised as [ɐ] (Erschler 2020: 644). The close-mid front rounded vowels [ø] are possibly only found in one suffix in a sub-dialect of Tat, where the influence of Azerbaijani vowel harmony has triggered rounding of the plural suffix –ho into –hö (Suleymanov 2020: 95).

Figure 7.40: Distribution of vowel height and vowel backness for all unrounded vowels in the Indo-European affixes.

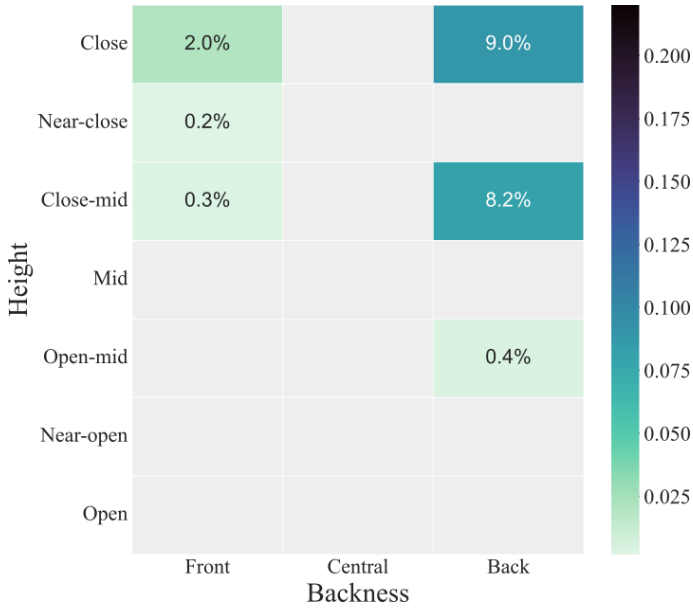


Figure 7.41: Distribution of vowel height and vowel backness for all rounded vowels in the Indo-European affixes.

7.9. Morphophonological results

7.9.1. Phoneme inventories and grammatical functions

In order to investigate the correlation between affixation and the size of a language's phoneme inventory, it is important to summarise the binarised grammatical functions described in section 7.1. This yields the types of grammatical functions expressed by affixation for each language, which are then compared to the number of consonant phonemes in each language (cf. appendix C).

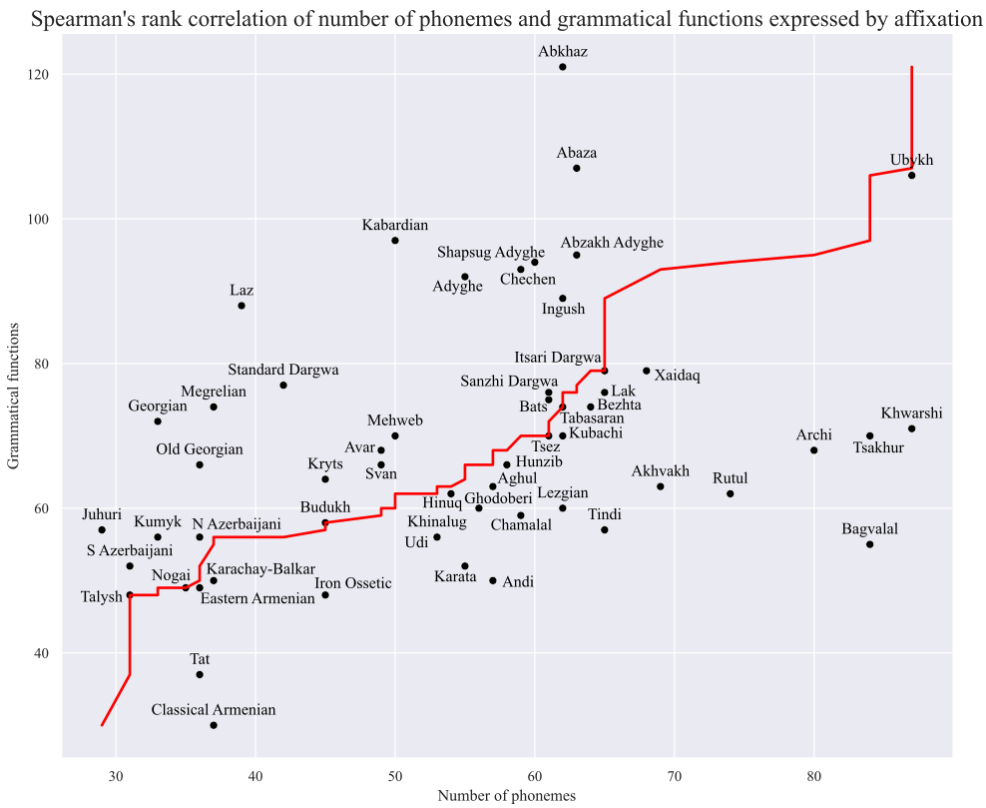


Figure 7.42: Spearman's rank correlation (red line) between the number of all phonemes, i.e. both consonants and vowels, and the number of grammatical functions expressed by affixation in all 56 languages.

The Spearman rank correlation coefficient, or Spearman's ρ, is a non-parametric measure that calculates 'the correlation of the ranks of the variables' (Kolassa 2020: 115). The Spearman rank correlation between only consonant phonemes and grammatical functions is moderately positive, $\rho = 0.51$, and significant, $p < .001$.

The correlation for of all phonemes, i.e. consonants and vowels, is less but still moderately positive, $\rho = 0.48$, and also statistically significant, $p < .001$, cf. figure 7.42. The large number of both consonant phonemes and grammatical functions in the Northwest Caucasian languages are worth testing for, as they will inevitably affect the results. If the Northwest Caucasian languages are excluded, the Spearman rank correlation for all phonemes is still moderately positive, $\rho = 0.46$, and significant, $p < .001$.

The Pearson correlation coefficient (Pearson's r) of these two variables indicates a significant positive correlation between the number of consonant phonemes and the number of grammatical functions expressed by affixation, as $r(55) = .51$ $p < .001$. If vowel phonemes are also included the Pearson correlation coefficient is slightly smaller but not significant, $r(55) = .41$ $p = .002$. The Pearson correlation coefficient of consonant inventories and grammatical functions for all non-Northwest Caucasian languages is weakly positive and non-significant, $r(48) = .33$ $p = .018$, while Pearson's r for all phonemes, i.e. consonants and vowels, is slightly more positive but still non-significant, $r(48) = .38$ $p = .007$.

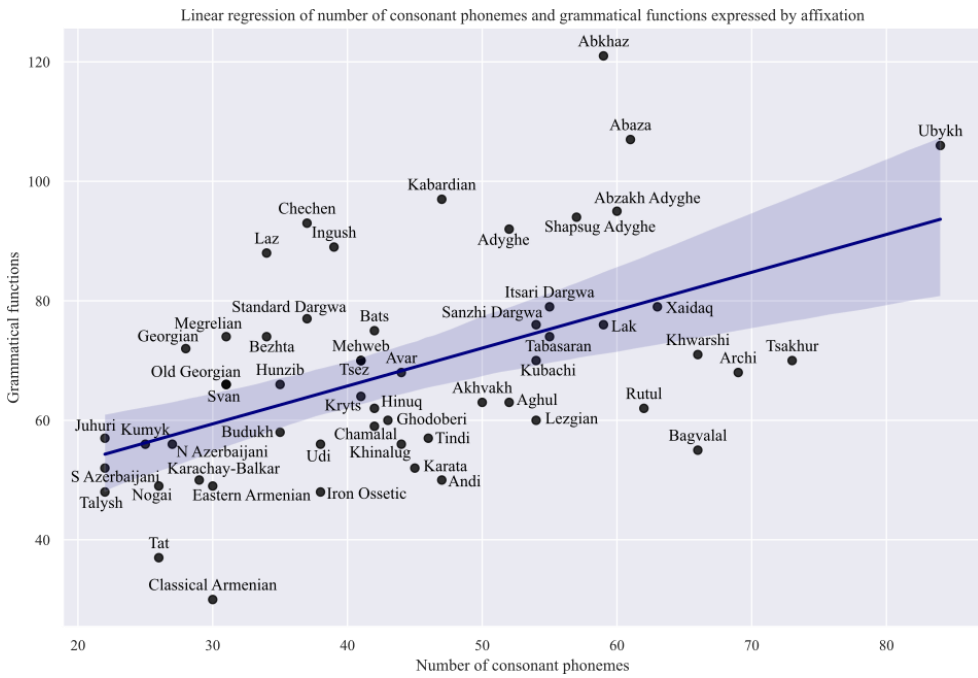


Figure 7.43: Linear regression of the consonant phoneme inventories and grammatical functions of all 56 languages.

A simple linear regression demonstrates a significant linear relationship between the number of consonant phonemes and grammatical functions in the languages of the Caucasus, as $R^2 = .26$, $F(1, 54) = 19.43$, $p < .001$, cf. figure 7.43. This indicates that

26% of the observed variance in grammatical functions can be explained by the number of consonant phonemes. If vowel phonemes are included, the linear relation is less positive and not significant, as $R^2 = .17$, $F(1, 54) = 10.95$, $p = .002$. If the Northwest Caucasian languages are excluded, there is again a stronger effect for all phonemes than for just consonants, but a simple linear regression of all non-Northwest Caucasian languages does not demonstrate a significant linear relationship, as $R^2 = .15$, $F(1, 47) = 7.99$, $p = .007$, and the regression for only consonant phonemes is clearly not significant, $R^2 = .11$, $F(1, 47) = 6.04$, $p = .018$.

This enables us to answer the sixth research question, i.e. whether there is a significant correlation between the phoneme inventories and the number of grammatical functions expressed by affixation in the languages of the Caucasus. The results above suggest a significant positive correlation between the number of consonant phonemes in a language and its number of grammatical functions. If both consonant and vowel phonemes are included, the results still suggest a significant yet smaller positive correlation. This indicates that both consonant and vowel phonemes should be considered when investigating the interaction between morphology and phonology. A simple linear regression also suggests a significant linear relation between the number of consonant phonemes in a language and the number of grammatical functions, at least in the languages of the Caucasus. For future research, it is therefore highly relevant to test whether these tendencies are specific to the Caucasus or if similar tendencies can be observed outside the Caucasus as well.

7.9.2. Morphophonological results by grammatical function

The data structure of this study enables analyses of the interaction between morphology and phonology on a functional level, as the segmental distributions of all grammatical functions can be analysed separately. As this is a topic in its own right, I only present some general tentative results to investigate this morphophonological interaction. It is only relevant to compare grammatical functions that are found across the five language families, while also acknowledging that the Nakh-Dagestanian languages will inevitably skew the results to a certain degree, as they constitute almost 60% of the 56 languages. I will present the results for the most common functions from the following grammatical categories: core cases (incl. genitive), non-core cases, local case orientations, tense functions and person-marking functions.

Table 7.18: Mean proportions of the ten most common trivariate consonants in the affixal data for the four most common case functions. The highest proportion per function is in bold.

Place and Manner	DAT	GEN	ERG	OBL
/n/	13.40%	**26.03%**	14.52%	9.35%
/r/	5.77%	3.92%	9.09%	11.06%
/d/	0.38%	4.46%	11.21%	10.21%
/l/	4.81%	15.61%	**17.33%**	**19.06%**
/j/	16.57%	8.61%	8.80%	8.14%
/b/	0%	5.87%	0.60%	4.84%
/m/	8.17%	9.35%	15.49%	17.62%
/s/	**19.87%**	8.06%	4.12%	0.85%
/t/	1.92%	1.02%	2.88%	1.13%
/w/	1.92%	1.89%	1.42%	0.46%

The results for the core case functions dative, ergative and oblique and the genitive case function are summarised in table 7.18, which includes the multifunctional oblique cases in the Northwest Caucasian and Iranian languages as datives, genitives and ergatives as well. Genitive has been included among the core cases, as the genitive is often syncretised with the other three functions. Although this is the case, there is still a remarkable variation between these four functions, and the difference between the dative and genitive functions is particularly interesting. The most common combinations of place and manner differ between the voiceless alveolar/dental fricative [s] for dative, the voiced alveolar/dental nasals [n] for genitive, and the voiced alveolar/dental lateral approximants [l] for ergative and oblique. The high mean proportion of alveolar/dental fricatives for the dative function could be explained by the Kartvelian languages, but they would just account for maximally 8.9% of these results as that is the Kartvelian share of the equalised data.

The result for the genitive function is also interesting, as the alveolar/dental nasals [n] are found in genitive affixes of all language families except Kartvelian, while the genitive affixes with alveolar/dental lateral approximants [l] are only found in the Nakh-Dagestanian languages. This also holds true for the ergative and oblique affixes, as the alveolar/dental lateral approximants [l] are only found in the Nakh-Dagestanian languages. This is unsurprising, as the genitive, ergative and oblique are often syncretised in some Nakh-Dagestanian languages, which also explains the similar distributions for the ergative and oblique.

Table 7.19: Mean proportions of the 15 most common trivariate consonants in the affixal data for the five most common non-core case functions. The highest proportion per function is in bold.

Place and Manner	INS	COM	LOC	ABL/ELA	ALL/LAT
/n/	**14.88%**	12.18%	3.84%	**17.34%**	7.01%
/r/	0.71%	4.27%	5.56%	13.50%	9.89%
/d/	13.15%	4.10%	**27.22%**	9.21%	5.16%
/l/	13.39%	**22.05%**	2.22%	9.41%	**12.30%**
/j/	5.62%	1.71%	2.30%	6.09%	8.62%
/b/	2.52%	0%	0%	0%	0.20%
/m/	0.66%	2.91%	5.56%	0%	2.22%
/s/	0%	0%	20.63%	8.93%	3.48%
/t/	9.26%	2.56%	2.78%	2.35%	1.09%
/w/	3.21%	0%	0.39%	1.34%	0.35%
/ʃ/	1.28%	2.14%	0%	5.63%	7.37%
/z/	6.05%	1.28%	2.78%	3.82%	5.82%
/k/	0.78%	4.27%	5.56%	1.76%	3.58%
/g/	2.78%	2.56%	1.85%	3.09%	4.95%
/ɬ/	0%	3.85%	0%	0.69%	0.40%

The results for the most common non-core cases (cf. table 7.19) appear to reveal a similar tendency, as the most common trivariate consonant is different for all five non-core case functions. These results are potentially skewed by the Nakh-Dagestanian languages, but they could also reflect a wider morphological tendency of the interaction of morphology and phonology, as morphology is reliant on phonology to uphold explicit grammatical distinctions. The distributional differences between the instrumental and the comitative are interesting, as they indicate that these are not just functionally separate but also phonologically differentiated in many languages of the Caucasus. The surprisingly low proportions of bilabial stops are likely not coincidental, as bilabial stops are also less frequent in the core case data, which would be interesting to test against non-Caucasian case systems.

