

**THE INSTITUTE OF LINGUISTICS OF THE RUSSIAN ACADEMY OF SCIENCES**

*as a manuscript*

**Timur Anatolievich Maisak**

**GRAMMATICALIZATION IN NAKH-DAGHESTANIAN LANGUAGES:  
TYPOLOGICAL, DIACHRONIC AND AREAL PERSPECTIVES**

**Dissertation Summary**  
for the purpose of obtaining academic degree  
Doctor of Science in Philology and Linguistics

Moscow 2023

The dissertation was prepared at the Institute of Linguistics of the Russian Academy of Sciences.

### **Publications**

Ten publications were selected for the defense. Seven articles are indexed in the Scopus or Web of Science database; all seven articles are published in journals included in the first or second quartile of Scopus.

1) Maisak T. Morphological fusion without syntactic fusion: The case of the "verificative" in Agul // *Linguistics*. 2016a. 54(4). 815–870. (Scopus Q2, List A)

2) Maisak T. Subject pronoun doubling in Agul: Spoken corpus data on a rare discourse pattern // *Studies in Language*. 2016b. 40(4). 955–987. (Scopus Q1, List A)

3) Maisak T. Borrowing from an unrelated language in support of intragenetic tendencies: The case of the conditional clitic =sa in Udi // *Diachronica*. 2019a. 36(3). 337–383. (Scopus Q1, List A)

4) Maisak T. Repetitive prefix in Agul: Morphological copy from a closely related language // *International Journal of Bilingualism*. 2019b. 23(2). 486–508. (Scopus Q1, List A)

5) Maisak T. Relative clauses in Agul from a corpus-based perspective // *STUF – Sprachtypologie und Universalienforschung*. 2020. 73(1). 113–158. (Scopus Q2, List B)

6) Maisak T. Endoclitics in Andi // *Folia Linguistica*. 2021a. 55(1). 1–34. (Scopus Q2, List A)

7) Maisak T. Numeral classifiers in Udi: A unique contact-induced development among Nakh-Daghestanian? // *Journal of Language Contact*. 2021b. 14(2). 330–367. (Scopus Q2, List A)

8) Majsak T. A. Pričastnye formy v vido-vremennoj sisteme agul'skogo jazyka (in Russian, Participial forms in the tense-aspect system of Agul) // *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies*. 2012. VIII(2). 228–289. (List D)

9) Majsak T. A. Tipologičeskoe, vnutrigenetičeskoe i areal'noe v grammatikalizacii: dannye lezgin'skix jazykov (in Russian, Typological, intragenetic and areal dimensions of grammaticalization: the Lezgian languages) // *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies*. 2016a. XII(1). 588–618. (List D)

10) Majsak T. A. Pričastnye otnositel'nye predloženiya v udinskom jazyke po korpusnym dannym) (in Russian, Participial relative clauses in Udi from a corpus perspective) // *Tomsk Journal of Linguistics and Anthropology*. 2020. 2(28). 46–65. (List D)

The results of the present study have also been presented in the following papers:

1) Majsak T. A., Merdanova S. R. Buduščee vremja v agul'skom jazyke v tipologičeskoj perspektive (in Russian, The future tense in Aghul in a typological perspective) // *Voprosy Jazykoznanija*. 2003. № 6. 76–107. (Scopus Q4, List B)

2) Majsak T. A. K publikacii kavkazsko-albanskix palimpsestov iz Sinaj'skogo monastyrja (in Russian, On the publication of Caucasian Albanian palimpsests from the Mount Sinai Monastery) // *Voprosy Jazykoznanija*. 2010. № 6. 88–107. (Scopus Q4, List B)

3) Majsak T. A. Perfekt i Aorist v nižskom dialekte udinskogo jazyka (in Russian, The Perfect and the Aorist in the Nizh dialect of Udi) // *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies*. 2016b. XII(2). 315–378. (List D)

4) Majsak T. A., Merdanova S. R. Perfekt i smežnye značeniya v agul'skom jazyke (in Russian, Perfect and related meanings in Aghul) // *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies*. 2016c. XII(2). 379–422. (List D)

5) Majsak T. A. Soglasovanie po množestvennomu čislu v udinskom jazyke (in Russian, Plural agreement in Udi) // *Acta Linguistica Petropolitana*. Transactions of the Institute for Linguistic Studies. 2017. XIII(1). 680–724. (List D)

#### **Conference presentations and grants**

The main results of the present study have been presented in 2013–2022 in 16 oral presentations at 10 international conferences:

- Conference on Typology and Grammar for Young Scholars (2013);
- System changes in the languages of Russia (2014);
- Typology of Morphosyntactic Parameters (2013, 2017, 2022);
- Annual Meeting of the Berkeley Linguistics Society (2018);
- Annual Meeting of the Societas Linguistica Europaea (SLE) (2017, 2019, 2020);
- Causal constructions in the world's languages (synchrony, diachrony, typology) (2021);
- Conference of the Association for Linguistic Typology (2017, 2019);
- International Morphology Meeting (2018);
- Syntax of the World's Languages (2014, 2018);
- Tense, aspect, modality and evidentiality in Nakh-Daghestanian languages (2013).

The studies collected in the dissertation were supported by the grants

- from the Russian Science Foundation #14-18-02624, #14-18-02429, #18-78-10128,
  - from the Russian Foundation for Fundamental Research #18-012-00852,
  - from the Russian Humanitarian Science Foundation #13-04-00345,
- and by the Basic Research Program at the National Research University Higher School of Economics (HSE).

## 1. Introduction

The present dissertation comprises several papers that examine typological, diachronic and areal aspects of grammaticalization phenomena in the Nakh-Daghestanian languages, mostly in the verbal system. The studies are mainly based on data from the Agul and Udi languages of the Lezgian group and the Andi language of the Avar-Andic group, although other languages of the family are also taken into account for comparison.

