



You have downloaded a document from
RE-BUŚ
repository of the University of Silesia in Katowice

Title: Turkic C- type reduplications

Author: Kamil Stachowski

Citation style: Stachowski Kamil. (2013). Turkic C- type reduplications.
Praca doktorska. Katowice : Uniwersytet Śląski

© Korzystanie z tego materiału jest możliwe zgodnie z właściwymi przepisami o dozwolonym użytku lub o innych wyjątkach przewidzianych w przepisach prawa, a korzystanie w szerszym zakresie wymaga uzyskania zgody uprawnionego.



UNIwersytet ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

Uniwersytet Śląski
Wydział Filologiczny

Kamil Stachowski

Turkic C-type reduplications

Promotor

dr hab. Kamilla Termińska-Korzon, prof. UŚ

Katowice 2013

Contents

| | |
|---------------------------------------|-----------|
| Preface | 1 |
| 1 Introduction | 3 |
| 1.1 Problem | 3 |
| 1.2 Technical | 12 |
| 2 Data | 19 |
| 2.1 The oldest attestations | 19 |
| 2.2 Azeri | 27 |
| 2.3 Bashkir | 34 |
| 2.4 Dolgan | 39 |
| 2.5 Gagauz | 42 |
| 2.6 Karaim | 47 |
| 2.7 Karakalpak | 57 |
| 2.8 Kazakh | 61 |
| 2.9 Khakas | 68 |
| 2.10 Kirghiz | 71 |
| 2.11 Kumyk | 80 |
| 2.12 Oirod | 83 |
| 2.13 Ottoman | 89 |
| 2.14 Shor | 102 |
| 2.15 Tatar | 104 |
| 2.16 Turkish | 108 |
| 2.17 Turkmen | 122 |
| 2.18 Tuvinian | 127 |
| 2.19 Uighur | 132 |
| 2.20 Uzbek | 139 |
| 2.21 Yakut | 145 |

| | |
|---------------------------------------|------------|
| 3 Analysis | 157 |
| 3.1 Recurring peculiarities | 157 |
| 3.2 Structure | 182 |
| 3.3 Semantics | 216 |
| 3.4 History | 227 |
| 4 Summary | 249 |
| 4.1 Summary | 249 |
| 4.2 Conclusions | 259 |
| A Müller 2004 | 263 |
| Indices | 273 |
| Semantic groupings | 273 |
| Subject index | 283 |
| Symbols and abbreviations | 288 |
| References | 288 |

List of Figures

| | | |
|------|--|-----|
| 1.1 | Number of entries and reduplications in dictionaries | 14 |
| 2.1 | Reduplications of Az. * <i>garyš</i> i.a. ‘mixed, scattered’ &c. | 32 |
| 2.2 | Reduplications of Kar.E <i>ansyz</i> ‘suddenly, unawares’ &c. | 52 |
| 2.3 | Reduplications of Oir. <i>ak</i> ‘white’ &c. | 86 |
| 2.4 | Reduplications of Oir. <i>kök</i> ‘blue’ &c. | 88 |
| 2.5 | Phonetic shapes of Ott. <i>topdolu</i> ‘full’ &c. | 99 |
| 2.6 | Reduplications of Shor <i>ak</i> ‘white’ &c. | 103 |
| 3.1 | Map of <i>pp</i> as a closer | 170 |
| 3.2 | Map of entropy of closing consonants | 185 |
| 3.3 | Entropy of closing consonants | 185 |
| 3.4 | Number of reduplications against entropy of closing consonants in different languages | 186 |
| 3.5 | Common reduplicated stems in different genealogical groups | 199 |
| 3.6 | MDS of common reduplicated stems | 201 |
| 3.7 | Semantic ranges in different languages | 220 |
| 3.8 | Number of common semantic units | 222 |
| 3.9 | MDS of common semantic units | 223 |
| 3.10 | Number of bases and of semantic units in different languages | 225 |
| 3.11 | Number of bases against number of semantic units | 225 |
| 3.12 | Map of reduplications in the Tungusic languages | 234 |
| 3.13 | Reduplications against Baskakov’s 1960 classification | 239 |
| 3.14 | Reduplications against Tekin’s 1990 classification | 241 |
| 3.15 | Hierarchical clustering of reduplications | 243 |
| 3.16 | Growth of the number of reduplications in Ottoman | 246 |
| A.1 | Distances between Turkish phonemes encoded for neural networks | 269 |