If these results are compared to the six most common local case orientations (cf. table 7.20), the same tendency becomes even stronger, as the most common consonants are different for all six orientations. The orientation IN is peculiar

however, as many Nakh-Dagestanian languages indicate inessives with gender or noun class markers, and these markers are among the most common combination, i.e. [j] and [w]. Particularly the orientations AD, SUB, INTER and CONT are remarkable, as the highest proportions for these functions are all found among the rather unusual consonants /χ/, /tɬ'/, /tɬ/, and /tʃ'/.

Table 7.20: Mean proportions of the 15 most frequent trivariate consonants and /χ/, /tɬ'/, /tɬ/, and /tʃ'/ for the six most common local case orientations. The highest proportion per function is in bold.

Place and Manner	IN	SUPER	AD	SUB	INTER	CONT
/n/	10.59%	5.08%	6.38%	1.58%	0.88%	0%
/r/	6.22%	1.91%	1.79%	2.13%	1.72%	5.00%
/d/	6.13%	1.94%	3.57%	1.44%	0.40%	0%
/l/	6.83%	**31.66%**	0.71%	4.89%	12.28%	0%
/j/	**12.50%**	6.03%	0%	0%	0%	0%
/b/	5.63%	1.61%	0%	2.41%	1.21%	0%
/m/	10.00%	0%	1.79%	0%	0%	0%
/s/	2.59%	3.43%	5.95%	1.15%	1.75%	1.67%
/t/	0.44%	3.65%	3.57%	0%	0%	0%
/w/	8.40%	0.24%	5.36%	1.15%	0.40%	0%
/ʃ/	4.00%	0%	14.29%	0%	0%	0%
/z/	1.67%	3.13%	0%	0%	8.10%	2.50%
/k/	1.11%	0.52%	0.60%	19.11%	0%	20.00%
/g/	0.67%	3.13%	1.02%	10.34%	0%	0%
/ɬ/	0%	0%	0%	17.24%	21.93%	10.00%
/χ/	0%	0%	**15.39%**	0%	1.32%	2.50%
/tɬ'/	0%	18.75%	0%	**25.29%**	0%	5.00%
/tɬ/	3.33%	0%	0%	17.24%	**24.56%**	0%
/tʃ'/	0%	2.08%	0%	0%	0%	**26.67%**

The remaining local case orientations suggest that the less common a function is, the higher the proportion of uncommon consonants. The high proportions of lateral fricatives, lateral affricates and various lateral ejective manners could play an integral part of the Nakh-Dagestanian local case systems, which is further supported by the lack of both local case and lateral fricatives, lateral affricates and all lateral ejectives in Chechen and Ingush (Komen, Molochieva & Nichols 2020). This perhaps suggests that distinctive consonant phonemes are a requisite for large local case systems of the Nakh-Dagestanian type, which could be tested by investigating local case systems outside the Caucasus.

The distributions of trivariate consonants in the tense affixes (cf. table 7.21) appear not to have the same tendency, as the distributions only differ marginally between the three tense functions, and the distribution for the specific tense functions are noticeably more balanced if compared to the case functions above.

This is highly relevant, as it indicates that TAM affixes possibly yield different results than case affixes, which might explain the weak and inconclusive results in the previous study by Bybee (2005).

Table 7.21: Mean proportions of the ten most common triavariate consonants in the affixal data for the three tense functions. The highest proportion per function is in bold.

Place and Manner	PST	PRS	FUT
/n/	**17.27%**	**16.17%**	**22.13%**
/r/	14.24%	13.51%	11.80%
/d/	12.73%	7.53%	6.93%
/l/	5.20%	5.52%	7.45%
/j/	9.91%	8.69%	8.32%
/b/	3.47%	1.68%	2.17%
/m/	2.29%	6.97%	1.91%
/s/	4.07%	5.89%	9.82%
/t/	5.10%	6.20%	3.32%
/w/	3.46%	5.14%	2.84%

The final point of morphophonological comparison is subject person-marking affixes (cf. table 7.22), as these are also found in all five language families. These results reveal a similar tendency to the case functions, as the most common consonant segments are different for all person functions. This warrants further research, as similar patterns are most likely also found in other language families.

Table 7.22: Mean proportions of the thirteen most common trivariate consonants in the affixal data and /v/ for the subject person-marking functions. The highest proportion per function is in bold.

Place and Manner	1SG	2SG	3SG	1PL	2PL	3PL
/n/	8.10%	12.95%	16.92%	6.25%	10.89%	**25.91%**
/r/	4.63%	6.03%	10.48%	4.14%	5.98%	8.83%
/d/	8.99%	3.71%	13.76%	**22.47%**	14.55%	9.73%
/l/	0.37%	1.03%	3.14%	1.43%	1.22%	6.69%
/j/	4.55%	6.02%	**19.89%**	5.43%	5.13%	14.46%
/b/	1.56%	5.46%	5.99%	2.74%	1.07%	9.73%
/m/	**20.42%**	2.20%	4.39%	7.80%	0.87%	2.59%
/s/	18.70%	10.83%	14.09%	0.12%	4.01%	4.45%
/t/	2.09%	8.96%	2.22%	14.93%	**15.21%**	4.36%
/w/	2.45%	**15.20%**	0.99%	1.29%	1.11%	0%
/ʃ/	0.67%	8.37%	1.97%	2.53%	9.41%	2.01%
/k/	0.84%	1.56%	0%	4.34%	4.24%	0.47%
/z/	14.19%	0.22%	0.49%	3.47%	6.57%	0.16%
/v/	4.59%	1.88%	0.86%	2.86%	0.61%	0%

8. Conclusions

I conclude this thesis by answering the research questions formulated in chapter 1 and discussed in chapter 7. Afterwards I will discuss the more overarching conclusions that can be drawn from the results of this thesis.

The first research question asked whether the affix inventories of the three endemic language families, i.e. Kartvelian, Nakh-Dagestanian and Northwest Caucasian, display sufficient morphological similarities to support the notion of a Caucasian Sprachbund. The morphological results could not support the presence of such a morphological Sprachbund, as the results for the three endemic language families presented almost diametrically opposite systems. The three language families only share a few functions that are expressed by affixation, apart from the expected categories of core cases, number, tense, aspect and mood. The particular grammatical functions generally expressed by affixation in the Caucasus are thus the optative, interrogative, subjunctive/irrealis and conditional moods, negative polarity, non-witnessed past, causatives and perhaps most interestingly, the preverbs, which encode a wide range of spatial functions found in all the three endemic language families. There is also a remarkable set of highly specialised converbs, i.e. adverbial subordinators, shared between the Nakh-Dagestanian and Northwest Caucasian languages. Many of these converbs are also found in the Turkic languages, suggesting that they are not specific to the Caucasus. However, the phonological results could potential support a phonological sprachbund, but it would primarily include the Kartvelian, Nakh, Lezgic and Dargic languages. The most remarkable finding was that the Northwest Caucasian languages and the Avar-Andic-Tsezic branch of the Nakh-Dagestanian family are both morphologically and phonologically highly divergent, which makes it questionable to group them either areally or genealogically.

The second research question asked whether the affixation patterns of the Turkic and Indo-European languages spoken in the Caucasus exhibit morphological similarities with the three endemic language families of the Caucasus. The morphological results showed that the Turkic and Indo-European languages are generally more similar to each other than the three endemic language families. However, certain Nakh-Dagestanian languages do instead exhibit morphological similarities with the Turkic languages. Further research is therefore needed to investigate the well-known and long-standing contact between Kartvelian, Armenian, Iranian, Turkic, Nakh and the south Dagestanian languages.

The third research question asked whether there are hierarchical patterns in the distribution of grammatical functions expressed by affixation across the five language families of the Caucasus. The distribution of grammatical functions did reveal a tendency for larger affix inventories to be built upon core functions found across all systems, while also adding more complex functions in an almost predictable order as certain functions are only present in the largest systems. The morphological hierarchies should therefore be tested against language families outside the Caucasus to examine if these are general or Caucasus-specific tendencies. The intrafamilial variation is surprisingly high, as the data suggest that even closely related languages differ morphologically with regard to how various grammatical functions are expressed. Comparing morphology based on grammatical functions therefore presents an interesting approach to the wider study of linguistic diversity and morphological change, as the data indicate that closely related languages can be identical except for certain grammatical functions. Changes in particular grammatical functions can potentially explain how languages and dialects diverge over time and space, function by function.

Continuing with the phonological research questions, the fourth research question intended to explore whether there are significant phonological differences between affixes and lexicon in the languages of the Caucasus, based on the articulatory variables of place, manner and voicing. The results showed that place and manner of articulation alone did not differ significantly, but combinations of place and manner differed significantly between affixes and lexicon in the Caucasus. The results for voicing alone differed significantly, which suggests that voicing might help speakers to differentiate between affixes and lexicon in these languages. This is an important finding, since the languages of the Caucasus are famous for their many ejective consonants and this could potentially explain why these languages also have such intricate affixation patterns. Having a systematic phonological difference between lexical stems and affixes potentially enable speakers of these languages to detect stems and predict what will come next, as particularly Kartvelian and Northwest Caucasian languages have verb morphology with stable affix ordering and a fixed stem slot. These patterns together with a systematic difference between stems and affixes are likely the prerequisites for producing, comprehending and predicting morphologically complex languages.

The fifth research question asked whether there are significant differences in the distribution of certain places and manners of articulation between affixes and lexicon in the languages of the Caucasus, as these languages have large inventories of places and manners of articulation. The results suggest that only certain combinations of place and manner differ between affixes and lexical stems. Particularly alveolar/dental nasals, uvulars and the various ejective manners all presented a significantly different distribution between the affixes and lexical data, suggesting that certain consonants are more or less frequent in affixes. When voicing was added, the results became more complex as voiced alveolar/dental stops [d] differed significantly while voiceless alveolar/dental stops [t] did not, while the

bilabial stops [b] and [p] presented an opposite situation. The large number of consonant segments only found in the lexical data further supported the observation that affixes and lexical stems differ phonologically.

The variation between the five language families is considerable for both consonant and vowel segments, although there are observable phonological similarities between particularly the Kartvelian, Turkic and Indo-European, and to a varying degree the Nakh, Dargic and Lezgic languages. The phonological results generally suggest that geography and language contact affect phonological distributions of both affixes and lexicon. The Northwest Caucasian and Avar-Andic-Tsezic languages are the most divergent both when it comes to phonology and morphology, which potentially could be explained by the same geographical factors. The results also suggest phonological changes might interact with morphological changes, e.g. the simultaneous lack of both local cases and lateral fricatives, lateral affricates and all lateral ejective manners in Chechen and Ingush. Another example of the potential interaction between phonology and morphology could be that [b] has among the lowest proportions of all Caucasian languages in the Lezgian lexical data, while they are completely absent in the affixal data. This could explain why Lezgian has lost grammatical gender or noun classes, as almost all Nakh-Dagestanian noun class systems include [b] as a marker. These results suggest that phonological variation and change could explain morphological variation between related languages, while the importance of diachronic data and reconstruction become apparent in order to differentiate innovation and loss from patterns inherited from earlier language stages.

The sixth and last research question asked whether there is a significant correlation between phoneme inventory size, i.e. the number of consonant and vowel phonemes, and the number of grammatical functions expressed by affixation in the languages of the Caucasus. The results suggested a significant positive correlation between phoneme inventory size and the number of grammatical functions expressed by affixes. The results also indicate that both the number of consonant and vowel phonemes correlate significantly with the number of grammatical functions, while the results for only consonant phonemes are less conclusive as there is a significant positive correlation for all five language families. If the Northwest Caucasian languages are excluded the correlation is still moderately positive and statistically significant. This suggests that particularly consonant phonemes play an integral part in affix inventories, which was the central hypothesis of this thesis. Are large consonant inventories a prerequisite for complex morphology? The significant correlation between consonant inventory size and affix inventory size seems to indicate this. However, factors such as vowel harmony will affect the number of affixes, which makes it important to differentiate between distinctive affix functions and the total number of allomorphs. These results should be compared to languages with complex morphology and small phoneme inventories, as they will either falsify this assumption or give us a better understanding of the interaction between morphology and phonology.

Furthermore, the observed general preference for mono-syllabic and mono-consonantal affixes in the Caucasus is likely an important factor. Is there a global preference for mono-syllabic affixes, similar to the observed global preference for suffixation? If so, then phoneme inventory size would be an important delimiting factor for morphological complexity. This calls for a wider investigation of how phoneme inventories and phonotactics correlate with morphology, to examine whether this is only a Caucasian tendency or whether it applies to language in general.

The final conclusions reflect upon the methodology of this thesis, as the data-driven approach of this study enables in-depth analyses of both morphology and phonology in a way that can likely be applied to any language, any language family or any linguistic area. The data structure should however be expanded by including tone to ensure validity for many language families outside the Caucasus. Secondary articulation was encoded in the data structure, but it should likely be treated as a separate variable for each trivariate consonant and vowel. The conclusions from the phonological results indicate that the trivariate consonant, i.e. place-manner-voicing, has a great potential for wide-scale phonological analyses and that all three variables need to be considered. The *type*-based approached of this thesis should however be compared to *token*-based data, e.g. by applying this methodology to corpus data.