Following leading scholars in the field, such as Christian Lehmann, Bernd Heine, Joan Bybee, Paul Hopper and Elizabeth Closs Traugott, among others, **grammaticalization** is understood, in the broadest sense, as the process by which grammatical structures are created or, in a more common wording, the process by which constructions typically including a lexical item become grammatical morphemes. Interest in the study of grammaticalization, which took shape as an independent sub-field of functional linguistics in the 1980s, does not subside now. In the last few years alone, a number of new general works within the framework of grammaticalization theory have appeared, including an extensively revised and updated edition of the *World lexicon of grammaticalization* (Kuteva et al. 2019), a new textbook on grammaticalization (Narrog, Heine 2021), and a number of fundamental collective volumes containing overviews of grammaticalization processes in particular language families (Narrog, Heine 2011; Narrog, Heine, 2018; Bisang, Malchukov 2020a, 2020b). The same is true of the field of **areal linguistics** and the processes associated with language contact, both between genealogically unrelated and related languages. This area of research has undergone a boom in recent years, not least due to the discovery of significant cross-linguistic variation in both matter and pattern borrowings in morphology, see especially (Johanson, Robbeets 2012; Wiemer et al. 2012; Vanhove et al. 2012; Gardani et al. 2015). An important phenomenon that lies at the intersection of grammaticalization studies and contact linguistics is contact-induced grammaticalization, namely grammaticalization processes that occur due to the influence of one language on another (Heine, Kuteva 2003). Unfortunately, the languages of the Caucasus, including those of the Nakh-Daghestanian family, still do not figure prominently neither in studies of grammaticalization, nor in areal studies.

The languages of the Nakh-Daghestanian (East Caucasian) family, one of the three indigenous language families of the Caucasus and the largest one in terms of the number of languages, at present remain underdescribed. Although new grammatical descriptions and dictionaries of the languages of the family appear almost every year, by no means all grammatical phenomena of these languages receive a detailed description and theoretical analysis. This applies both to the core aspects of grammar, including, e.g., the paradigmatic organization of the verb system, and, to an even greater extent, to peripheral areas such as the behaviour of particular syntactic constructions or the use of clitics. The present dissertation partly fills the existing gap by offering a description and theoretical treatment of a number of phenomena found in the Nakh-Daghestanian languages that previously did not attract sufficient attention. These phenomena are considered both from an intragenetic perspective (as understood by Alexander Kibrik [1998]), i.e., against the background of languages of the same family or group, and from a broader typological perspective. In a number of cases, when we are dealing with contact-induced developments, the focus is on areal interaction, which can involve contacts either between unrelated languages (cf. the influence of Azerbaijani on Lezgian), or between languages of the same genealogical grouping.

The ten papers included in the dissertation are divided into three thematic parts. The three works of the first part analyze intragenetic variation in the use of participles in the Lezgian languages, both in periphrastic forms and in modifying participial clauses. The second part, which includes three articles, examines the areal aspects of grammaticalization in the Lezgian languages, namely borrowing of a numeral classifier and a conditional clitic into Udi as examples of material copying from unrelated languages of the area, and borrowing of a repetitive prefix from Lezgian to the southern dialects of Agul as an example of “intra-group”

contact between the Lezgian languages. Finally, in the four articles included in the third part, case studies dealing with the emergence of not only intragenetically, but also typologically rare structures are presented: these are “verificative” verb forms and post-verbal doubling of pronouns in Agul, and endoclitisis in negative verb forms of Andi.

Some of the phenomena considered in the dissertation, namely the use of a numeral classifier and conditional clitics in Udi, “verificative” forms and pronoun doubling in Agul, and Andi endoclititics, had been virtually unknown before the respective publications of the author, in which they received the first detailed treatment. The general structure of verbal paradigms or the general principles of the functioning of participial clauses in the Lezgian languages had been discussed to some extent in previous literature. However, it is in the present research by the author that they received a more detailed description (relying on corpus data), as well as a theoretical and typological interpretation. Thus, the **novelty** of the dissertation is due to the fact that the vast majority of phenomena presented here have been for the first time described and placed in a theoretical and typological context by the author (a number of phenomena of the Agul language were studied in cooperation with Solmaz Merdanova). Of particular note is the fact that a significant part of the research was carried out based on the material from oral corpora created with the participation of the author: these are the Oral corpus of the Huppuq’ dialect of Agul, developed by Dmitry Ganenkov, Timur Maisak and Solmaz Merdanova, and the Oral corpus of the Nizh dialect of Udi, developed by Dmitry Ganenkov, Yury Lander and Timur Maisak; both corpora are transcripts of the creators’ own audio recordings from the 2000s.

In addition to the descriptive component, the works included in the dissertation aim at the incorporation of the Nakh-Daghestanian data into the context of modern linguistic theory and typology. The studies presented here allow us to deepen our understanding of the functional typology of tense-aspect systems, frequency patterns in the use of modifying participial clauses, pronominal doubling, plural agreement, contact-induced development of classifier systems and conditional/indefiniteness marking, as well as more exotic phenomena such as morphologically bound complementation (cf. “verificatives”) and endoclitisis. These aspects determine the **relevance** and **theoretical significance** of the dissertation for the theory of grammar and typology.

The unifying principle of the studies presented in the dissertation is the combination of typological, diachronic and areal perspectives in the description of the Nakh-Daghestanian languages, as well as the search for theoretical generalizations. In terms of theoretical orientation, the papers are based on fundamental works on the grammaticalization theory and contemporary approaches to areal linguistics and language convergence. Concerning the treatment of linguistic data, I follow the tradition of typologically-oriented descriptions of the Nakh-Daghestanian languages laid down by Alexander Kibrik’s school of field linguistics (see, in particular, Kibrik, Testelets 1999 and Kibrik et al. 2001). As regards issues related to the historical grammar of Nakh-Daghestanian, and especially Lezgian, my approach can be considered as a continuation of the works of such researchers as Mikhail Alekseev, Wolfgang Schulze, Alice Harris (Alekseev 1985; Schulze 1982; Harris 2002), among others.