List of Tables

| | | |
|------|---|-----|
| 1.1 | Types of reduplications in the Turkic languages | 6 |
| 1.2 | Transcription | 15 |
| 2.1 | Reduplications of Yak. <i>soyotoχ</i> ‘lone(ly)’ &c. | 151 |
| 3.1 | Reduplications with alternative closers | 159 |
| 3.2 | Closing consonant vs. labial C_1 | 162 |
| 3.3 | Closing consonant vs. dental C_1 | 163 |
| 3.4 | Initial consonants of the bases | 165 |
| 3.5 | Closing consonant vs. different C_2 | 166 |
| 3.6 | Use of different closing consonants in various languages | 183 |
| 3.7 | Common reduplicated stems | 193 |
| 3.8 | Common reduplicated stems in the Kipchak languages | 194 |
| 3.9 | Presence of reduplicated stems in different genealogical groups | 197 |
| 3.10 | Common reduplicated stems in different genealogical groups | 200 |
| 3.11 | Common reduplicated stems in the <i>mprs</i> -languages | 203 |
| 3.12 | Common reduplicated stems in Azeri, Turkish, Turkmen, and Uzbek | 205 |
| 3.13 | Presence patterns of common reduplicated stems in Azeri, Turkish, Turkmen, and Uzbek | 206 |
| 3.14 | Oghuz common reduplicated stems in Gagauz, Eastern Karaim, Ottoman, and Yakut | 209 |
| 3.15 | Closing consonants other than <i>p</i> in the <i>p</i> -languages | 210 |
| 3.16 | Number of common semantic units | 221 |
| 4.1 | Typological diversification of Turkic collections | 250 |
| A.1 | Müller’s prediction of closing consonants | 266 |
| A.2 | Phonological encoding of Turkish for neural networks | 269 |
| A.3 | Mean accuracy of neural networks in Müller’s test | 270 |
| A.4 | Standard deviation of neural networks in Müller’s test | 270 |

Preface

In line with the classical Middle Eastern tradition of somewhat inflated, but rhyming, titles, the present work could be named *The Book of Explication of Turkic Reduplication*. In line with the more prosaic reality, however, it shall be said that its main goal is to draw a general outline of the past and the present of one type of Turkic reduplications whereby, primarily, adjectives and adverbs are intensified by having their initial mora repeated and prepended to the base with a lexically determined consonant in between, e.g. Tksh. *kara* ‘black’ → *kapkara* ‘jet-black’, Trkm. *gūry* ‘dry’ → *gu.s.gūry* ‘completely dry’.

The novelty of the present work consists in a different attitude. It assembles possibly complete collections of examples from more than twenty Turkic languages, and analyses them as a whole from a diachronic perspective, and combining etymological, historical-comparative, and quantitative methodology. Previous works, on the other hand, typically only focused either on the synchronic state in modern Turkish, or on the general theoretical picture, and were based on highly selective data.

The book begins with an introduction to the problem and setting of general guidelines of composition (chapter 1), then proceeds to present and comment on the material (chapter 2), and then to analyse it and make general observations (chapter 3). Conclusions scattered across the book are then collected and summarized in chapter 4. Finally, accessory material and considerations are presented in appendix A.

The current shape of this work is the result of a collective effort of many people. Several of them have influenced it directly. In particular, my thanks are due to Professors Kamilla Termińska-Korzon (Katowice, Poland) and Marek Stachowski (Cracow, Poland), and to my friends (alphabetically) José Andrés Alonso de la Fuente (Vitoria-Gasteiz, Spain), Tomasz Majtczak, and Michał Németh (both Cracow, Poland). I would also like to express my gratitude to my wife, without whose loving support this work would have forever remained a wishful plan.

Needless to say, all the remaining errors, inaccuracies and weaknesses are my own.