The phonological coding could therefore be applied to almost any language family if it is expanded according the suggestions above, while the coding of grammatical functions would need to be adjusted to fit a global sample. However, the inventory of grammatical functions described in this thesis should be seen as a baseline for any morphological description or comparison, as any function that is expressed by morphology in one language can be morphology in any language. Systematically comparing how these grammatical functions are expressed and realised in the languages of the world would potentially reveal even greater insights into the interaction between morphology and phonology. This approach might be problematic for languages with little or no morphology, which would be highly interesting to investigate further. Do languages with no morphology differ phonologically from languages with complex morphology? If so, could phonological systems predict morphological patterns, or are these variables completely independent? The results from this thesis suggest that the latter would be unlikely, since the observed correlation between phoneme inventory size and grammatical functions expressed by affixation in the Caucasus indicates that morphological complexity depends on phonological factors.

Another interesting observation that can be drawn from the phonological results is that phoneme distributions are remarkably uniform across both related and unrelated languages, particularly for the lexical data. Although the specific phonemes differ, there appears to be general distributional patterns, as e.g. the languages without ejectives have distributions of stops that are comparable to the combined proportions of stops and ejectives in the endemic languages of the

Caucasus. Similar patterns are also potentially observable for velars and uvulars in the Kartvelian, Indo-European and Turkic languages, while the high proportions of uvulars in the Nakh-Dagestanian and Northwest Caucasian languages potentially obscure any general tendencies. This warrants further investigation into phoneme distribution in languages outside the Caucasus.

The methodology used in this thesis also enables phonological analyses of lexical data, which could be applied to purely lexical studies. The lexical data were also coded for semantic meaning, and the possibility of analysing larger lexical data sets both phonologically and semantically seems promising as another development. The data structure developed in this thesis can therefore be expanded to investigate languages outside the Caucasus, since the morphological and phonological complexity of the languages of the Caucasus was chosen as a formidable challenge. Past linguistic descriptions have given us enormous amounts of potential data, which we ought to turn into analysable data, as this is truly an underexploited resource available for future linguistic research.

Subject index

References

Abaev, V. I. (1964). ‘A Grammatical Sketch of Ossetic’, Publication of the Indiana University Research Center, *Anthropology, Folklore, and Linguistics*, 35. The Hague: Mouton & Co.

Abdullaev, S. (1954). *Grammatika darginskogo jazyka*. Maxačkala: Dagestanskij filial Akademii nauk SSSR.

Abdullaev, Z. G. (1971). *Očerki po sintaksisu darginskogo jazyka*. Moscow: Akademija Nauk SSSR.

Abdullaeva, A. Z., Gadžiaxmedov, K. S., Kadyradžiev, I. A., Kerimov, N. X., Olmesov, D. M. & Xangišev, D. M. (2014). *Sovremennyj kumykskij jazyk*. Maxačkala: RAN

Adelaar, Willem F. H. & Muysken, Pieter C. (2004). *The Languages of the Andes*. Cambridge: Cambridge University Press.

Aglarov, Mamaykhan A. (1994). ‘Andis’, Paul Friedrich & Norma Diamond (eds.), *Russia and Eurasia / China*. Encyclopedia of World Cultures VI. Boston: G. K. Hall, pp. 23-27.

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

Aikhenvald, Alexandra Y. (2007), ‘Typological distinctions in word-formation’, Timothy Shopen (ed.), *Language Typology and Syntactic Description*. Volume III: Grammatical Categories and the Lexicon. Cambridge: Cambridge University Press, pp. 1–65.

Aikhenvald, Alexandra Y. (2010). *Imperatives and Commands*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. (2018). ‘Evidentiality: The framework’, Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*. Oxford: Oxford University Press, pp. 1-46.

Ajello, Roberto (1997). ‘Armenian’, Giacalone Ramat & Paulo Ramat (eds.), *The Indo-European Languages*. London: Routledge, pp. 197-227.

Alekseev, M. E. (1994a). ‘Rutul’, Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 213-258.

Alekseev, M. E. (1994b). 'Budukh', Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus*, volume 4. North East Caucasian Languages Part 2. Delmar, New York: Caravan Books, pp. 259-296.

Alekseev, M. E. (1999). ‘Naxsko-dagestanskie jazyki’, M. E. Alekseev (ed.) *Jazyki mira. Kavkazskie jazyki*. Moscow: Academia, pp. 156-165.

Alekseev, M. E, Ataev, Boris, Magomedov, M. A., Magomedov, M. I., Madieva, G. I., Saidova , P. A. & Samedov, Dž. S. (2012). *Sovremennyj avarskij jazyk*. Maxačkala: ALEF.

Alekseev, M. E., Maisak, T. A., Ganenkov, D. S., Lander, J. A. (eds) (2008). *Udinskij sbornik: grammatika, leksika, istorija jazyka*. Moscow: Academia.

Alekseev, M. E. & Radžabov, R. N. (2004). 'Tsez', Michael Job (ed.), *The Indigenous Languages of the Caucasus, volume 3. The North East Caucasian Languages: Part 1*. Ann Arbor, Michigan: Caravan Books, pp. 115-168.

Alekseev, M. E. & Šixalieva, Sabrina X. (2003). *Tabasaranskij jazyk*. Moscow: Nauka.

Aliev, U. B. (1973). *Sintaksis karačaevo-balkarskogo jazyka*. Moscow: Nauka.

Aliroev, I. Ju. (1999). *Čečenskij jazyk*. Moskva: Academia.

Anderson, Ralph Dewitt (1963). *A Grammar of Laz*. Diss. Austin: University of Texas at Austin.

Anderson, Stephen R. (2015). 'Dimensions of morphological complexity', Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds), *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press, pp. 11–26.

Applebaum, Ayla Ayda Bozkurt (2013). *Prosody and Grammar in Kabardian*. Diss. Santa Barbara: University of California.

Arche, María J., Fábregas, Antonio & Marín, Rafael (eds). *The Grammar of Copulas Across Languages*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Aristava, Š. K. (1968). *Grammatika abxazskogo jazyka*. Suxumi: Alašara.

Ariztimuño López, Borja (2014). 'The origin of the Basque partitive', Silvia Luraghi & Tuomas Huumo (eds.), *Partitive Cases and Related Categories*, Berlin: De Gruyter Mouton, pp. 323-346.

Arkadiev, Peter & Lander, Yury (2020). 'The Northwest Caucasian Languages', Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 369-446.

Arkadiev, Peter, Lander, Yury & Bagirokova, Irina (2024). 'Applicative constructions in the Northwest Caucasian languages', Fernando Zúñiga & Denis Creissels (eds.), *Applicative Constructions in the World's Languages*. Berlin: De Gruyter Mouton, pp. 869-912.

Aronson, Howard I. (1990). *Georgian: a reading grammar*. Corr. ed. Columbus, Ohio: Slavica Publishers.

Asatrian, Garnik & Borjian, Habib (2005). 'Talish and the Talishis (The State of Research)'. *Iran and the Caucasus*, Vol. 9, 1. Leiden: Brill, pp. 43-72.

Authier, Gilles (2009). *Grammaire kryz: Langue caucasique d'Azerbaïdjan, dialecte d'Alik*. Collection linguistique de la Société de linguistique de Paris. Paris: Peeters.

Authier, Gilles (2010). 'Finite and non-finite: Prosodic distinctions on Budugh verb stems', I. Bril (ed.), *Clause linking and clause hierarchy: syntax and pragmatics*. Amsterdam/Philadelphia: John Benjamins, pp. 143-164.

Authier, Gilles (2012). *Grammaire juhuri, ou judéo-tat, langue iranienne des Juifs du Caucase de l'est*. Wiesbaden: Reichert.

Babaliyeva, Ayten (2007). *Présentation du dialecte lezgi de Yargun (Azerbaïdjan): grammaire, textes glosés et traduits*. MA thesis, École Pratique des Hautes Études.

Babaliyeva, Ayten (2013). *Études sur la morphosyntaxe du tabasaran littéraire*. Diss. Paris: École Pratique des Hautes Études.

Baerman, Matthew (2009). 'Case syncretism', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 219-230.

Baerman, Matthew, Brown, Dunstan & Corbett, Greville G. (eds.) (2015). *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press.

Baerman, Matthew, Brown, Dunstan & Corbett, Greville G. (2017). *Morphological Complexity*. Cambridge: Cambridge University Press.

Bagaev, N. K. (1965). *Sovremennyj osetinskij jazyk*. Ordžonikidze: Severo-Osetinskoe knižnoe izd.

Bakuradze, Lia, Beridze, Marine & Pourtskhvanidze, Zakharia (2020). 'A Georgian Language Island in Iran: Fereydani Georgian', *Iranian Studies*, 53(3–4), pp. 489–550.

Baldi, Philip (2002). *The Foundations of Latin*. Berlin: Mouton.

Beguš, Gašper (2020). 'Segmental Phonetics and Phonology in Caucasian Languages', Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 689-728.

Belyaev, Oleg (2020). 'Indo-European languages of the Caucasus', Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 574-639.

Bennett, C. E. (1908). *A Latin Grammar*. 2nd ed. Boston and Chicago: Allyn and Bacon.

Bentz, C., Gutierrez-Vasques, X., Sozinova, O. & Samardžić, T. (2023). 'Complexity trade-offs and equi-complexity in natural languages: a meta-analysis', *Linguistics Vanguard*, vol. 9, s1, pp. 9-25. https://doi.org/10.1515/lingvan-2021-0054

Berikashvili, Svetlana (2017). *Morphological Aspects of Pontic Greek Spoken in Georgia*. München: LINCOM.

Berta, Árpád & Csató, Éva Á. (2022). 'West Kipchak languages', Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 320-336.

Bickel, Balthasar & Nichols, Johanna (2007). 'Inflectional morphology', Timothy Shopen (ed.) *Language Typology and Syntactic Description*. Volume III: Grammatical Categories and the Lexicon. Cambridge: Cambridge University Press, pp. 169-240.

Bisang, Walter (1995). 'Verb serialization and converbs – differences and similarities', Martin Haspelmath & Ekkehard König (eds), *Converbs in Cross-linguistic Perspective*. Berlin: Mouton de Gruyter, pp. 137–188.

Blake, Barry J. (2001). *Case*. 2nd ed. Cambridge: Cambridge University Press.

Blevins, James P., Ackerman, Farrell & Malouf, Robert (2019). 'Word and Paradigm Morphology', Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*. Oxford: Oxford University Press, pp. 265-284.

Boeder, Winfried (2000). 'Evidentiality in Georgian', Lars Johanson & Bo Utas (eds.), *Evidentials: Turkic, Iranian and Neighbouring Languages*. Berlin: Mouton de Gruyter, pp. 275-328.

Boeder, Winfried (2005). ‘The South Caucasian Languages’. *Lingua*, volume 115, Issues 1–2, pp. 5-89.

Boeschoten, Hendrik (2022). ‘The speakers of Turkic languages’, Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 1-12.

Bokarev, Anatolij A. (1949a) *Očerk grammatiki čamalinskogo jazyka*. Moscow: Akademija nauk SSSR.

Bokarev, Anatolij A. (1949b) *Sintaksis avarskogo jazyka*. Moscow: Akademija nauk SSSR.

Bokarev, E. A. (1959). *Tsezskie (didojskie) jazyki Dagestana*. Moscow: Akademija nauk SSSR.

Bolkvadze, Tinatin & Kiziria, Dodona (2023). *Georgian: A Comprehensive Grammar*. London: Routledge.

Borise, Lena (2020). ‘Tone and Intonation in Languages of the Caucasus’, Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 757-782.

Borjian, Habib (2015). ‘Judeo-Iranian Languages’, Lily Kahn & Aaron D. Rubin (eds.), *Handbook of Jewish languages*. Brill's handbooks in linguistics 2. Leiden: Brill, pp. 234-296.

Bulut, Christiane (2022). ‘Turkic languages of Iran’, Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 287-302.

Butskhrikidze, Marika (2002). *The Consonant Phonotactics of Georgian*. Diss. Leiden: Universiteit Leiden.

Butt, Miriam (2009). ‘Modern approaches to case: an overview’, Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 27-43.

Bybee, Joan (1985). *Morphology : a study of the relation between meaning and form*. Amsterdam ; John Benjamins.

Bybee, Joan (2005). ‘Restrictions on phonemes in affixes: A crosslinguistic test of a popular hypothesis’. *Linguistic Typology*, 9, pp. 165–222.

Bybee, Joan, Perkins, Revere & Pagliuca, William (1994). *The Evolution of Grammar: tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.

Camaj, Martin (1984). *Albanian grammar: with exercises, chrestomathy, and glossaries*. Wiesbaden: Harrassowitz.

Carling, Gerd (ed.) (2019). *The Mouton Atlas of Languages and Cultures*. Berlin: De Gruyter Mouton.

Carling, Gerd (ed.) (2024). *Diachronic Atlas of Comparative Linguistics Online*. Frankfurt am Main: Goethe University: https://diacl.uni-frankfurt.de/.

Carlson, Greg (2012). ‘Habitual and generic aspect’, Robert I.Binnick (ed.), *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp. 828-851.

Catford, J. C. (1977). ‘Mountain of Tongues: The Languages of the Caucasus’, *Annual Review of Anthropology* 6, pp. 283–314.

Charachidzé, Georges (1981). *Grammaire de la langue avar: langue du Caucase nord-est*. Saint-Sulpice de Favières: Jean-Favar.

Chechuro, Ilya (2019). 'Nominal morphology of Mehweb', Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 39-72.

Cherchi, Marcello (1999). *Georgian*. Languages of the World/Materials 147. München: LINCOM Europa.

Chirikba, Viacheslav A. (1996). *Common West Caucasian: the reconstruction of its phonological system and parts of its lexicon and morphology*. Leiden: Research School CNWS.

Chirikba, Viacheslav A. (2003a). *Abkhaz*. München: LINCOM Europa.

Chirikba Viacheslav A. (2003b). 'Evidential category and evidential strategy in Abkhaz', Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Studies in Evidentiality*. Amsterdam: John Benjamins.

Chirikba Viacheslav A. (2008). 'The problem of the Caucasian Sprachbund', Pieter Muysken (ed.), *From Linguistic Areas to Areal Linguistics*. Amsterdam; John Benjamins, pp. 25-93.

Chukhua, Merab (2019). *Georgian-Circassian-Apkhazian Etymological Dictionary*. Tbilisi: TSU Arnold Chikobava Institute of Linguistics.

Chumakina, Marina (2020). 'Archi', Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 281-316.