To put it generally, the studies included in the dissertation: a) show the systematicity and significant predictability in the semantic development of verb forms, in case the structure of the underlying periphrastic constructions is taken into account; b) confirm the tendencies in the use of modifying participial clauses outlined in earlier literature; c) elaborate the scenarios of matter and pattern borrowing of grammatical markers both from genealogically related and unrelated languages; and d) offer explanations of the ways in which typologically rare phenomena arise. Based on the results of the study, the following specific **theses** are proposed for the defense:

1) A peculiarity of the Agul tense-aspect system in the indicative is the existence of a series of periphrastic forms based on participles. Both structurally and semantically, forms of the “participial” subsystem are parallel to those of the main subsystem.

2) Both in Agul and Udi, participial clauses in the modifying function instantiate what is known in the literature as *general noun-modifying clause constructions* (GNMCCs), which allow a wider range of possible associations between a head noun and a clause than canonical relative clauses. For each language, the counts based on both written and oral corpora show a number of peculiar properties with respect to the frequency of relativization of particular arguments or adjuncts (e.g., very high frequency of peripheral arguments/adjuncts relativization in Agul, higher frequency of agent relativization than patient relativization in Udi, among many others).

3) Borrowing of the Azerbaijani conditional mood enclitic =*sa* helped Udi to fill certain functional “lacunae” like the marking of realis conditional protases or the existence of a dedicated marker of indefinite pronouns. At the same time, at the level of grammatical patterns, this borrowing made Udi more, and not less, similar to its Lezgian relatives.

4) Borrowing of the Iranian numeral classifier *dänä* contributed to the divergence of Udi from the other Lezgian languages and Nakh-Daghestanian languages in general, making Udi the only language of the family with a (reduced) system of numeral classification.

5) In southern dialects of Agul, the repetitive prefix most probably represents a borrowing from the neighboring (and closely related) language Lezgian, where the repetitive prefix has grammaticalized from a locative prefix with the POST (‘behind’) meaning. In Agul, this borrowing resulted in the coexistence of two cognate prefixes, both ultimately going back to the Proto-Lezgian marker of localization POST.

6) The series of “verificative” forms in Agul, which are arguably the result of morphologization of the indirect question complement and the matrix verb (most probably, ‘see’), represents an instance of morphologically bound complementation: although the two verbal heads are fused in one synthetic form, they still enjoy syntactic autonomy to a large extent.

7) The postverbal doubling of subject pronouns in Agul (e.g. ‘I said I, ...’) represents a cross-linguistically unusual pattern, being restricted to just one verb (‘say’), and can be viewed as a first step towards the grammaticalization of subject agreement of the type attested in the closely related language Tabasaran.

8) The placement of two enclitics in Andi, namely the additive clitic and the intensifier, inside synthetic negative verb forms should be analysed as an instance of a cross-linguistically rare type of endoclitics, i.e. clitics that can be found inside words.

## 2. Functions of participial verb forms in the Lezgian languages

The Lezgian languages, spoken in southern Dagestan and northern Azerbaijan, are the most southern branch of the Nakh-Dagestani family. There are nine languages in this group, with seven ‘core’ languages and two outliers. The core languages are located on both sides of the Great Caucasus range in the area surrounding the Samur river: Lezgian, Tabasaran and Agul (Aghul) belong to the eastern subbranch, Tsakhur and Rutul belong to the western subbranch, and Kryz (Kryts) and Budugh (Budukh) constitute the southern subbranch. Archi is the northwestern outlier, spoken in the Avar-dominated area of central Dagestan, and Udi is the southernmost outlier, spoken in the Azeri-dominated area of northern Azerbaijan. Udi is also the closest living relative of the extinct Caucasian Albanian language, which can be considered the tenth Lezgian language, and which is the only Nakh-Dagestani language with an ancient written tradition.

The Lezgian languages are predominantly agglutinative with ergative case alignment, extraordinarily rich case systems, nominal gender agreement, and elaborate consonant inventories. The tense and aspect systems of the Lezgian languages are rich and include dozens of both synthetic and periphrastic forms. The most important categories of the verb are tense, aspect, mood/modality and evidentiality. The perfective vs. imperfective aspectual opposition is expressed by inflection within verb stems and is quite archaic. Like other Nakh-Dagestani languages, the Lezgian languages are predominantly left-branching with dominant SOV word order.

In the three papers presented in this part of the dissertation, the use of finite and non-finite forms of the verb in two Lezgian languages, Agul and Udi, are investigated: the general organization of periphrastic forms in Agul and the role of forms based on participles in the tense-aspect system (Section 2.1), and the use of participles in modifying clauses, with a focus on quantitative corpus data (Section 2.2).

### 2.1. Participial forms in the tense-aspect system of Agul

Paper selected for the defense: (Majsak 2012)

In this study, I focus on the indicative tense-aspect system of Agul, whose core part includes the Present (e.g. *ruχaa / ruχaja* ‘is reading, reads’), the Habitual (e.g. *ruχaje* ‘reads’), the Aorist (e.g. *ruχune* ‘read’) and the Perfect (e.g. *ruχunaa / ruχunaja* ‘has read’), all of which have periphrastic origin and are built on the imperfective and perfective converbs. The postposed auxiliaries are the identification copula or the locative stative verb ‘be inside’, which occur in the present and past tenses. Thus, “perfective converb + locative verb in the present” is the source construction for the perfect/resultative, and “imperfective converb + locative verb in the present” is the source construction for the present. “Perfective converb + copula in the present” is the aorist (perfective past), “imperfective converb + copula in the present” is the present habitual. The Future (e.g. *ruχase* ‘will read’) is based on the infinitive: while “infinitive + copula in the present” is the indicative future tense, “infinitive + copula in the past” is the irrealis.