Chapter 1

Introduction

This chapter introduces reduplications, gives an overview of the state of the art, sets the objectives for the present work (1.1), and explains the guidelines observed during its composition (1.2).

1.1 Problem

This section introduces the problem discussed in this book. First, it explains what the main subject is (1.1.1), then it gives an overview of previous research into the phenomenon (1.1.2), and finally it uses this information as background against which to set the primary objective of the work (1.1.3).

1.1.1 Subject

The term *reduplication* is used in the literature to refer to a wide array of repetitions. The present work only discusses one type in the Turkic languages. Primarily, it is a method of intensification of adjectives and adverbs, which yields a form composed of the initial mora of the original word with a lexically determined consonant appended to it, and prepended as a whole to the original word itself, for example: Tksh. *bejaz* ‘white’ → *be.m.bejaz* ‘snow-white’, Trkm. *dōly* ‘full’ → *dosdōly* ‘absolutely full’.

Since the element inserted between the doubled anlaut and the original word is (typically) a single consonant, this type will be called the *C*-type here. Other elements are also possible in the Turkic languages (see tab. 1.1), but they are far less numerous and, it might be suspected, secondary to the *C*-type.

Some terms are occasionally used as mental abbreviations in the present work. For clarity, their slightly more formal but still readable definitions are provided below. (They are given in the alphabetical order, but the reader might want to begin with the term *reduplication*.) Note that they are limited to the usage in the present work, and do not aspire to capture all the senses in which these terms can be found employed in

the literature (see below). Also, the term *homolocal* has been introduced for use in the present work.

anlaut The first segment in a word, or less commonly, in a syllable.¹

auslaut The last segment in a word, or, less commonly, in a syllable.¹

base The word that undergoes \rightarrow reduplication₁. The repeated part of the anlaut of the base will be called the \rightarrow head, and the not-repeated part of the auslaut the \rightarrow tail. A secondary phonetic modification may be applied to the base regardless of the reduplicated anlaut (see 3.1.19).

base meaning (as a mental abbreviation) The meaning of the \rightarrow base.

C₁ (as a mental abbreviation) The initial (not: *first*) consonant of the \rightarrow base. In bases with a vocalic \rightarrow anlaut, C₁ is null (\emptyset).

C₂ (as a mental abbreviation) The first postvocalic consonant of the \rightarrow base.

closer The segment inserted between the \rightarrow reduplicated anlaut and the \rightarrow base during \rightarrow reduplication₁. The closer can be null (\emptyset), although, it seems, only in Mongolic.

closing consonant (as a mental abbreviation)

1. A single consonant acting as a \rightarrow closer.
2. A double consonant, especially *pp*, acting as a \rightarrow closer.

head (of the base) The part of the anlaut of the \rightarrow base that is repeated during \rightarrow reduplication₁. The head cannot be null (\emptyset).

homolocal Pronounced at the same place of articulation.

This term has been introduced to avoid the inaccurate term *homorganic* which might be also misleading, especially for speakers of German (compare the definitions in e.g. Crystal 2008 and Trask 1996 versus those in e.g. Bußmann 1990 and Glück 1993).

mprs-language (mental abbreviation) Any Turkic language in which at least three of *m*, *p*, *r*, and *s* are attested as a \rightarrow closer. See 3.2.1.

p-language (mental abbreviation) Any Turkic language that is not an \rightarrow mprs-language.

reduplicated anlaut Copy of the \rightarrow head created during \rightarrow reduplication₁, and prepended to the \rightarrow base with a \rightarrow closer in between. A secondary phonetic modification may be applied to the reduplicated anlaut regardless of the head (see 3.1.19).

reduplicated meaning (as a mental abbreviation) The meaning of a \rightarrow reduplication₂.