Chumakina, Marina, Bond, Oliver & Corbett, Greville G. (2016) "Essentials of Archi grammar', Oliver Bond, Greville G. Corbett, Marina Chumakina, and Dunstan Brown (eds), *Archi: Complexities of Agreement in Cross-Theoretical Perspective*. Oxford: Oxford University Press, pp. 17-42.

Čikobava, Arnold (1965). *Iberiul-k'avk'asiur enata šesc'avlis ist'oria*. Tbilisi: Ganatleba.

Clackson, James P. T. (2008). 'Classical Armenian', Roger D. Woodard (ed.), *The Ancient Languages of Asia Minor*. Cambridge: Cambridge University Press, pp 124-144.

Coene, Frederik (2010). *The Caucasus: An introduction*. Abingdon: Routledge.

Čokaev, Kati Z. (1970). *Affiksacija*. Morfologija čečenskogo jazyka II. Groznyj: Čečeno-Ingušskoe knižnoe izdatel′stvo.

Colarusso, John (1992). *A Grammar of the Kabardian Language*. Calgary: University of Calgary Press.

Comrie, Bernard (1976). *Aspect*. Cambridge: Cambridge University Press.

Comrie, Bernard (1985). *Tense*. Cambridge: Cambridge University Press.

Comrie, Bernard (2000). 'Valency-changing derivations in Tsez', Alexandra Y. Aikhenvald & Robert M. W. Dixon, *Changing Valency: Case Studies in Transitivity*. Cambridge: Cambridge University Press, pp. 360-374.

Comrie, Bernard, Forker, Diana & Khalilova, Zaira (2012). 'Adverbial clauses in the Tsezic languages', Gast, V. & Diessel, H. (ed.), *Clause Linkage in Cross-Linguistic Perspective: Data-Driven Approaches to Cross-Clausal Syntax*. Berlin: De Gruyter Mouton, pp. 157-190.

Comrie, Bernard & Polinsky, Maria (1998). 'The Great Daghestanian case hoax', A. Siewierska & J. J. Song (eds.), *Case, Typology and Grammar*. Amsterdam: John Benjamins, pp. 95-114.

Comrie, Bernard & Polinsky, Maria (2007). 'Evidentials in Tsez', Zlatka Guentchéva & Jon Landaburu (eds.), *L'énonciation médiatisée II: le traitement épistémologique de l'information: illustrations amérindiennes et caucasiennes*. Leuven: Peeters, pp. 335-350.

Comrie, Bernard & Khalilov, Madzhid (2009) 'Loanwords in Bezhta, a Nakh-Dagestanian language of the North Caucasus', Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: a comparative handbook*. Berlin: Mouton, pp. 414-429.

Comrie, Bernard, Xalilov, Madžid & Xalilova, Zaira (2015). *Grammatika bežtinskogo jazyka*. Leipzig-Maxačkala: MPIEVA.

Corbett, Greville G. (1991). *Gender*. Cambridge: Cambridge University Press.

Creissels, Denis (2008). 'Person variations in Akhvakh verb morphology : functional motivation and origin of an uncommon pattern'. *Sprachtypologie und Universalienforschung* 61(4), pp. 309-325.

Creissels, Denis (2009a). 'Spatial cases', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 609-625.

Creissels, Denis (2009b). 'Participles and finiteness: The case of Akhvakh'. *Linguistic Discovery* 7.1, pp. 106-130.

Creissels, Denis (2010). 'Specialized converbs and adverbial subordination in Axaxdere Akhvakh', in I. Bril (ed.) *Clause linking and clause hierarchy: syntax and pragmatics*. Amsterdam/Philadelphia: John Benjamins, pp. 105-142.

Creissels, Denis (2018). 'Perfective Tenses and Epistemic Modality in Northern Akhvakh', Diana Forker & Timur Maisak, *The Semantics of Verbal Categories in Nakh-Dagestanian Languages*. Leiden: Brill, pp. 166-187.

Csató, Éva Á. & Karakoç, Birsel (1998). 'Noghay', Lars Johanson & Éva Ágnes Csató (eds.), *The Turkic Languages*. London: Routledge, pp. 333-343.

Daniel, Mixail A. (2001a). 'Imja suščestvitel′noe', A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 127-150.

Daniel, Mixail A. (2001b). 'Padež i lokalizacija', A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 203-230.

Daniel, Michael (2019). 'Mehweb verb morphology', Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 73-116.

Daniel, Michael, Dobrushina, Nina & Ganenkov, Dmitry (2019). *The Mehweb language: Essays on phonology, morphology and syntax*. Berlin: Language Science Press.

Daniel, Michael & Ganenkov, Dmitry (2009). ‘Case marking in Daghestanian: limits of elaboration’, Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 668-685.

Daniel, Michael & Spencer, Andrew (2009), ‘The vocative – an outlier case’, Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 626-634.

Daragan, J. V. & Maisak, Timur A. (2001). ‘Časticy, sojuzy i transkategorial′nye pokazateli’, A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 173-184.

de Haan, Ferdinand (2013). ‘Coding of Evidentiality’, Matthew S. Dryer & Martin Haspelmath (eds.), *WALS Online* (v2020.3): https://wals.info.

de Hoop, Helen & Zwarts, Joost (2009). ‘Case in formal semantics’, Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 170-184.

de Swart, Henriëtte (2012). ‘Verbal Aspect’, Robert I.Binnick (ed.), *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp. 752-780.

Dehghani, Yavar (2000). *A grammar of Iranian Azari: including comparisons with Persian.* München: LINCOM Europa.

Dellert, J., Daneyko, T., Münch, A. et al. (2019). *NorthEuraLex 0.9*. Lang Resources & Evaluation: https://doi.org/10.1007/s10579-019-09480-6 (version 0.9).

Dešeriev, Junus D. (1953). *Bacbijskij jazyk. Fonetika, morfologija, sintaksis, leksika*. Moscow: Akademija Nauk SSSR.

Dešeriev, Junus D. (1959). *Grammatika xinalugskogo jazyka*. Moscow: Akademija nauk SSSR.

Dešerieva, T. I. (1999a) ‘Naxskie jazyki’, M. E. Alekseev (ed.) *Jazyki mira. Kavkazskie jazyki*. Moscow: Academia, pp. 166-172.

Dešerieva, T. I. (1999b) ‘Čečenskij jazyk’, M. E. Alekseev (ed.) *Jazyki mira. Kavkazskie jazyki*. Moscow: Academia, pp. 173-185.

Dixon, Robert M. W. & Aikhenvald, Alexandra Y. (2000). *Changing Valency: Case Studies in Transitivity*. Cambridge: Cambridge University Press.

Dobrushina, Nina (2019a). ‘The language and people of Mehweb’, Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 1-16.

Dobrushina, Nina (2019b). ‘Moods in Mehweb’, Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 117-166.

Dobrushina, Nina R., Kibrik, A. E., Tatevosov, S. G. (2001). ‘Glagol’, A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 67-125.

Dobrushina, Nina, van der Auwera, Johan & Goussev, Valentin (2005). 'The Optative', Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press, pp. 298-301.

Dobrushina, Nina, Daniel, Michael & Koryakov, Yuri (2020). 'Languages and Sociolinguistics of the Caucasus', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus,* 31. Oxford: Oxford University Press, pp. 27-66.

Donaldson, Bruce (2006). *German: An essential grammar*. London: Routledge.

Donohue, Mark, Hetherington, Rebecca, McElvenny, James & Dawson, Virginia (2013). *World Phonotactics Database*. Department of Linguistics, The Australian National University.

Dryer, Matthew S. & Haspelmath, Martin (eds.) (2013). *WALS Online* (v2020.3): https://wals.info.

Dum-Tragut, Jasmine (2009). *Armenian: modern Eastern Armenian.* Amsterdam: John Benjamins.

Dumézil, Georges (1931). *La Langue des Oubykhs*. Paris: Champion.

Durkin-Meisterernst, Desmond (2014). *Grammatik des Westmitteliranischen: Parthisch und Mittelpersisch*. Veröffentlichungen zur Iranistik 73. Wiesbaden: Harrassowitz.

Erschler, David (2020). 'Iron Ossetic', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus,* 31. Oxford: Oxford University Press, pp. 641-688.

Fähnrich, Heinz (1991). 'Old Georgian', Alice C. Harris (ed.), *The Kartvelian Languages. The Indigenous Languages of the Caucasus 1*, Delmar, New York: Caravan Books, pp. 129-218.

Fähnrich, Heinz (1994). *Grammatik der altgeorgischen Sprache.* Hamburg: Helmut Buske.

Fähnrich, Heinz (2001). *Batsisch (Zowatuschisch)-Deutsches Wörterbuch*. Jena: Friedrich-Schiller Universität.

Fähnrich, Heinz (2007). *Kartwelisches Etymologisches Wörterbuch*. Leiden: Brill.

Fähnrich, Heinz. (2012) *Die georgische Sprache*. Leiden: Brill.

Federal State Statistic Service (2020). 'Tom 5: Nacional′nyj sostav i vladenie jazykami', *Vserossijskaja perepis' naselenija 2020 goda* [All-Russian population census 2020]. Moscow: Rosstat: https://rosstat.gov.ru/vpn/2020/Tom5_Nacionalnyj_sostav_i_vladenie_yazykami

Fenwick, Rohan S.H. (2011). *A Grammar of Ubykh.* LINCOM Studies in Caucasian Linguistics. München: LINCOM Europa.

Fillmore, Charles J. (1968). 'The Case for Case', Emmon, Bach & Robert T. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, pp. 1-88.

Filonenko, V. I. (1940). *Grammatika balkarskogo jazyka.* Nal'čik: Kabardino-Balkarskoe Gosudarstvennoe Izdatel'stvo.

Forker, Diana (2013). *A grammar of Hinuq.* Mouton Grammar Library, 63. Berlin: De Gruyter Mouton.

Forker, Diana (2018a). 'Evidentiality in Nakh-Dagestanian languages', Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*. Oxford: Oxford University Press, pp. 490-509.

Forker, Diana (2018b). 'The Semantics of Evidentiality and Epistemic Modality in Avar', Diana Forker & Timur Maisak, *The Semantics of Verbal Categories in Nakh-Dagestanian Languages*. Leiden: Brill, pp. 188-214.

Forker, Diana (2020a). *A grammar of Sanzhi Dargwa.* Languages of the Caucasus 2. Berlin: Language Science Press.

Forker, Diana (2020b). 'Avar grammar sketch', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus,* 31. Oxford: Oxford University Press, pp. 243-280.

Forsyth, James (2013). *The Caucasus: A History*. Cambridge: Cambridge University Press.

Friedman, Victor A. (1992). 'Lak Substantival Declension: 40 cases or 50?', *The Non Slavic Languages of the USSR: Linguistic Studies Second Series*, 1992. Chicago: Chicago Linguistic Society, pp. 113-33.

Friedman, Victor A. (2018). 'Where do evidentials come from?', Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*. Oxford: Oxford University Press, pp. 124-147.

Friedman, Victor A. (2020). 'Lak', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus,* 31. Oxford: Oxford University Press, pp. 201-242.

Fritz, Catherine O., Morris, Peter E., & Richler, Jennifer J. (2012). 'Effect size estimates: Current use, calculations, and interpretation'. *Journal of Experimental Psychology: General*, 141(1), pp. 2–18.

Ganenkov, Dmitry (2008). 'Morfologičeskaja i semantičeskaja xarakteristika padežej v udinskom jazyke', M. E. Alekseev, T. A. Maisak, D. S. Ganenkov, J. A. Lander (eds), *Udinskij sbornik: grammatika, leksika, istorija jazyka*. Moscow: Academia, pp. 27-61.

Ganenkov, Dmitry (2019). 'Case and agreement in Mehweb', Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 189-234.

Ganenkov, Dmitry & Maisak, Timur (2020). 'Nakh-Dagestanian Languages', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus*, 31. Oxford: Oxford University Press, pp. 87-146.

Ganieva, Faida A. (2002) *Xinalugsko-russkij slovar'.* Maxačkala: Inst. jazyka, literatury i iskusstva im. G. Cadasy Dagestanskogo naučnogo centra RAN.

Ganieva, Faida A. (2007). *Džabinskij dialekt lezginskogo jazyka*. Maxačkala: Rossijskaja Akademija Nauk.

Ganieva, Faida A. (2008). *Kurušskij govor Lezginskogo jazyka*. Maxačkala: Rossijskaja Akademija Nauk.

Genko, Anatolij N. (1955). *Abazinskij jazyk*. Moscow: Akademia Nauk SSSR.

Gippert, Jost, Schulze, Wolfgang, Aleksidze, Zaza & Mahe, Jean-Pierre (2008). *The Caucasian Albanian Palimpsests of Mount Sinai*. Turnhout: Brepols.

Göksel, Aslı & Kerslake, Celia (2005). *Turkish: a comprehensive grammar*. London: Routledge.

Goodwin, William W. (1897). *Syntax of the Moods and Tenses of the Greek Verb*. Rewritten and enlarged. Boston: Ginn & Company.

Gordon, Matthew K. (2016). *Phonological Typology*. Oxford Survey In Phonology & Phonetics. Oxford: Oxford University Press.

Gudava, T'ogo (1962). *Botlixuri ena: gramat'ik'uli analizi, t'ekst'ebi, leksik'oni*. Tbilisi: Sakartvelos SSR mecnierebata ak'ademiis gamomcemloba.

Gudjedjiani, Chato & Palmaitis, Mykolas L. (1986). *Upper Svan: Grammar and texts*. Vilnus: Mokslas.

Güldemann, Tom (2013). 'Typology', Rainer Vossen (ed.), *The Khoesan languages*. London: Routledge, pp. 25-37.

Gvozdanović, Jadranka (2012). 'Perfective and Imperfective Aspect', Robert I. Binnick (ed.), *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp. 781-802.

Haig, Geoffrey (2018). 'Northern Kurdish (Kurmanjî)', Geoffrey Haig & Geoffrey Khan (eds.), *The Languages and Linguistics of Western Asia*. Berlin: DeGruyter Mouton, pp. 106-158.

Harris, Alice C. (1985). *Diachronic syntax: The Karvelian Case*. Syntax and Semantics, vol. 18. Orlando: Academic Press.