Special attention is given to the participial subsystem, comprising periphrastic forms built on four substantivized participles. Morphologically, participles are a heterogeneous class in Agul. There is no single “participle marker” common to all participles. In general, one can distinguish between three formal types of participles: unmarked, suffixal and periphrastic. The two unmarked participles of standard verbs are simply identical to the two aspectual stems. Statives have an affirmative participle marked suffixally. Finally, there is a series of participles based on periphrastic combinations (“converb + auxiliary in the participle form”). They follow the periphrastic patterns attested in the indicative system both formally and functionally.

Four finite periphrastic forms based on participles, with the copula as the auxiliary verb, are the “participial” Present (e.g. *ruχajef-e* ‘is reading, reads’), Habitual (e.g. *ruχaf-e* ‘reads’), Aorist (e.g. *ruχuf-e* ‘has read’) and Perfect (e.g. *ruχunajef-e* ‘has read’). It is argued that these “participial” periphrastic forms are semantically parallel to the main (“converbial”) forms, and the peculiarities of their meaning should be explained by the semantics of the participles. In particular, while the main forms are basically used to introduce a new situation, participial forms refer to the characteristics of participants and to facts from some stable “fund of knowledge”. Thus, while the Aorist is a typical perfective past used in narratives, the “participial” Aorist has an experiential or existential meaning. The Perfect is a polyfunctional form which expresses both perfect and resultative meanings, as well as indirect evidentiality in the past, whereas the “participial” Perfect has a narrower perfect meaning and introduces a currently relevant situation as already known, the function of the corresponding clause being explanatory or confirmatory.

Also considered are the functions of participles in non-finite clauses (relative, complement and adverbial), as well as their use in the focus construction, where participles head the presupposed part of the utterance. The analysis is based on the corpus of spontaneous oral narratives representing the Huppuq’ dialect.

## 2.2. Participial relative clauses in Agul and Udi from a corpus-based perspective

Papers selected for the defense: (Maisak 2020; Majsak 2020)

The studies present the analysis of quantitative data on modifying participial clauses in two Lezgiic languages, Agul and Udi, based on text corpora. In both languages, modifying participial clauses normally precede nominal heads, and the participle is clause-final within its clause. Like in other Nakh-Daghestanian languages, modifying participial clauses in Agul and Udi are close equivalents of relative clauses proper. However, as they allow a wider range of possible associations between a head noun and a clause and do not show syntactic restrictions on what can be relativized, they can be rather assigned to what is known as *general noun-modifying clause constructions* (GNMCCs, see Matsumoto et al. 2017). The main goal of the study is the analysis of frequencies of different associations between participial clauses and head nouns in terms of arguments, adjuncts, etc.

For Agul, the study is based on a sample of 858 headed noun-modifying clauses taken from two text corpora, one spoken and one written. As the text counts show, intransitive verbs are more frequent than transitive and experiential verbs in participial clauses, and among intransitive verbs, locative statives with the roots ‘be’ and ‘stay, remain’ account for half of all uses. The asymmetry between different relativization targets is also significant. Among the core arguments, the intransitive subject (S) is the most frequent target, patient (P) occupies the second place, and agent (A) is comparatively rare. The preference of S and, in general, of S and P over A also holds true for most of the other Nakh-Daghestanian languages for which comparable counts are available. At the same time, Agul stands apart from the other languages by its high ratio of non-core relativization which accounts for 42% of all participial clauses. Addressees in the dative (1), arguments and adjuncts encoded with a locative case, as well as more general PLACE and TIME relativizations show especially high frequencies, outnumbering such arguments as experiencers, recipients, and predicative and adnominal possessors. Possible reasons for the high ratio of non-argument relativization are discussed in the paper.

- (1) \_\_\_\_\_ *lemert=na žumart ab.a žu ču*  
 [GAP(DAT) Lemert=and Dzhumart say.IPF(PT)] two brother  
 ‘two brothers who are called Lemert and Dzhumart’



For Udi, about 1,000 occurrences of participial clauses were retrieved from three corpora of the Nizh dialect, one spoken corpus, one written corpus comprising a translated text of the Gospel of Luke, and another written corpus comprising two collections of original folklore. The Udi data were compared to data on relativization frequencies available for a few other Nakh-Daghestanian languages, including Agul, Archi, Lezgian etc. The main generalizations which can be made from the counts are as follows. In participial clauses, intransitive predicates (especially the verb ‘be’) turned out to be more frequent than transitive ones. Relativization of the three core arguments S, A and P accounts for the vast majority of all occurrences (more than 80%), with the intransitive subject S by far outnumbering agent and patient (the same is true of the other languages of the family). Unlike in some other related languages, relativization of the agent in Udi is more frequent than that of the patient. Relativization frequencies for peripheral arguments and adjuncts is small compared to some other languages of the family (less than one fifth of all occurrences). Among the non-core relativizations, the locative and the temporal ones are the most common. Also, corpus data confirm the impression that, although “extended” uses (i.e. non-syntactic associations) typical of GNMCCs are indeed attested in Udi, their frequency is very low.

### 3. Areal aspects of grammaticalization in the Lezgian languages

Among the languages of the Nakh-Daghestanian family, the Lezgian languages seem to have been influenced by an unrelated language, namely Azerbaijani (< Turkic), to the most degree. This is especially true for those languages that are spoken in the northern part of Azerbaijan, namely Kryz, Budugh, Udi, as well as southern dialects of Tsakhur, Rutul and Lezgian. Udi, which genealogically is a peripheral member of the Lezgian branch, has been in contact with unrelated languages of the area for centuries: besides Azerbaijani, Udi was influenced by Armenian and Iranian languages (Schulze 2016). Azerbaijani influence can be seen at various linguistic levels, including phonology, the lexicon, and morphosyntax. There is a whole number of morphemes and morphological patterns borrowed from Azerbaijani, such as the suffix *-(i)mž̄i* which derives ordinal numerals from cardinal bases, the adjectival suffix *-lu* ‘related to X’ and the privative suffix *-suz* ‘not having X’, the suffix of abstract and status nouns *-luɤ* and the very frequent complementizer *ki* (ultimately of Iranian origin). Iranian influence is not restricted to Persian: Schulze (2001) identifies several chronological layers of Udi contacts, including those with Old and Middle Iranian varieties, Talyshi and Jewish Tati. Among the linguistic phenomena that might reflect the Iranian influence on Udi, the system of “mobile” subject agreement markers and Differential Object Marking (Schulze 2015), as well as the present tense based on the dative form of the infinitive (Majsak 2016a) have been discussed in the literature.