¹ After Trask 1996, who considers the terms to be one of those “largely confined to the older philological literature” and “often maddeningly difficult to look up” (p. viii). It is nevertheless preferred in the present work as being more practical and more widely applicable than “onset” or “the beginning of the word/syllable/...”. Mutatis mutandis, the same applies to *auslaut*.

reduplication

1. (in the present work) Word-derivative method yielding a form composed of the →reduplicated anlaut, plus the →closer, plus the →base.

$$kara \rightarrow \underbrace{ka}_{\substack{\text{reduplicated} \\ \text{anlaut}}} \underbrace{p}_{\text{closer}} \overbrace{\underbrace{ka}_{\text{head}} \underbrace{ra}_{\text{tail}}}_{\text{base}}}$$

Primarily, the meaning of reduplication₁ is intensification of adjectives and adverbs. The reduplicated anlaut is typically identical to the head; in the C-type, the closer is necessarily a single or double consonant, and most commonly *p*; the head is almost always the initial mora of the base (see 3.2.6 on the use of morae in the description).

2. (as a mental abbreviation) The form resulting from →reduplication₁.

tail (of the base) The part of the anlaut of the →base that is not repeated during →reduplication₁. The tail cannot be null (∅).

Note that the above definition of *reduplication* refers to the result rather than to the process. This is because the actual mechanics of the phenomenon are not, in fact, known. Traditionally, descriptions assume that the initial syllable is doubled, its vowel shortened, and, if it existed, its final consonant dropped. However, the same final form of reduplication₂ can be also created by other processes, and there is no actual reason to believe that the customarily assumed one is really the one; see 3.2.6.

In the literature, the term *reduplication* can be found referring to a number of formations that the above set of definitions does not cover. Güler 2003: 67 gives a list of thirteen types of repetitions in Turkish (of one of which the C-type is a subtype); a much shorter list can also be found in Müller 2004: 15. It is my belief that such liberal use of the term has resulted in obfuscating it, and that it would be beneficial if it were reserved for what is sometimes called “partial reduplications”, i.e. those where not the entire word is repeated (as e.g. in Tksh. *kara (my) kara* ‘very black’, *güzel.ler güzel.i* lit. ‘beauty of beauties’, *var.yr var.maz* lit. ‘comes not-comes’ → ‘as soon as [he/she/it] comes’, &c.). Especially, petrified nominal compositions would be better left excluded (e.g. Tksk. *eski püskü* lit. ‘old shabby’, *güçlü kuvvetli* lit. ‘strong powerful’, &c.).

Still, not all types of partial reduplications are covered by the above set of definitions. In particular, the so-called “*m*-reduplications” (e.g. Tksh. *kitap mitap* ‘books and such’) and “echo-words” (e.g. Tksh. *delik deşik* ‘all in holes’, *ufak tefek* ‘tiny’, see also Schönig 1988) deserve, I believe, a more specialized term than the simple *repetition* or *doubling*, but they are not discussed in the present work, and therefore excluded from the special definition of the term provided here.

Out of the many arbitrary ways to classify all the possible reduplications into types, it appears that the closer is the most useful criterion for the Turkic languages. Five main

types can be discerned in the Turkic material, see tab. 1.1. Several unclear examples, possibly not in fact reduplications at all, do not fit into any of them, e.g. Tksh. *čyrylčy* ≠ *plak* ‘stark-naked’, or *paramparča* ‘shattered, in pieces’. Not recognized as a separate type are here those reduplications where the closer is identical to the first postvocalic consonant of the stem, e.g. Az. *jumjumšag* ‘very soft’, Uzb. *japjalâq* ‘completely flat’, Yak. *čepčepčeki* ‘very cheap’, &c., see 3.1.6.

| Type | Closer | Example closer | Example reduplication |
|------------|-------------------------|---------------------|--|
| <i>C</i> | <i>C</i> or \bar{C} | <i>p, pp</i> | Tksh. <i>kapkara, appak</i> |
| <i>V</i> | <i>V</i> | <i>a</i> | Bshk. <i>karakaršy</i> |
| <i>CV</i> | <i>CV</i> or $\bar{C}V$ | <i>pa, ppa, ry</i> | Tksh. <i>güpegündüz</i> , Uzb. <i>jâppajâlyiz</i> , Yak. <i>örüöhös</i> |
| <i>CVC</i> | $C_{\alpha}VC_{\beta}$ | <i>bys</i> | Yak. <i>debisdeŋ</i> |
| <i>ma</i> | <i>ma, Vma, or Cma</i> | <i>ma, ama, pma</i> | Tskh. <i>karmakaryšyk</i> , Kklp. <i>kara</i> ≠ <i>makaršy</i> , Bshk. <i>kapmakaršy</i> |
| <i>rV</i> | <i>rV</i> | <i>ru</i> | Yak. <i>čuručulbugur</i> |