Harris, Alice C. (1991a). 'Overview on the history of the Kartvelian languages', Alice C. Harris (ed.), *The Indigenous Languages of the Caucasus, volume 1. The Kartvelian Languages*. Delmar, New York: Caravan Books, pp. 9-83.

Harris, Alice C. (1991b). 'Mingrelian', Alice C. Harris (ed.), *The Indigenous Languages of the Caucasus, volume 1. The Kartvelian Languages*. Delmar, New York: Caravan Books, pp. 313-394.

Harris, Alice C. (2002). *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.

Harris, Alice C. (2009). *Georgian syntax: a study in relational grammar*. Cambridge studies in linguistics, 33. Cambridge: Cambridge University Press.

Harris, C. R., Millman, K. J., van der Walt, S. J. et al. (2020). 'Array programming with NumPy', *Nature* 585, pp. 357–362. https://doi.org/10.1038/s41586-020-2649-2

Haspelmath, Martin (1993). *A Grammar of Lezgian.* Berlin: Mouton de Gruyter.

Haspelmath, Martin (1995a). 'The converb as a cross-linguistically valid category', Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: Mouton de Gruyter, pp. 1-56.

Haspelmath, Martin (1995b). 'Contextual and specialized converbs in Lezgian', Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: Mouton de Gruyter, pp. 415-440.

Haspelmath, Martin (2009). 'Terminology of case', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 505-517.

Haspelmath, Martin & König, Ekkehard (1998). 'Concessive conditionals in the languages of Europe', Johan van der Auwera (ed.), *Adverbial Constructions in the Languages of Europe*. Berlin: Mouton de Gruyter, pp. 563-640.

Haspelmath, Martin & Sims, Andrea D. (2010). *Understanding morphology*. 2nd ed. London: Routledge.

Hauk, Bryn. (2020). *Deixis and Reference Tracking in Tsova-Tush*. Diss. Honolulu: University of Hawai'i at Mānoa.

Heine, Bernd (2009). 'Grammaticalization of cases', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 458-469.

Henze, Paul B. (1992). 'Circassian Resistance to Russia', Marie B. Broxup (ed.), *The North Caucasus Barrier. The Russian Advance towards to Muslim World*. London: Hurst & Co., pp. 62-111.

Hewitt, George B. (1989) 'Abkhaz', George B. Hewitt (ed.), *The North West Caucasian Languages (The Indigenous Languages of the Caucasus 2).* Delmar, New York: Caravan Books, pp. 38-88.

Hewitt, B. G. (1995). *Georgian: A Structural Reference Grammar.* London Oriental and African Language Library, volume 2. Amsterdam/Philadelphia: John Benjamins.

Hewitt, George (2004). *Introduction to the Study of the Languages of the Caucasus.* München: LINCOM Europa.

Hewitt, George (2005). *Georgian: A Learner's Grammar.* 2nd ed. London: Routledge.

Hewitt, George B. (2010). *Abkhaz: A Comprehensive Self-tutor.* München: LINCOM Europa.

Hewson, John (2012). 'Tense', Robert I.Binnick (ed.), *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp. 507-535.

Himmelmann, Nikolaus P. (2014). 'Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference'. *Language* 90(4), pp 927–960.

Hockett, C. F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.

Höfler, Concha Maria, Böhm, Stefanie, Jungbluth, Konstanze & Skopeteas, Stavros (2016). 'Urum and Pontic Greeks: Communities and language situations', *STUF - Language Typology and Universals*, Vol. 69 (2), pp. 171-174.

Holisky, Dee Ann (1991). 'Laz', Alice C. Harris (ed.), *The Indigenous Languages of the Caucasus, volume 1*. The Kartvelian Languages. Delmar, New York: Caravan Books, pp. 395-472.

Holisky, Dee Ann & Gagua, Rusudan (1994). 'Tsova-Tush (Batsbi)', Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 147-212.

Hunter, J. D. (2007). 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95.

Ibragimov, Garun X. (1978). *Rutul'skij jazyk*. Maxačkala: Moscow: Akademija nauk SSSR.

Ibragimov, Garun X. (1990). *Caxurskij jazyk*. Moscow: Nauka.

Imnaišvili, David S. (1963). *Didojskij jazyk: v sravnenii s ginuxskim i xvaršijskim jazykami*. Tbilisi: Akad. nauk Gruzinskoj SSR.

Inkelas, Sharon (2011). 'The Interaction Between Morphology and Phonology', John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The Handbook of Phonological Theory*, 2nd Ed. Chichester: Blackwell Publishing Ltd, pp. 68-102.

Isaev, M.-Š. A. (2004). 'Dargwa', Michael Job (ed.), *The Indigenous Languages of the Caucasus, volume 3. The North East Caucasian Languages: Part 1*. Ann Arbor, Michigan: Caravan Books, pp. 299-346.

Isakov, Isak A. & Xalilov, Madžid S. (2012). *Gunzibskij jazyk*. Moscow: Maxačkala: Rossijskaja Akademija Nauk.

Ismailova, E. I. (2011). *Russko-rutul'skij slovar'*. Maxačkala: Institut jazyka, literatury i iskusstva im. G. Cadasy Dagestanskogo naučnogo centra RAN.

Jaimoukha, Amjad (2001). *The Circassians: a handbook*. London: Curzon Press.

Jakobson, Roman (1965). 'Quest for the Essence of Language', *Diogenes*, 13 (51), pp. 21-37.

Jakovlev, N. F. (1940). *Sintaksis čečenskogo literaturnogo jazyka*. Moscow: Akademija nauk SSSR.

Jakovlev, N. F. (1960). *Morfologija čečenskogo jazyka*. Groznyj: Čečeno-Ingušskoe knižnoe izdatel′stvo.

Jakovlev, N. F. (2001). *Sintaksis ingušskogo literaturnogo jazyka*. Jazyki Narodov Rossii. Moscow: Academia.

Johanson, Lars (1995). 'On Turkic converb clauses'. Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: Mouton de Gruyter, pp. 313-347.

Johanson, Lars (2018). 'Turkic indirectivity', Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*. Oxford: Oxford University Press, pp. 510-524.

Johanson, Lars (2022a). 'The Structure of Turkic', Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 26-59.

Johanson, Lars (2022b). 'The History of Turkic', Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 83-120.

Junker, Marie-Odile, Quinn, Conor M. & Valentine, Randolph J. (2018). 'Evidentiality in Algonquian', Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*. Oxford: Oxford University Press, pp. 431-462.

Kajaia, Otar (2001). *Megrul-kartuli leksik'oni*. Volume 1. Tbilisi: NEKERI.

Kajaia, Otar (2008). 'Mingrelisch', Heinz Fähnrich (ed.), *Kartwelsprachen*. Wiesbaden: Reichert.

Karakoç, Birsel (2005). *Das finite Verbalsystem im Nogaischen*. Wiesbaden: Harrassowitz Verlag.

Karakoç, Birsel (2022). 'Noghay', Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 354-369.

Karlsson, Fred (2018). *Finnish: A comprehensive grammar*. London: Routledge.

Károly, László (2022). 'West Old Turkic', Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 145-151.

Kazenin, Konstantin I. (2001). ‘Deepričastnye konstrukcii’, A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 554-594.

Kazenin, Konstantin I. (2013). *Sintaksis sovremennogo lakskogo jazyka*. Maxačkala: ALEF.

Kazenin, Konstantin (2020). ‘North Caucasus: Regions and their demography’, Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus*, 31. Oxford: Oxford University Press, pp. 67-84.

Kenesei, István, Vago, Robert M. & Fenyvesi, Anna (1998). *Hungarian*. Descriptive Grammars Series. London: Routledge.

Key, Mary Ritchie & Comrie, Bernard (eds.) (2023). *The Intercontinental Dictionary Series*. Leipzig: Max Planck Institute for Evolutionary Anthropology: https://ids.clld.org.

Khalilova, Zaira (2009). *A grammar of Khwarshi*. 234. Diss. Leiden: Rijksuniversiteit te Leiden.

Khalilova, Zaira (2011). ‘Evidentiality in Tsezic languages’, *Linguistic Diversity*, 9, pp. 30-48.

Khvtisiashvili, Tamrika (2013). *Principal aspects of Xinaliq phonology and morphosyntax*. Diss. Salt Lake City, Utah: University of Utah.

Kibrik, A. E. (1977). *Opyt strukturnogo opisanija arčinskogo jazyka*. Tom II, taksonomičeskaja grammatika. Publikacii otdelenija strukturnoj i prikladnoj lingvistiki 12. Moscow: Izdatel′stvo Moskovskogo Universiteta.

Kibrik, A. E. (1988). ‘The resultative in Archi’, Vladimir P. Nedjalkov (ed.) (1988). *Typology of resultative constructions*. Amsterdam: John Benjamins, pp. 167-184.

Kibrik, A. E. (1994a). ‘Archi’, Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 297-366.

Kibrik, A. E. (1994b). ‘Khinalug’, Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 367-406.

Kibrik, A. E. (1996) *Godoberi*. LINCOM Studies in Caucasian Linguistics 02. München: LINCOM.

Kibrik, A. E., Kazenin, K. I., Ljutikova, E. A., & Tatevosov, S. G. (eds.) (2001). *Bagvalinskij jazyk: Grammatika. Teksty. Slovari.* Moscow: Nasledie.

Kibrik, A. E., Kodzasov, S. V., & Olovjannikova, I. P. (1972). *Fragmenty grammatiki xinalugskogo jazyka*. Moscow: Izdatel′stvo Moskovskogo universiteta.

Kibrik, A. E. & Testelets, J. G. (1999). *Elementy caxurskogo jazyka v tipologičeskom osveščenii*. Moscow: Nasledie.

Kibrik, A. E. & Testelets, J. G. (2004). ‘Bezhta’, Michael Job (ed.), *The Indigenous Languages of the Caucasus, volume 3. The North East Caucasian Languages: Part 1*. Ann Arbor, Michigan: Caravan Books, pp. 217-298.

King, Charles (2008). *The Ghost of Freedom: A History of the Caucasus*. Oxford: Oxford University Press.

Kiparsky, Paul (2018). ‘Formal and empirical issues in phonological typology’, Larry M. Hyman & Frans Plank (eds.) *Phonological Typology*. Berlin: De Gruyter Mouton, pp. 54-106.

Klimov, G. A. (1965). *Kavkazskie jazyki*. Moscow: Nauka.

Klimov, G.A. (ed.) (1978). *Strukturnye obščnosti kavkazskix jazykov*. Moscow: Nauka.

Klimov, G. A. (1999). ‘Kartvelskie jazyki’ in M. E. Alekseev (ed.), *Jazyki mira. Kavkazskie jazyki*. Moscow: Academia, pp. 14-19.

Klyčev, Rauf N. (1995). *Slovar' sočetaemosti lokal'nyx preverbov s suffiksoidami i glagol'nymi kornjami v abazinskom jazyke*. Čerkessk: Karačaevo-čerkesskoe knižnoe izdatel′stvo.

Kodzasov, S. V. (1977). ‘Fonetika arčinskogo jazyka’, A. E. Kibrik, S. V. Kodzasov, I. P. Olovjannikova & D. S. Samedov (eds.), *Opyt strukturnogo opisanija arčinskogo jazyka*, vol. 1. Moscow: Izdatel'stvo Moskovskogo Universiteta, pp. 185–352.

Kodzasov, S. V. (2001). ‘Segmentnaja fonetika’, A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 35-43.

Koenig, Jean-Pierre & Michelson, Karin (2015). ‘Morphological complexity à la Oneida’, Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press, pp. 69–92.

Kolassa, John E. (2020). *An Introduction to Nonparametric Statistics*. New York: Chapman and Hall/CRC Press.

Komen, Erwin R., Molochieva, Zarina & Nichols, Johanna (2020). ‘Chechen and Ingush’, Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus*, 31. Oxford: Oxford University Press, pp. 317-368.

König, Ekkehard (1995). ‘The meaning of converb constructions’, Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: Mouton de Gruyter, pp.1-56.

Konuk, Mezane (2022). *Description grammatical de l’abzakh – tcherkesse occidental*. Diss. Paris: Université Sorbonne Nouvelle-Paris 3.

Koryakov, Y. B. (2002). *Atlas of Caucasian languages with language guide*. Moscow: Institute of Linguistcs, Russian Academy of Sciences.

Kozintseva, Natalia A. (2000). ‘Perfect forms expressing evidentiality’, Lars Johanson & Bo Utas (eds.), *Evidentials: Turkic, Iranian and Neighbouring Languages*. Berlin: Mouton de Gruyter, pp. 401-417.

Kuipers, Aert H. (1960). *Phoneme and morpheme in Kabardian (Eastern Adyghe)*. Janua Linguarum 8. s'Gravenhage: Mouton.

Kulikov, Leonid (2011). ‘Voice Typology’, Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, pp. 368-398.

Kumakhov, Mukhadin & Vamling, Karina (2009). *Circassian clause structure*. Malmö: Culture and Society, Department of International Migration and Ethnic Relations (IMER), Malmö University.

Kumaxov, M. A. (2006). *Kabardino-čerkesskij jazyk*. Volume I. Nal′čik: Izdatel′skij centr El′-Fa.

Kumaxov, M. A. (2013). *Kabardino-čerkesskij jazyk*. Moscow: Inst. Jazykoznanija, RAN.

Kurdiani, Mixeil (2016). *Osnovy iberijsko-kavkazskogo jazykoznanija*. Tbilisi: Izdatel'stvo Tbilisskogo gosudarstvennogo universiteta.

Kustova, Marina (2019). 'General converbs in Mehweb', Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 255-270.

Kutscher, Silvia (2001). *Nomen und nominales Syntagma im Lasischen: Eine deskriptive Analyse des Dialekts von Ardeşen*. LINCOM Studies in Caucasian Linguistics 17. München: LINCOM.

Kutscher, Silvia (2008). 'The language of the Laz in Turkey: Contact-induced change or gradual language loss?', *Turkic Languages* 12(1), pp 82-102.

L'Huillier, Monique. (1999) *Advanced French Grammar*. Cambridge: Cambridge University Press.