Being genealogically a Nakh-Daghestanian language, in the areal dimension Udi can be seen as belonging to the north-eastern periphery of Araxes-Iran linguistic area (AILA) covering the South Caucasus, Northern Iran, Northern Iraq and Eastern Turkey, and comprising Kartvelian languages, Armenian and its dialects, Azerbaijani and Turkish varieties, Neo-Aramaic and Arabic dialects, and many Iranian languages (Stilo 2015).

The three papers included in this part of the dissertation present three case studies of contact-induced developments in the grammatical structures of the two Lezgian languages Udi and Agul. Two case studies focus on grammatical copying from unrelated languages: the borrowing of the Azerbaijani conditional marker into Udi with a range of functions related to conditional mood and indefiniteness (Section 3.1), and the borrowing of the Iranian classifier *dänä*, which made Udi the only Nakh-Daghestanian language with a small-inventory classifier system (Section 3.2). The third study is devoted to borrowing from a closely related language, namely the borrowing of the repetitive prefix from Lezgian into the southern dialects of Agul (Section 3.3).

#### 3.1. A borrowed Azerbaijani conditional marker in the grammatical structure of Udi

Paper selected for the defense: (Majsak 2019a)

In this study, I describe the ways in which the conditional marker *=sa* from the sociolinguistically dominant Turkic language Azerbaijani was adopted by the grammatical system of the minor Nakh-Daghestanian language Udi. Although *-sa* (*-sə*) in Azerbaijani exists both as a suffix and as a reduced cliticised form of the conditional copula *isə*, only the latter variant was copied into Udi: the Udi marker *=sa* is an enclitic, which can be hosted by a fully inflected indicative verb form or by a non-verbal predicate. Unlike its Azerbaijani source, *=sa* in Udi has an invariable form: there are no vowel harmony variants. The functional range of the clitic has also been copied from the source language. In both Udi and Azerbaijani, *=sa* is used on verbs in conditional protases proper and in a number of other dependent clause types (indirect questions, correlatives, parametric concessive-conditionals). Before Udi had borrowed the Azerbaijani morpheme, it seems to have lacked a comparable conditional marker that could be combined with various indicative verb forms to create a whole paradigm of conditional mood forms. At the same time, for the other languages of the

Lezgian branch it is typical to use conditional markers (or auxiliaries in the conditional form) in the functions that =*sa* now fulfills in Udi. The borrowing of the conditional clitic thus brought to Udi grammatical structures that are common to the Lezgian languages, making Udi more similar to its relatives.

Outside the verbal domain, the borrowed conditional =*sa* in Udi became part of the indefinite pronoun series marker copied from Azerbaijani. While in Azerbaijani indefinites are derived from interrogatives by means of the conditional copula, there is no copula in modern Udi, so the indefiniteness marker =*esa* / =*nesa* integrated the 3rd singular clitic =*e* / =*ne* in combination with the conditional =*sa* (e.g. *šu* ‘who’ > *šu=ne=sa* [who=3SG=SA] ‘whoever it is, lit. who if is’ > *šu=nesa* [who=INDEF] ‘someone’). Prior to the calquing of the Azerbaijani pattern, there was no general indefiniteness marker in Udi based on a conditional morpheme. The borrowing of this pattern from Azerbaijani has also made Udi more structurally similar to other Lezgian languages, for which the derivation of indefinites from interrogatives by means of a conditional clitic or copula is typical.

Some functions of =*sa* in Udi do not match those of native conditional markers in the other Lezgian languages. In accordance with the corresponding use in Azerbaijani, =*sa* in Udi is employed as an optional marker on the standard of comparison. This use has no known parallels among the other languages of the Lezgian branch. In addition, the Azerbaijani conditional copula in its full form *isə* has been copied into Udi as a contrastive topic marker *isä*. To my knowledge, no instances of the use of the (native) conditional copulas as topic markers have been attested in any other Lezgian language.

From the available data one can draw the conclusion that the borrowing and expansion of =*sa* has occurred mainly in the Nizh dialect of Udi and rather recently (probably starting no earlier than the second half of the 20th century). In published texts in the Nizh dialect, the use of =*sa* is only attested starting from the 2010s. This might be due to a faster expansion of this clitic both in spoken discourse and in written literature after Azerbaijani became the only official language of the country, or to the overall low frequency of the marker (which could have been overlooked by the authors of grammar sketches), or maybe to an idiolectal aversion to =*sa*, which might be felt to be clearly Azerbaijani by some Udi authors. However, it is not the very fact of borrowing this item that is important in the case of Udi. Most Lezgian languages share certain morphosyntactic strategies using conditional markers or auxiliary verbs (e.g. copulas) bearing them. Udi seems to have lacked these strategies, at least during some period of its history. This lack of a regular conditional marker (which is able to combine with various tense-aspect forms) and the lack of a productive indefiniteness marker made Udi’s typological profile less similar to that of related languages. The borrowing of =*sa* does not merely fill certain functional “lacunae” in Udi grammar. It also led to the emergence of patterns that are common for other Lezgian (and many other Nakh-Daghestanian) languages. The difference is that in Udi, the morphological material recruited for the corresponding functions not native, but borrowed. Thus, the convergence of Udi with Azerbaijani at the same time resulted in the convergence of Udi with the Lezgian languages in this particular part of the grammar.