Lacroix, René (2009). *Description du dialecte laze d'Arhavi (caucasique du sud, Turquie): Grammaire et textes*. Diss. Lyon: Université Lumière Lyon 2.

Lacroix, René (2018). 'Laz', Geoffrey Haig & Geoffrey Khan (eds.), *The Languages and Linguistics of Western Asia*, 830-850. Berlin: DeGruyter Mouton.

Larsson, Filip (2013). *Two Millennia of Lexical and Typological Change in Western Europe - a quantitative geographical approach*. MA thesis, Lund University.

Lauer, M. (1869). *Grammatik der classischen Armenischen Sprache*. Wien: Wilhelm Braumüller.

Lee, Sooman Noah (2008). *A Grammar of Iranian Azerbaijani*. Seoul: Thaehaksa.

Lehmann, Christian (2015). *Thoughts on grammaticalization.* 3rd ed. Berlin: Language Science Press.

Loladze, Nika (2016). 'Greeks in Georgia: Main factors and motivations of emigration', *STUF - Language Typology and Universals*, Vol. 69 (2), pp. 175-192.

Lomtadze, Amiran (1987). *Brunebis taviseburebata ist'oriisatvis megrulši.* Tbilisi: Tbilisis universit'et'is gamomcemloba.

Lomtadze, Tamari & Enoch Reuven (2019). 'Judeo-Georgian Language as an Identity Marker of Georgian Jews (The Jews Living in Georgia)'. *Journal of Jewish Languages*. 2019; 7(1), pp. 1-26.

Lomtatidze, Ketevan & Klychev, Rauf (1989). 'Abaza', George B. Hewitt (ed.), *The Indigenous Languages of the Caucasus, volume 2. The North West Caucasian Languages*. Delmar, New York: Caravan Books, pp. 89-154.

Luraghi, Silvia (2009). 'Case in cognitive grammar', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 136-150.

Luraghi, Silvia & Huumo, Tuomas (2014). *Partitive Cases and Related Categories*. Berlin: De Gruyter Mouton.

Luraghi, Silvia & Kittilä, Seppo (2014). ‘Typology and diachrony of partitive case markers’, Silvia Luraghi & Tuomas Huumo (eds.), *Partitive Cases and Related Categories*, Berlin: De Gruyter Mouton, pp. 17-62.

Lyons, Christopher (1999). *Definiteness*. Cambridge: Cambridge University Press.

Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Magomedbekova, Zagidat M. (1967). *Axvaxskij jazyk: Grammatičeskij analiz, teksty, slovar'*. Tbilisi: Izdatel′stvo Akademii Nauk Gruzinskoj SSR.

Magomedbekova, Zagidat. M. (1971). *Karatinskij jazyk. Grammatičeskij analiz, teksty, slovar'.* Tbilisi: Izdatel′stvo Akademii Nauk Gruzinskoj SSR.

Magomedova, Patimat T. (2003). *Tindinsko-russkij slovar'*. Maxačkala: Institut jazyka, literatury i iskusstva im. G. Cadasy.

Magomedova, Patimat T. (2004). ‘Chamalal’, Michael Job (ed.), *The Indigenous Languages of the Caucasus, volume 3. The North East Caucasian Languages: Part 1*. Ann Arbor, Michigan: Caravan Books, pp. 3-68.

Magomedova, Patimat T. (2012). *Tindinskij jazyk*. Maxačkala: Institut jazyka, literatury i iskusstva im. G. Cadasy.

Magomedova, Patimat T. & Abdulaeva, I. A. (2007) *Axvaxsko-russkij slovar'*. Maxačkala: Institut jazyka, literatury i iskusstva im. G. Cadasy.

Magometov, Aleksandr A. (1963). *Kubačinskij jazyk: Issledovanie i teksty*. Tbilisi: Izdatel′stvo Akademii Nauk Gruzinskoj SSR.

Magometov, Aleksandr A. (1970). *Agul'skij jazyk: Issledovanie i teksty*. Tbilisi: Mecniereba.

Magometov, Aleksandr A. (1982). *Megebskij dialekt darginskogo jazyka: Issledovanie i teksty*. Tbilisi: Mecniereba.

Maiden, Martin & Robustelli, Cecilia (2013). *A reference grammar of modern Italian*. 2nd ed. London: Routledge.

Maisak, Timur (2018). ‘The Aorist/Perfect Distinction in Nizh Udi’, Diana Forker & Timur Maisak, *The Semantics of Verbal Categories in Nakh-Dagestanian Languages*. Leiden: Brill, pp. 120-166.

Maisak, Timur A. & Tatevosov, S. G. (2001a). ‘Vremja’, A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 273-292.

Maisak, Timur A. & Tatevosov, S. G. (2001b). ‘Evidencial′nost′’, A. E. Kibrik, K. I. Kazenin, E. A. Ljutikova & S. G. Tatevosov (eds.), *Bagvalinskij jazyk: Grammatika. Teksty. Slovari*. Moscow: Nasledie, pp. 293-318.

Makharoblidze, Tamar (2012). *The Georgian verb*. München: LINCOM.

Malchukov, Andrej (2015). ‘Valency classes and alternations: parameters of variation’, Andrej Malchukov & Bernard Comrie (eds.), *Valency Classes in the World’s Languages. Volume 1: Introducing the Framework, and Case Studies from Africa and Eurasia*. Berlin: De Gruyter Mouton, pp. 73-130.

Marr, N. & Brière, M. (1931). *La langue géorgienne*. Paris: Firmin-Didot.

Martínez-Celdrán, E., Fernández-Planas, A.Ma. & Carrera-Sabaté, J. (2003) 'Castilian Spanish', *Journal of the International Phonetic Association*, 33 (2), pp. 255-259.

Mauri, Caterina & Sansò, Andrea (2016). 'The Linguistic Marking of (Ir)Realis and Subjunctive', Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, pp. 166-195.

Maxmudova, S. M. (2001). *Morfologija rutul'skogo jazyka*. Moscow: Rossijskaja Akademija Nauk.

McKinney, Wes (2010). 'Data Structures for Statistical Computing in Python', *Proceedings of the 9th Python in Science Conference*, Volume 445, pp. 56-61.

Meillet, Antoine (1936). *Esquisse d'une grammaire comparée de l'arménien classique*. Wien: Imprimerie des PP. Mekhitharistes.

Miestamo, Matti (2008). 'Grammatical complexity in a cross-linguistic perspective', Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language Complexity. Typology, contact, change*. Amsterdam: John Benjamins, pp. 23-42.

Miestamo, Matti (2014). 'Partitives and negation: A cross-linguistic survey', Silvia Luraghi & Tuomas Huumo (eds.), *Partitive Cases and Related Categories*, Berlin: De Gruyter Mouton, pp. 63-88.

Miller, B. V. (1953). *Talyšskij jazyk*. Moscow: Akademia Nauk SSSR.

Molochieva, Zarina (2010). *Tense, aspect, and mood in Chechen*. Diss. Leipzig: Universität Leipzig.

Molochieva, Zarina & Nichols, Johanna (2018). 'Tense, Aspect, Mood and Evidentiality in Chechen and Ingush', Diana Forker & Timur Maisak, *The Semantics of Verbal Categories in Nakh-Dagestanian Languages*. Leiden: Brill, pp. 26-48.

Moran, Steven & Blasi, Damián (2014). 'Cross-linguistic comparison of complexity measures in phonological systems', Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring Grammatical Complexity*. Oxford: Oxford University Press, pp. 217-240.

Moran, Steven & McCloy, Daniel (eds.) (2019). *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. Available online: http://phoible.org

Moroz, George (2019). 'Phonology of Mehweb', Daniel, Michael, Nina Dobrushina & Dmitry Ganenkov (eds.). 2019. *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 17-38.

Morwood, James (2002). *Oxford Grammar of Classical Greek*. Oxford: Oxford University Press.

Murkelinskij, Gadži B. (1971). *Grammatika lakskogo jazyka*. Maxačkala: Dagestanskoe Učebno-pedagogičeskoe Izdatel'stvo.

Musaev, Magomed-Said (2002). *Darginskij jazyk*. Moscow: Nauka.

Mutalov, Rasul (2018). 'The Tense/Aspect System of Standard Dargwa', Diana Forker & Timur Maisak, *The Semantics of Verbal Categories in Nakh-Dagestanian Languages*. Leiden: Brill, pp. 49-79.

Mzamane, Godfred Isaac Malunga (1962). *A comparative phonetic and morphological study of the dialects of southern Nguni including lexical influences of non-Bantu languages*. Diss. University of South Africa.

Nakayama, Toshihide (2001). *Nuuchahnulth (Nootka) Morphosyntax*. University of California Publications in Linguistics 134. Berkeley: University of California Press.

National Statistics Office of Georgia (2014). *2014 General Population Census Results*. Tbilisi: Geostat: https://www.geostat.ge/en/modules/categories/739/demographic-and-social-characteristics.

Nedjalkov, Vladimir P. (1995). 'Some typological parameters of converbs', Martin Haspelmath & Ekkehard König (eds.), *Converbs in Crosslinguistic Perspective*. Berlin: Mouton de Gruyter, pp. 97-136.

Nedjalkov, Vladimir P. & Jaxontov, Sergej E. (1988). 'The Typology of Resultative Constructions', Vladimir P. Nedjalkov (ed.) (1988). *Typology of resultative constructions*. Amsterdam: John Benjamins, pp. 3-62.

Nichols, Johanna (1994). 'Chechen', Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 1-78.

Nichols, Johanna (2011). *Ingush Grammar*. Berkeley: University of California Press.

Nichols, Johanna (2020). 'Canonical complexity', Peter Arkadiev & Francesco Gardani (eds.), *Morphological Complexity*. Oxford: Oxford University Press, pp. 163-192.

Nichols, Johanna & Vagapov, A. D. (2004). *Chechen-English and English-Chechen dictionary*. London: Routledge-Curzon.

Nikolaev, Dmitry, Nikulin, Andrey & Kukhto, Anton (2015). *The database of Eurasian phonological inventories*. Available online: https://eurphon.info/

Nikolaeva, Irina (2016). 'Analyses of the Semantics of Mood', Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, pp. 68-88.

Noonan, Michael (2007). 'Complementation', Timothy Shopen (ed.), *Language Typology and Syntactic Description. Volume II: Complex constructions*. Cambridge: Cambridge University Press, pp. 52-150.

Nuyts, Jan (2016). 'Surveying Modality and Mood: An Introduction', Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, pp. 1-8.

O'Herin, Brian (2002). *Case and Agreement in Abaza*. Dallas, Texas: SIL International.

O'Herin, Brian (2020). 'Abaza and Abkhaz', Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 447-488.

Öztürk, Balkız & Pöchtrager, Markus A. (2011). *Pazar Laz*. Languages of the World/Materials, 484. München: LINCOM Europa.

Palmer, F. R. (2001). *Mood and Modality*. 2nd ed. Cambridge: Cambridge University Press.

Paris, Catherine (1989). 'West Circassian (Adyghe: Abzakh Dialect)', George B. Hewitt (ed.), *The Indigenous Languages of the Caucasus, volume 2. The North West Caucasian Languages*. Delmar, New York: Caravan Books, pp. 155-260.

Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, Vanderpals, Jake, Passos, Alexandre, Cournapeau, David, Brucher, Matthieu, Perrot, Matthieu & Duchesnay, Édouard (2011). 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research,* 12, pp. 2825-2830.

Polinsky, Maria (2015). *Tsez Syntax: A Description*. Manuscript. Harvard University, Cambridge, MA.

Polinsky, Maria (ed.) (2020). *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press.

Ragagnin, Elisabetta (2022). 'Azeri', Lars Johanson & Éva Á. Csató (eds.), *The Turkic Languages*. 2nd ed. London: Routledge, pp. 242-261.

Reseck, Tamar (2015). *Präverbien im Megrelischen*. Diversitas Linguarum, vol. 37. Bochum: Universitätsverlag Dr. N. Brockmeyer.

Richmond, Walter (2013). *The Circassian genocide*. New Brunswick, New Jersey: Rutgers University Press.

Rind-Pawlowski, Monika (2023). 'Verbal Roots and Verbal Stems in Khinalug', *Millennium*. Vol. 1. Tbilisi: Academy of Digital Humanities – Georgia, pp. 87-124.

Ringe, Donald A. (2017). *From Proto-Indo-European to Proto-Germanic*. 2nd edition. Oxford: Oxford University Press.

Ritz, Marie-Eve (2012). 'Perfect tense and aspect', Robert I.Binnick (ed.), *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp. 881-907.

Rogava, G. V. & Keraševa, Z. I. (1966). *Grammatika adygejskogo jazyka*. Majkop: Krasnodarskoe knižnoe izdatel′stvo.

Rostovtsev-Popiel, Alexander (2016). 'Kartvelian Preverbs in Cross-Linguistic Perspective', VII International Symposium on Kartvelian Studies.

Rostovtsev-Popiel, Alexander (2020). 'Megrelian', Maria Polinsky (ed.), *The Oxford Handbook of Languages of the Caucasus*. Oxford: Oxford University Press, pp. 529-572.

Saadiev, Š. M. (1994). 'Kryts', Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 407-446.

Šagirov, A. K. (1999). 'Abxazsko-adygskie jazyki', M. E. Alekseev (ed.) *Jazyki mira. Kavkazskie jazyki*. Moscow: Academia, pp. 80-85.

Sagot, Benoît (2013). 'Comparing complexity measures'. Paper presented at the Workshop 'Computational Approaches to Morphological Complexity', 22 February, 2013, Paris.

Saidova, Patimat A. (2004). 'Ghodoberi', Michael Job (ed.), *The Indigenous Languages of the Caucasus, volume 3. The North East Caucasian Languages: Part 1*. Ann Arbor, Michigan: Caravan Books, pp. 69-114.

Saidova, Patimat A. (2006) *Godoberinsko-russkij slovar'*. Maxačkala: Institut jazyka, literatury i iskusstva im. G. Cadasy.

Salimov, X. S. (2010). *Gagatlinskij govor andijskogo jazyka*. Maxačkala: IYALI.

Sapir, Edward (1921). *Language*. New York: Harcourt, Brace and World.