### **3.2. The rise of a reduced classifier system in Udi**

Paper selected for the defense: (Maisak 2021b)

In this study, an account of numeral classifiers in Udi is presented following Stilo’s (2018) study of small-inventory classifier systems in a number of Indo-European, Turkic, Kartvelian and Semitic languages of the Araxes-Iran linguistic area. Being a peripheral member of the linguistic area in question, Udi possesses an even more reduced version of a small-classifier system, comprising one optional classifier *dänä* (Iranian borrowing, most likely via Azerbaijani) used with both human and inanimate nouns. A dedicated classifier for humans is lacking, although there is a word *tan* (also of Iranian origin) only used after

numerals or quantifiers, but predominantly as a noun phrase head. The behaviour of *dänä* and *tan* is scrutinized according to a set of parameters, in both spoken and written textual corpora of the Nizh dialect of Udi. Drawing on data from related Nakh-Daghestanian languages, the paper shows that among the languages of the family Udi may be unique in possessing classifiers (albeit as a result of contact), Khinalug possibly being the only other exception.

- (2) *göj-n-aχun χib dänä eʃʃʃ=e bist:a*  
 sky-OBL-ABL three CLSF apple=3SG fall+PRS  
 ‘Three apples fall from the sky.’
- (3) *va-χun sa šej čur-eb-ala här tan-a tad-a.*  
 you-ABL one thing ask-LV-PT:IPF [every person-DAT] give-IMP  
 ‘Give to everyone who asks you.’ (Lk. 6:30)

The reduced, even with respect to the “small-inventory classifier systems”, classifier system of Udi with its single universal numeral classifier *dänä* cannot be attributed the function of noun categorization. The Udi classifier co-occurs with both human and inanimate noun phrases and is not obligatory. It rather looks like an additional optional element of a numeral phrase, lacking any particular semantic content. It is also far from clear at the present stage of research whether the use of *dänä* in Udi has a clear-cut discourse/pragmatic function.

The position of Udi as a language with a small, even extremely small, classifier system, is intermediate between the other Nakh-Daghestanian languages, on the one hand, and languages of the Araxes-Iran linguistic area, on the other. Among the Nakh-Daghestanian languages, Udi may be unique in having at least one classifier, as the languages of the family typically lack them altogether. Among the AILA languages with their predominantly two-classifier systems, Udi belongs to the peripheral zone and, like some other peripheral languages, can be said to be “transitional to areas where there are no vestiges of classifiers” (Stilo 2018: 136).

### 3.3. Morphological copying of the reflexive prefix from Lezgian into Agul

Papers selected for the defense: (Maisak 2019b; Majsak 2016a)

The study describes the reflexive, or repetitive, prefix *q-* / *qV-* in Agul (4), focusing on the grammaticalization path of this morpheme. The main question that is being addressed is the hypothesis that the prefix has been copied from the closely related Lezgian language.

- (4) Reflexive prefix in Huppuq’ Agul
- |  |                       |
|--|-----------------------|
| <b>q-awas</b> ‘say again’                  | ← <i>awas</i> ‘say’   |
| <b>q-aq’as</b> ‘do again’, ‘fix’           | ← <i>aq’as</i> ‘do’   |
| <b>q-a-ic’as</b> ‘give again’, ‘give back’ | ← <i>ic’as</i> ‘give’ |
| <b>qu-hatas</b> ‘send back’                | ← <i>hatas</i> ‘send’ |

First, I provide a detailed description of the morphology and semantics of the repetitive prefix (‘again’, ‘back’) in comparison to the system of locative prefixes in Agul, showing that, despite the formal similarity with the ‘POST’-localization prefix (meaning ‘behind’), the repetitive prefix is distinct both functionally and paradigmatically. I then compare the repetitive and ‘POST’-prefix in Agul with their counterparts in Lezgian and other Lezgian languages. While the relation of the meaning ‘behind’ to ‘again’ is cross-linguistically common, the development of a special repetitive prefix in Agul is only attested in the two southern dialects, whose speakers have been in long-term contact with Lezgian – a language which possesses a productive repetitive prefix/inflix. It is thus natural to assume that the Agul

prefix has a contact-induced origin. I also show that this is not the only grammatical phenomenon of southern Agul that has been influenced by Lezgian.

The morphological properties and functions of the repetitive prefix in Agul have not been described in detail before. The conclusion that this prefix is a morphological copy is remarkable, because Lezgian is one of the two languages most closely genetically related to Agul, and also because the borrowing of the prefix has led to the emergence of affixal “etymological doublets” in Agul.

## 4. Grammaticalization and the rise of cross-linguistically rare structures

In the four papers included in this part of the dissertation, I analyse three developments that have resulted in cross-linguistically rare structures. This is the development of a series of morphologically synthetic but syntactically biclausal “verificative” verb forms in Agul, a Lezgian language (Section 4.1), the postverbal doubling of free personal pronouns with the verb ‘say’ in Agul (Section 4.2), and the endoclititic behaviour of the additive and intensifying clitics in negative verb forms of Andi, an Avar-Andic language (Section 4.3).

### 4.1. “Verificative” verb forms in Agul: a case of morphologically bound complementation

Papers selected for the defense: (Maisak 2016a; Majsak 2016a)

In this study, I describe a cross-linguistically unusual case of morphological fusion found in Agul, in which two clauses fuse morphologically in the absence of preceding syntactic fusion or clause union. This phenomenon involves a set of “verificative” verbal forms (forms that seek ‘to find out the truth value, or the value of an unknown variable’). Various tense and aspect inflections are possible for the semantically embedded portion of this construction that denotes the proposition to be ‘checked’. The operator-like verificative meaning is expressed by a fused segment that serves as the locus of morphosyntactic marking for the main clause. While verificatives are completely morphologically bound, they manifest clear biclausal properties: in particular, the introduction of a new agentive argument by the verificative (the ergative “verifier”) causes no change in the argument structure of the embedded clause. Also, it is possible to have two adverbials in one sentence, semantically modifying different parts of a single verificative form.

- (5) *zun jaša gada.ji naq’ dars ruχ.u-naj-čuk’.a-s-e.*  
I(ERG) today [boy(ERG) yesterday lesson read.PF-PRF]-VERIF.IPF-INF-COP  
‘I will check today whether the boy learnt his lesson yesterday.’

The paper argues that the Agul verificative has grammaticalized from the matrix verb ‘see’ plus an indirect question complement in the conditional form: over time, the two verbal heads have fused into one form. Thus, in (5) the verificative form is based on the perfect conditional form *ruχ.u-naj-či* [read.PF-PRF-COND] ‘if s/he has read’. Although this situation has a natural diachronic explanation, synchronically, the morphological verificative looks like a cross-linguistically unusual inversion of the much better-known phenomenon of clause union, in which two morphologically independent verbs comprise a single syntactic predicate. Partial parallels to this development can be found in the related languages Archi and Lezgian, where a semantic shift from ‘see’ to ‘check, find out’ is attested, together with a change in subject encoding from typically experiential (dative) to canonically agentive (ergative). Still, the complete morphologization of the verificative structure in Agul dialects remains exceptional given its comparatively recent origin, the infrequency of the construction, and the general absence of observed cases in which matrix verbs become fused with their complements.

### 4.2. Subject pronoun doubling in Agul

Papers selected for the defense: (Majsak 2015; Maisak 2016b)

The study presents the first account of the doubling of free personal pronouns in Agul based on data from the spoken corpus of the Huppuq’ variety of Agul. The doubling construction consists of a subject pronoun in the canonical preverbal position, paired with an identical instance of the same pronoun immediately following the verb. The first pronoun is

usually adjacent to the “verb–pronoun” combination, though it can optionally be separated by another constituent. In the oral corpus consulted for the analysis, the construction is found most often with the primary verb of speech in clauses introducing a quote (e.g. ‘I said I, ...’). 1<sup>st</sup>-person pronouns, 2<sup>nd</sup>-person pronouns, and various kinds of pronouns referring to non-locutors (demonstratives, logophors) can appear in the doubling construction, but the 1<sup>st</sup>-singular subject pronoun is especially frequent.

- (6) *zun* *ʁ-a-a* *zun*, “*ha-mi-šti* *še?* *ag-u-ne* *za-s*”.  
 I(ERG) say-IPF-PRS I(ERG) EM-DEMM-ADV(GEN) thing(ABS) see-PF-AOR I-DAT  
 ‘(And then) I say: “I saw such-and-such thing”.’

I argue that the doubling pattern originated as the conflation of a preverbal subject with a very frequent “verb–subject” word order used with highly topical referents. The function of the doubling construction is therefore postulated to draw additional attention to the referent. A brief comparison of Agul doubling and related phenomena in other languages (e.g. person agreement and clitic doubling) is also offered. As I show, the pronoun-doubling construction in Agul can be seen as a member of a whole “family” of related phenomena found in both genetically close languages (e.g. Tabasaran) and in unrelated languages outside the Caucasus. At the same time, the construction seems to be unique among similar instances described in the literature, and does not match precisely any of them.

#### 4.3. Endoclitics in the negative verb forms of Andi

Paper selected for the defense: (Maisak 2021a)

Endoclitics (also known as intraclitics or mesoclitics), a rare type of clitics occurring inside words, have attracted substantial attention in the literature over the last two decades. For the first time, in this study I provide evidence for the existence of endoclitics in Andi. In Andi, the additive marker (‘also’) and the intensifying marker (‘even, at all’) behave as enclitics on various types of hosts and as endoclitics on negative verb forms. In the latter case, the additive and intensifying markers break up the word form and appear before the negation marker. For example, in (7) the additive clitic =*lo* breaks up the negative perfect forms with the suffix *-č’igu* ‘PRF.NEG’, which is only found in this particular verb form (cf. the affirmative perfect forms *hago-č’igu* ‘has seen’, *anli-č’igu* ‘has heard’).

- (7) *hege-š-bo* *hege-j* *hago=lo=č’igu*, *anli=lo=č’igu*  
 DEM-OBL.M-AFF DEM-F see=ADD=PRF.NEG hear=ADD=PRF.NEG  
 ‘He has neither seen her, nor heard her.’

I argue that both the additive and the intensifier are clitics, especially in view of their highly promiscuous attachment. I also show that negative verb forms are morphologically synthetic, so the additive and the intensifier are genuine endoclitics, i.e. clitics that occur inside morphological words. In addition, I provide a few parallels for the unusual morphosyntactic behaviour of additive and intensifying clitics in some other Nakh-Daghestanian languages as well as in some languages of Northern Eurasia. Although in these cases the corresponding markers do not qualify as endoclitics proper, the available data hint at a cross-linguistic tendency towards word-internal placement of morphemes with meanings like ‘also’, ‘even’ or ‘only’. Thus, it may not be accidental that among the Andi clitics only the additive and the intensifier came to be used as endoclitics: Andi follows an existing tendency, bringing it to the ultimate stage of placing these clitic elements inside a word.

## 5. Conclusions

The studies collected in this dissertation analyse the typological, diachronic and areal aspects of grammaticalization in the Lezgian and Avar-Andic groups of the Nakh-Daghestanian family, mainly in the verb system. Most of the phenomena presented in the papers are described and placed in a theoretical and typological context for the first time. Most of the case studies are based on the corpora of Agul and Udi, which were developed with the participation of the author. Among the topics discussed in the dissertation are the structure and use of verb forms (with a focus on participles), the impact of borrowed grammatical markers (e.g., a numeral classifier or a conditional mood clitic) on the grammatical structure of the recipient language, the creation of grammatical etymological doublets resulting from borrowing from a closely related language, the origin of morphologically bound complements and the cross-linguistically rare case of endoclitisis, among others.