Sapir, Edward & Hoijer, Harry (1967). *The Phonology and Morphology of the Navaho Language*. University of California Publications in Linguistics 50. Berkeley: University of California Press.

Savelyev, Alexander & Robbeets, Martine (2020). 'Bayesian phylolinguistics infers the internal structure and the time-depth of the Turkic language family'. *Journal of Language Evolution* 5, pp. 39–53.

Schmidt, Karl Horst (1991). 'Svan', Alice C. Harris (ed.), *The Indigenous Languages of the Caucasus, volume 1*. The Kartvelian Languages. Delmar, New York: Caravan Books, pp. 473-556.

Schmitt, Rüdiger (2007). *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. Innsbrucker Beiträge zur Sprachwissenschaft 32. 2nd edn. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Schönig, Claus (1998). 'Azerbaijanian', Lars Johanson & Éva Ágnes Csató (eds.), *The Turkic Languages*. London: Routledge, pp. 248-260.

Schulze, Wolfgang (1982). *Die Sprache der Uden in Nord-Azerbajdžan: Studien zur Synchronie und Diakronie einer Süd-ostkaukasischen Sprache*. Wiesbaden: Otto Harrassowitz.

Schulze, Wolfgang (1997). *Tsakhur*. Languages of the World/Materials, 133. München: LINCOM Europa.

Schulze, Wolfgang (2000). *Northern Talysh*. Languages of the World/Materials, 380. München: LINCOM Europa.

Schulze, Wolfgang (2007). 'The Lak Language: A Quick Reference'. Manuscript.

Schulze-Fürhoff, Wolfgang (1994). 'Udi', Rieks Smeets (ed.), *The Indigenous Languages of the Caucasus, volume 4. North East Caucasian Languages Part 2*. Delmar, New York: Caravan Books, pp. 447-514.

Seegmiller, Steve (1996). *Karachay*. München, LINCOM Europa.

Shanidze, Akaki (1976). *Dzveli kartuli enis gramat'ik'a*. Tbilisi: Tbilisi University Press.

Shanidze, Akaki (1980). *Kartuli enis gramat'ik'is sapudzvlebi*. Tbilisi: Tbilisi University Press.

Shanidze, Akaki (1982). *Altgeorgisches Elementarbuch. 1. Theil. Grammatik der altgeorgischen Sprache*. Tbilisi: Tbilisi University Press.

Sheyanova, Maria (2019). 'Specialized converbs in Mehweb', Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Languages of the Caucasus 1. Berlin: Language Science Press, pp. 235-254.

Shosted, Ryan K., Chikovani, Vakhtang (2006). 'Illustrations of the IPA: Standard Georgian', *Journal of the International Phonetic Association*, 36. Cambridge: Cambridge University Press, pp. 255–264.

Širaliev, M. Š. (1971). *Grammatika azerbajdžanskogo jazyka: fonetika, morfologija i sintaksis*. Baku: Elm.

Smeets, Henricus Joannes (1984). *Studies in West Circassian phonology and morphology*. Leiden: Hakuchi Press.

Smolina, Marija (2008). *Urumskij jazyk: priazovskij variant*. Kyiv: Blank-Press.

Snyman, Jannie W. (1970). *An Introduction to the !Xũ (!Kung) Language*. Cape Town: University of Cape Town.

Song, Jae Jung (2013). 'Nonperiphrastic Causative Constructions', Matthew S. Dryer & Martin Haspelmath (eds.), *WALS Online* (v2020.3): https://wals.info.

Spencer, Andrew (2009). 'Case as a morphological phenomenon', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 185-199.

Spencer, Andrew & Luís, Ana R. (2012). *Clitics: An Introduction*. Cambridge: Cambridge University Press.

Spruit, Arie (1986). *Abkhaz studies*. Diss. Leiden: Rijksuniversiteit te Leiden.

Squartini, Mario (2016). 'Interactions between Modality and Other Semantic Categories', Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, pp. 50–67.

State Statistical Committee (2019). *Population census 2019*. Baku: State Statistical Committee of the Republic of Azerbaijan: https://www.stat.gov.az/?lang=en

Steele, Susan (1995). 'Towards a theory of morphological information', *Language* 71, pp. 260-309.

Stilo, Donald (2009). 'Case in Iranian: From Reduction and Loss to Innovation and Renewal', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 700-715.

Stilo, Donald (2015). 'The Polygenetic Origins of the Northern Talyshi Language', Uwe Bläsing, Victoria Arakelova & Matthias Weinreich (eds.), *Studies on Iran and The Caucasus*. Leiden: Brill, pp. 411-454.

Stolz, Thomas, Stroh, Cornelia & Urdze, Aina (2009). 'Varieties of comitative', Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. Oxford: Oxford University Press, pp. 601-608.

Story, Gillian L. (1979). *A morphological study of Tlingit*. Dallas, Texas: Summer Institute of Linguistics.

Stump, Gregory T. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. (2019). 'Paradigm Function Morphology', Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*. Oxford: Oxford University Press, pp. 285-304.

Sulejmanov, N. D. (2000). 'Axvaxskij jazyk', V. P. Neroznak (ed.), *Jazyki Dagestana*. Jazyki Narodov Rossii. Maxačkala/Moscow: Rossijskaja Akademija nauk, pp. 306-318.

Suleymanov, Murad (2020). *A Grammar of Şirvan Tat*. Wiesbaden: Reichert.

Sumbatova, Nina (2020). 'Dargwa', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus*, 31. Oxford: Oxford University Press, pp. 147-200.

Sumbatova, Nina R. & Lander, Yury A. (2014). *Darginskij govor selenija Tanty: grammatičeskij očerk, voprosy sintaksisa*. Moscow: Jazyki slavjanskoj kultury.

Sumbatova, Nina R. & Mutalov, Rasul O. (2003). *A Grammar of Icari Dargwa*. München: LINCOM Europa.

Tabulova, N. T. (1976). *Grammatika abazinskogo jazyka: fonetika i morfologija*. Čerkessk: Karačaevo-Čerkesskoe otdelenie Stavropols′kogo knižnogo izdatel′stva.

Talibov, Bukar B. (2004). 'Tsakhur', Michael Job (ed.), *The Indigenous Languages of the Caucasus, volume 3. The North East Caucasian Languages: Part 1*. Ann Arbor, Michigan: Caravan Books, pp. 347-419.

Talibov, Bukar B. (2007). *Buduxskij jazyk*. Moscow: Akademia.

Temirbulatova, S. M. (2004). *Xajdakskij dialekt darginskogo jazyka*. Diss. Maxačkala: Izdatel′stvo tipografii DNC RAN.

Testelets, Yakov G. (2020). 'Kartvelian (South Caucasian) Languages', Maria Polinsky (ed.), *Handbook of the Languages of the Caucasus*, 31. Oxford: Oxford University Press, pp. 491-529.

Thordarson, Fridrik (2009). *Ossetic grammatical studies*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Timberlake, Alan (2007). 'Aspect, tense, mood', Timothy Shopen (ed.) *Language Typology and Syntactic Description*. Volume III: Grammatical Categories and the Lexicon. Cambridge: Cambridge University Press, pp. 280-333.

Trubetzkoy, N. S. (1928). 'Proposition 16', *Actes du 1er Congrès international de linguistes*, Leiden: A. W. Sijthoff's, pp. 17-18.

Tsagareli, A. (1880). *Mingrel'skie etjudy. Pervyj vypusk. Mingrel'skie teksty s perevodom i objasnenijami*. St. Petersburg: Izd. Imperatorskoj akademii nauk.

Tschenkéli, Kita (1958). *Einführung in die georgische Sprache*. Band I: Theoretischer Teil. Zürich: Amirani Verlag.

Tuite, Kevin (1997). *Svan*. Languages of the World/Materials, 139. München: LINCOM Europa.

Tuite, Kevin (1998). *Kartvelian morphosyntax: number agreement and morphosyntactic orientation in the South Caucasian languages*. München: LINCOM Europa.

Tuite, Kevin (2008a). 'Early Georgian', Roger D. Woodard (ed.), *The Ancient Languages of Asia Minor*. Cambridge: Cambridge University Press, pp. 145-165.

Tuite, Kevin (2008b). 'The Rise and Fall and Revival of the Ibero-Caucasian Hypothesis', *Historiographia Linguistica* XXXV:1/2. John Benjamins, pp. 23-82.

Tuite, Kevin (2017). 'Language and Emergent Literacy in Svaneti', Ramazan Kormaz & Gürkan Doğan (eds.), *Endangered Languages of the Caucasus and Beyond*. Languages of Asia, volume 5. Leiden: Brill, pp. 226-243.

Tuite, Kevin (2024). 'Applicative constructions in Kartvelian', Fernando Zúñiga & Denis Creissels (eds.), *Applicative Constructions in the World's Languages*. Berlin: De Gruyter Mouton, pp. 913-942.

Ulakov, M. Z. & Guseev, Ž. M. (2016). *Sovremennyj karačaevo-balkarskij jazyk*. Nal'čik: Pečatnyj Dvor.

Uslar, Petr K. (1888). *Etnografija Kavkaza. Jazykoznanie, II. Čečenskij jazyk*. Tbilisi: Upravlenie Kavkazskago Učebnago Okruga.

Uslar, Petr K. (1890). *Lakskij jazyk.* Etnografija Kavkaza 4. Tbilisi: Izdanie Upravlenija Kavkazskago Učebnago Okruga.

Väänänen, Veikko (1981). *Introduction au Latin vulgaire*. Paris: Klincksieck.

Vamling, Karina (1989). *Complementation in Georgian*. Travaux de l'institut de linguistique de Lund, 23. Lund: Lund University Press.

Vamling, Karina & Tchantouria, Revaz (1991). 'Sketch of the Grammar of Kubachi: The Simple Sentence', *Lund University, Dept. of Linguistics / Working Papers 38.*, pp. 221-236.

Vamling, Karina & Tchantouria, Revaz (1993). 'On subordinate clauses in Megrelian', K. Hengeveld (ed.), *The internal structure of adverbial clauses*. Eurotyp Working Papers V, pp. 67-86.

Van Damme, Dirk (2004). *Altarmenische Kurzgrammatik*. Göttingen: Vandenhoeck & Ruprecht.

van den Berg, Helma (1995). *A grammar of Hunzib: with texts and lexicon*. LINCOM Studies in Caucasian Linguistics, 01. München: LINCOM Europa.

van den Berg, Helma (2001). *Dargi folktales: oral stories from the Caucasus and an introduction to Dargi grammar*. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), Universiteit Leiden.

van der Auwera, Johan (1998). 'Defining converbs', Leonid I. Kulikov & Heinz Vater (eds.), *Typology of verbal categories*. Tübingen: Niemeyer, pp. 273–282.

Van Olmen, Daniël & van der Auwera, Johan (2016). 'Modality and mood in Standard Average European', Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, pp. 362–384.

Vaux, Bert (2007). 'Homshetsma: The language of the Armenians of Hamshen', Hovann H. Simonian (ed.), *The Hemshin: History, society and identity in the Highlands of Northeast Turkey*. London: Routledge, pp. 257-278.

Virtanen, P., Gommers, R., Oliphant, T.E. et al. (2020). 'SciPy 1.0: fundamental algorithms for scientific computing in Python', *Nature Methods*, 17, pp. 261–272.

Vogt, Hans (1936). 'Esquisse d'une grammaire du géorgien moderne', *Norsk tidsskrift for sprogvidenskap*, 9–10. Oslo: A.W. Brøggers Boktrykkeri.

Vogt, Hans (1963). *Dictionnaire de la langue oubykh*. Oslo: Universitetsforlaget.

Vogt, Hans (1971). *Grammaire de la langue géorgienne*. Instituttet for Sammenlignende Kulturforskning. Ser. B, Skrifter 57. Oslo: Universitetsforlaget.

Voßen, Rainer (1997). *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Quellen zur Khoisan-Forschung/Research in Khoisan Studies 12. Köln: Rüdiger Köppe.

Wade, Terence, Gillespie, David, Gural, Svetlana & Korneeva, Marina (2020). *A comprehensive Russian grammar*. 4th ed. Oxford: Wiley-Blackwell.

Waskom, M. L., (2021). 'seaborn: statistical data visualization', *Journal of Open Source Software*, 6(60), 3021. https://doi.org/10.21105/joss.03021.

Weiss, Daniel (1995). ‘Russian converbs: A typological outline’, Martin Haspelmath & Ekkehard König (eds.), *Converbs in Crosslinguistic Perspective*. Berlin: Mouton de Gruyter, pp. 239-282.

Wichers Schreur, Jesse (2024). *Intense language contact in Eastern Georgia: The case of Tsova-Tush*. Diss. Frankfurt am Main: Goethe-Universität.

Woodard, Roger D. (2008). ‘Attic Greek’, Roger D. Woodard (ed.), *The Ancient Languages of Europe*. Cambridge: Cambridge University Press, pp. 14-49.

Xalilov, Madžid S. (1995). *Beštinsko-russkij slovar'*. Maxačkala: Tsadasa Institute of Language, Literature, and Arts.

Xajdakov, S. M. (1999). ‘Lakskij jazyk’, M. E. Alekseev (ed.), *Jazyki mira. Kavkazskie jazyki*. Moscow: Academia, pp. 347-357.

Yousef, Saeed (2018). *Persian: A comprehensive grammar*. London: Routledge.

Žirkov, Lev Ivanovič (1955). *Lakskij jazyk: fonetika i morfologija*. Moscow: Nauka.

Zúñiga, Fernando & Kittilä, Seppo (2019). *Grammatical Voice*. Cambridge: Cambridge University Press.

Zwicky, Arnold M. (1985). ‘How to describe inflection’, Mary Niepokuj, Mary Van Clay, Vassiliki Nikiforidou & Deborah Feder (eds.), *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*. Berkeley Linguistics Society, pp. 372-386.

Zwicky, Arnold M. (1994). ‘What is a clitic?’, Joel A. Nevis, Briand D. Joseph, Dieter Wanner & Arnold M. Zwicky (eds.), *Clitics: A Comprehensive Bibliography 1892-1991*. Amsterdam: John Benjamins, pp. xii-xx.

Zwicky, Arnold M. & Pullum, Geoffrey K. (1983). ‘Cliticization vs. Inflection: English *n’t*’, *Language*, 59, pp. 502-513.

Appendix A: Table of languages with sources (affixal data)

Language family	Family branch	Language	Sources
Kartvelian	Karto-Zan	Georgian	Vamling 1989; Aronson 1990; Hewitt 1995; Tuite 1998; Bolkvadze & Kiziria 2023
		Old Georgian	Shanidze 1982; Fähnrich 1991; Fähnrich 1994; Tuite 2008a; Fähnrich 2012
		Megrelian	Harris 1991; Vamling & Tchantouria 1993; Reseck 2015; Rostovtsev-Popiel 2020
		Laz	Anderson 1963; Holisky 1991; Lacroix 2009; Öztürk & Pöchtrager 2011; Lacroix 2018
	Svan	Svan	Gudjedjiani & Palmaitis 1986; Schmidt 1991; Tuite 1997
Nakh-Dagestanian	Nakh	Chechen	Jakovlev 1940; Jakovlev 1960; Čokaev 1970; Nichols 1994; Aliroev 1999; Nichols & Vagapov 2004; Molochieva & Nichols 2018; Komen, Molochieva & Nichols 2020
		Ingush	Jakovlev 2001; Nichols 2011; Komen, Molochieva & Nichols 2020
		Bats	Dešeriev 1953; Holisky & Gagua 1994; Hauk 2020
	Avar-Andic	Avar	Charachidzé 1981; Alekseev et al. 2012; Forker 2018b; Forker 2020
		Andi	Salimov 2010
		Tindi	Magomedova 2012
		Bagvalal	Kibrik et al. 2001
		Chamalal	Bokarev 1949a; Magomedova 2004
		Karata	Magomedbekova 1971
		Akhvakh	Magomedbekova 1967; Creissels 2008; Creissels 2009; Creissels 2010, Creissels 2018
		Ghodoberi	Kibrik 1996; Saidova 2004
	Lezgic	Lezgian	Haspelmath 1993
		Tabasaran	Alekseev & Shixalieva 2003; Babaliyeva 2013
		Rutul	Ibragimov 1978; Alekseev 1994a; Maxmudova 2001
		Aghul	Magometov 1970
		Tsakhur	Ibragimov 1990; Kibrik & Testelets 1999; Schulze 1997; Talibov 2004
		Udi	Schulze 1982; Schulze-Fürhoff 1994; Harris 2002; Alekseev et al. 2008; Ganenkov 2008; Maisak 2018
		Kryts	Saadiev 1994; Authier 2009
		Budukh	Alekseev 1994b; Talibov 2007

Nakh-Dagestanian	Lezgic (continued)	Archi	Kibrik 1977; Kibrik 1994a; Chumakina, Bond & Corbett 2016
	Dargic	Standard Dargwa	Abdullaev 1954; Abdullaev 1971; Van den Berg 2001; Musaev 2002; Isaev 2004; Sumbatova 2020
		Xaidaq	Temirbulatova 2004
		Kubachi	Magometov 1963; Vamling & Tchantouria 1991
		Itsari Dargwa	Sumbatova & Mutalov 2003
		Sanzhi Dargwa	Forker 2020a
		Mehweb	Magometov 1982; Daniel, Dobrushina & Ganenkov 2019
	Tsezic	Tsez	Imnaišvili 1963; Alekseev & Radžabov 2004
		Khwarshi	Khalilova 2009
		Hinuq	Forker 2013
		Bezhta	Kibrik & Testelets 2004; Comrie, Khalilov & Khalilova 2015
		Hunzib	Isakov & Xalilov 2012; Berg 1995
	Lak	Lak	Žirkov 1955; Murkelinskij 1971; Friedman 1992; Schulze 2007; Kazenin 2013; Friedman 2020
	Khinalug	Khinalug	Kibrik 1994b; Khvtisiashvili 2013
Northwest Caucasian	Circassian	Kabardian	Colarusso 1992; Kumaxov 2006; Kumakhov & Vamling 2009; Arkadiev & Lander 2020
		Adyghe	Rogava & Keraševa 1966; Kumakhov & Vamling 2009; Arkadiev & Lander 2020
		Abzakh Adyghe	Paris 1989; Konuk 2022
		Shapsug Adyghe	Smeets 1984
	Abkhaz-Abaza	Abkhaz	Aristava 1968; Hewitt 1989; Chirikba 2003a; Hewitt 2010; O'Herin 2020
		Abaza	Genko 1955; Tabulova 1976; Lomtatidze & Klychev 1989; O'Herin 2020
	Ubykh	Ubykh	Dumézil 1931; Vogt 1963; Fenwick 2011
Indo-European	Armenian	Eastern Armenian	Dum-Tragut 2009
		Classical Armenian	Meillet 1936; Van Damme 2004; Schmitt 2007; Clackson 2008
	Iranian	Iron Ossetic	Abaev 1964; Bagaev 1965; Thordarson 2009; Erschler 2020
		Tat	Suleymanov 2020
		Juhuri	Authier 2012
		Talysh	Miller 1953; Schulze 2000
Turkic	Oghuz	North Azerbaijani	Širaliev 1971; Schönig 1998; Ragagnin 2022
		South Azerbaijani	Dehghani 2000; Lee 2008
	Kipchak	Karachay-Balkar	Filonenko 1940; Aliev 1973; Seegmiller 1996; Ulakov & Guseev 2016; Berta & Csató 2022
		Kumyk	Abdullaeva et al. 2014; Berta & Csató 2022
		Nogai	Csató & Karakoç 1998, Karakoç 2005, Karakoç 2022

Appendix B: Table of lexical data per language with sources

Family	Branch	Language	Nouns	Verbs	Other	Total	Sources
Kartvelian	Karto-Zan	Georgian	403	494	63	960	Fähnrich (2007); Carling (2024)
		Old Georgian	178	298	50	526	Fähnrich (2007); Carling (2024)
		Megrelian	335	459	51	845	Fähnrich (2007); Carling (2024)
		Laz	289	327	81	697	Fähnrich (2007); Carling (2024)
	Svan	Svan	321	320	49	690	Fähnrich (2007); Carling (2024)
Nakh-Dagestanian	Nakh	Chechen	147	207	0	354	Key & Comrie (2023); Carling (2024)
		Ingush	153	156	1	310	Key & Comrie (2023); Carling (2024)
		Bats	209	243	50	502	Key & Comrie (2023); Carling (2024)
	Avar-Andic	Avar	158	157	0	315	Key & Comrie (2023); Carling (2024)
		Andi	110	179	0	289	Key & Comrie (2023); Carling (2024)
		Tindi	289	186	0	475	Key & Comrie (2023)
		Bagvalal	292	151	0	443	Key & Comrie (2023)
		Chamalal	309	187	0	496	Key & Comrie (2023)
		Karata	275	155	0	430	Key & Comrie (2023)
		Akhvakh	296	176	0	472	Key & Comrie (2023)
		Ghodoberi	262	163	0	425	Key & Comrie (2023)
	Lezgic	Lezgian	207	189	47	443	Key & Comrie (2023); Carling (2024)
		Tabasaran	115	137	0	252	Key & Comrie (2023); Carling (2024)
		Rutul	122	83	0	205	Key & Comrie (2023); Carling (2024)
		Aghul	198	126	0	324	Key & Comrie (2023); Carling (2024)
		Tsakhur	134	148	0	282	Key & Comrie (2023); Carling (2024)
		Udi	137	155	0	292	Key & Comrie (2023); Carling (2024)

Nakh-Dagestanian	Lezgic (continued)	Kryts	132	172	0	304	Key & Comrie (2023); Carling (2024)
		Budukh	105	113	0	218	Key & Comrie (2023); Carling (2024)
		Archi	259	120	0	379	Key & Comrie (2023)
	Dargic	Standard Dargwa	175	203	0	378	Key & Comrie (2023); Carling (2024)
		Xaidaq	285	167	0	452	Key & Comrie (2023)
		Kubachi	238	127	0	365	Key & Comrie (2023)
		Itsari Dargwa	252	148	0	400	Key & Comrie (2023)
		Mehweb	290	147	0	437	Key & Comrie (2023)
	Tsezic	Tsez	276	139	0	415	Key & Comrie (2023)
		Khwarshi	117	165	0	282	Key & Comrie (2023); Carling (2024)
		Hinuq	296	145	0	441	Key & Comrie (2023);
		Bezhta	175	154	0	329	Key & Comrie (2023); Carling (2024)
		Hunzib	306	150	0	456	Key & Comrie (2023)
	Lak	Lak	128	99	0	227	Key & Comrie (2023); Carling (2024)
	Khinalug	Khinalug	100	179	0	279	Key & Comrie (2023); Carling (2024)
Northwest Caucasian	Circassian	Kabardian	204	149	17	370	Kumakhov & Vamling (2009); Dellert et al (2019); Carling (2024)
		Adyghe	358	194	36	588	Dellert et al (2019); Carling (2024)
	Abkhaz-Abaza	Abkhaz	213	288	52	553	Dellert et al (2019); Carling (2024)
		Abaza	104	35	3	142	O'Herin (2020); Carling (2024)
	Ubykh	Ubykh	158	128	2	288	Fenwick (2011); Carling (2024)
Indo-European	Armenian	Eastern Armenian	168	225	39	432	Key & Comrie (2023); Carling (2024)
		Classical Armenian	129	33	32	194	Carling (2024)
	Iranian	Iron Ossetic	108	235	42	385	Key & Comrie (2023); Carling (2024)
		Juhuri	389	122	0	511	Key & Comrie (2023)
		Persian	124	32	35	191	Carling (2024)
	Slavic	Russian	225	48	43	316	Carling (2024)
Turkic	Oghuz	North Azerbaijani	368	208	8	584	Key & Comrie (2023); Carling (2024)
	Kipchak	Karachay-Balkar	98	83	55	236	Savelyev & Robbeets (2020)
		Kumyk	356	221	4	581	Key & Comrie (2023); Carling (2024)
		Nogai	354	225	0	579	Key & Comrie (2023)

Appendix C: Table of number of phonemes per language

Language	C	V	Source
Abaza	61	2	O'Herin 2002
Abkhaz	59	3	Hewitt 2010
Abzakh Adyghe	60	3	Paris 1989
Adyghe	52	3	Rogava & Keraševa 1966
Aghul	52	5	Magometov 1970
Akhvakh	50	19	Magomedova & Abdulaeva 2007
Andi	47	10	Salimov 2010
Archi	69[257]	11	Chumakina, Bond & Corbett 2016
Avar	44	5	Forker 2020; Alekseev et al. 2012
Bagvalal	66	18	Kodzasov 2001
Bats	42	19[258]	Fähnrich 2001
Bezhta	34	30	Comrie, Xalilov & Xalilova 2015
Budukh	35	10	Alekseev 1994b
Chamalal	42	17	Magomedova 2004
Chechen	37	22	Komen, Molochieva & Nichols 2020
Classical Armenian	30	7	Schmitt 2007
Eastern Armenian	30	6	Dum-Tragut 2009
Georgian	28	5	Shosted & Chikovani 2006
Ghodoberi	43	13	Saidova 2006
Hinuq	42	12[259]	Forker 2013
Hunzib	35	23	Isakov & Xalilov 2012
Ingush	39	23	Nichols 2011
Iron Ossetic	38	7	Erschler 2020
Itsari Dargwa	55[260]	10	Sumbatova & Mutalov 2003
Juhuri	22	7	Authier 2012

[257] Kodzasov (1977) describes 81 consonants, but this includes the pharyngealised uvulars which should be analysed as a suprasegmental feature indicating stress (Kibrik 1994a: 303; Chumakina, Bond & Corbett 2016: 20-21).

[258] Hauk (2020) does not analyse neither the nasalised or reduced vowels as phonemes however.

[259] Not including pharyngealised vowels as they are described as optional (Forker 2013: 26).

[260] Including labialised velar and uvular consonants.

Kabardian	47[261]	3	Kumaxov 2006
Karachay-Balkar	29	8	Ulakov & Guseev 2016
Karata	45	10	Magomedbekova 1971
Khinalug	44	9	Khvtisiashvili 2013
Khwarshi	66	21	Khalilova 2009
Kryts	41[262]	4	Authier 2009
Kubachi	54	8	Magometov 1963
Kumyk	25	8	Abdullaeva et al. 2014
Lak	59[263]	6	Murkelinskij 1971
Laz	34	5	Öztürk & Pöchtrager 2011
Lezgian	54	8	Haspelmath 1993
Megrelian	31	6	Rostovtsev-Popiel 2020
Mehweb	41	9	Moroz 2019
Nogai	26	9	Karakoç 2022
North Azerbaijani	27	9	Ragagnin 2022
Old Georgian	31	5	Tuite 2008a
Rutul	62	12	Ibragimov 1978
Sanzhi Dargwa	54	7	Forker 2020a
Shapsug Adyghe	57	3	Smeets 1984
South Azerbaijani	22	9	Dehghani 2000
Standard Dargwa	37	5	Isaev 2004; Musaev 2002
Svan	31	18[264]	Tuite 1998
Tabasaran	55	7	Babaliyeva 2013
Talysh	22	9	Schulze 2000
Tat	26	10	Suleymanov 2020
Tindi	46	19	Magomedova 2012
Tsakhur	73[265]	11	Kibrik & Testelets 1999
Tsez	41[266]	20	Imnaišvili 1963
Ubykh	84	3	Fenwick 2011
Udi	38	15	Schulze-Fürhoff 1994
Xaidaq	63	5	Temirbulatova 2004

[261] The consonant inventory differs between the different Kabardian dialects, as Standard Kabardian is described as having 47 consonants, while Kuipers (1960) describes 49 consonant phonemes.

[262] Saadiev (1994) describes a larger inventory of 46 consonant phonemes and 9 vowels.

[263] Friedman (2020: 203-204) only describes 41 consonant phonemes as he does not include the labialised consonants.

[264] Upper Svan.

[265] Descriptions of the Tsakhur consonant inventory range from 62 to 88 consonant phonemes (Kibrik & Testelets 1999; Schulze 1997; Ibragimov 1990).

[266] Consonant and vowel phonemes for all dialects (Imnaišvili 1963).