Besides providing new data as part of the description of minority Nakh-Daghestanian languages, the dissertation offers additional support to the scenarios of grammaticalization, both language-internal and contact-induced, that have been discussed in the literature. The generalizations about the synchronic behaviour and diachronic development of various grammatical structures made in the dissertation enrich our understanding of cross-linguistic variation in such areas as tense-aspect systems, relative clauses, person agreement, indefiniteness marking, complementation strategies, numeral classifier systems and the placement of clitics.



## **Abbreviations**

3SG – 3rd person singular person agreement; ABL – ablative; ABS – absolutive; ADD – additive; ADV – adverb; AFF – affective; AOR – aorist; CLSF – classifier; COP – copula; DAT – dative; DEM / DEMM – demonstrative series; EM – emphatic; ERG – ergative; F – feminine; GAP – gap (in a relative clause); GEN – genitive; IMP – imperative; INF – infinitive; IPF – imperfective; LV – light verb; M – masculine; NEG – negation; OBL – oblique stem; PF – perfective; PRF – perfect; PRS – present; PT – participle; VERIF – verificative.

## References

- Alekseev M. E. 1985. *Voprosy sravnitel'no-istoričeskoj grammatiki lezginšix jazykov. Morfologija. Sintaksis* [The problems of comparative-historical reconstruction of the grammar of Lezgian languages: Morphology. Syntax]. M.: Nauka.
- Bisang, Walter & Andrej Malchukov (eds.). 2020a. *Grammaticalization Scenarios: Cross-linguistic Variation and Universal Tendencies. Vol. 1: Grammaticalization Scenarios from Europe and Asia*. Berlin/Boston: de Gruyter.
- Bisang, Walter & Andrej Malchukov (eds.). 2020b. *Grammaticalization Scenarios: Cross-linguistic Variation and Universal Tendencies. Vol. 2 Grammaticalization Scenarios from Africa, the Americas, and the Pacific*. Berlin/Boston: de Gruyter.
- Gardani, Francesco, Arkadiev, Peter & Nino Amiridze (eds.). 2015. *Borrowed Morphology*. Berlin, München, Boston: De Gruyter Mouton.
- Harris, Alice C. 2002. *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.
- Heine, Bernd & Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529–572.
- Johanson Lars & Martine Robbeets (eds.). 2012. *Copies versus Cognates in Bound Morphology*. Leiden: Brill.
- Kibrik, Alexandr E. 1998. Does intragenetic typology make sense? In Winfried Boeder, Christoph Schroeder, Karl H. Wagner, & Wolfgang Wildgen (eds.), *Sprache in Raum und Zeit: In memoriam Johannes Bechert*, Bd. 2: Beiträge zur empirischen Sprachwissenschaft, 61–68. Tübingen: Narr.
- Kibrik A. E., Kazenin K. I., Ljutikova E. A., Tatevosov S. G. (red.). 2001. *Bagvalinskij jazyk. Grammatika. Teksty. Slovare* [Bagvalal: Grammar, texts, dictionaries]. M.: Nasledie.
- Kibrik A. E., Testelec Ja. G. (red.). 1999. *Élementy tsachurskogo jazyka v tipologičeskom osveščeenii* [Aspects of Tsakhur from a typological perspective]. M.: Nasledie.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog and Seongha Rhee. 2019. *World Lexicon of Grammaticalization*. Second, extensively revised and updated edition. Cambridge: Cambridge University Press.
- Matsumoto, Yoshiko, Bernard Comrie & Peter Sells (eds.). 2017. *Noun-modifying clause constructions in languages of Eurasia: Rethinking theoretical and geographical boundaries*. Amsterdam: John Benjamins.
- Narrog, Heiko & Bernd Heine. 2021. *Grammaticalization*. Oxford: Oxford University Press.
- Narrog, Heiko & Bernd Heine (eds.). 2011. *The Oxford handbook of grammaticalization*. New York: Oxford University Press.
- Narrog, Heiko & Bernd Heine (eds.). 2018. *Grammaticalization from a typological perspective*. Oxford: Oxford University Press.
- Schulze, Wolfgang. 1982. *Die Sprache der Uden in Nordazerbajdžan. Studien zur Synchronie und Diachronie einer süd-ostkaukasischen Sprache*. Wiesbaden: Harrassowitz.
- Schulze, Wolfgang. 2001. *The Udi Language: A grammatical description with sample text*. (Available online at <http://wschulze.userweb.mwn.de/Udigen1.htm>)
- Schulze, Wolfgang. 2015. Aspects of Udi-Iranian language contact. In Uwe Bläsing, Victoria Arakelova & Matthias Weinreich (eds.), *Studies on Iran and the Caucasus. In honour of Garnik Asatrian*, 373-401. Leiden: Brill.
- Schulze, Wolfgang. 2016. How much Udi is Udi? In Ramazan Korkmaz & Gürkan Doğan (eds.), *Endangered languages of the Caucasus and beyond*, 187-208. Leiden: Brill.
- Stilo, Donald L. 2015. An introduction to the Atlas of the Araxes-Iran Linguistic Area. In Bernard Comrie & Lucía Golluscio (eds.), *Language contact and documentation. Contacto lingüístico y documentación*, 343-355. Berlin: Walter de Gruyter.

Stilo, Donald L. 2018. Numeral classifier systems in the Araxes-Iran linguistic area. In William B. McGregor & Søren Wichmann (eds.), *The diachrony of classification systems*, 135–164. Amsterdam: John Benjamins.

Vanhove, Martine, Stolz, Thomas, Urdze, Aina & Hitomi Otsuka (eds). 2012. *Morphologies in Contact*. Berlin: Akademie Verlag.

Wiemer, Björn, Wälchli, Bernhard & Hansen, Björn (eds). 2012. *Grammatical replication and borrowability in language contact*, Berlin, Boston: De Gruyter Mouton.