Table 1.1: Classification of the most common types of Turkic reduplications by the closer. Note that reduplications with a long consonant are not considered separate types here. This is for two reasons: 1. historically, they are most probably just emphatic variants of reduplications with a single closing consonant, which 2. came about through lengthening (doubling), not elision (i.e. *appak* ‘snow-white’ < *apak* id. rather than **apapak* id.), see 3.1.8.

By far the most common and most numerous represented in the Turkic languages, is the *C*-type, and it is to it, that the present work is exclusively devoted. See 1.1.3 for a more detailed statement of the objectives.

1.1.2 State of the art

Over a thousand years, study of Turkic reduplication appears to have intensified at an almost exponential pace. The first description of reduplication comes from Maḥmūd al-Kāšyarī, and is followed by a six hundred years long pause. (Chaghatai material is not included in the present work, see 1.2.1 for the reason.) François à Mesgnien Meninski’s account precedes a two hundred years long gap. In the 19th century, research finally gains impetus, and it does not appear to be slowing down today. This subsection briefly outlines the background against which this book is set (see 1.1.3).

11th century

The earliest description of reduplication, greatly predating all the others, was given by the brilliant Maḥmūd al-Kāšyarī in his *Compendium*. He described the process so: “The rule about colors and exaggerating the description of things is to take the first letter of the word and join it to *bā*’ in most of the Turkic dialects, but to *mīm* in Oğuz” (Dankoff/Kelly 1982: 261) which, apart from setting the terminus ante quem for both reduplication itself, and the diversification of closing consonants, shows that al-Kāšyarī viewed reduplication as a morphological or word derivative phenomenon rather than a simple combination of an, admittedly, somewhat limited in scope but otherwise independent intensifier, and an adjective. In this, he surpassed some of the modern scholars by almost a millennium.

To be fair, it must be mentioned that reduplicated anlauts are in fact called “exaggerative particles” or just “exaggeratives” in other parts of the book (*äp* in *äp ädgü* ‘very good’ (p. 87), *köm* in *köm kök* ‘deep gray’ (p. 267), possibly also *čim* in *čim jīg* ‘very raw’ (p. 267)), but since the same term is used to refer to *köp* in *köp kök* ‘very blue’ and *sap* in *sap saryg* ‘very yellow’ in the same entry where the previously adduced description is given, I am convinced that it is simply al-Kāšyarī’s shorthand term for ‘reduplication’ and ‘reduplicated anlaut’, which does not imply that his understanding of the phenomenon was any different from what he had stated explicitly.

17th–19th century

The next grammatical description that I am aware of was given as much as six hundred years later for Ottoman by François à Mesgnien Meninski (note that Chaghatai material is not included in the present work, see 1.2.1), who says: “Ad intendendam autem lignificationem Adjectivorum utuntur particulis certis ad fonum quali effectis, quas Adjectivo separatim præponunt” (Meninski 1680: V 39). It seems that, while noticing the oddity of reduplicated anlauts, Meninski was not prepared to abandon the classical Graeco-Roman perspective on grammar.

As far as Ottoman is concerned, a slightly more definite statement can be found in Jehlitschka 1895: 56f, where the reduplicated anlauts are described as “Vorsatzsilben [...], welche mit demselben Konsonanten und Vokale beginnen, sonst aber ziemlich willkürlich sind”, and it is only in Németh 1916: 41 and Deny 1921: 236, that the phenomenon is fully recognized as a reduplication. The great majority of 17th–19th century grammars, however, either do not seem to mention it at all, or provide effectively no grammatical commentary at all (see 2.13.1 for a list).

20th century – grammars and dictionaries

The turn of the 19th and 20th centuries, it appears, brought about the change in interpretation of reduplication. Pekarskij’s Yakut dictionary of 1907–30 well exemplifies

the doubts of the 19th century. Reduplicated anlauts are typically presented as separate entries in it, and commented with more than ten different formulas, composed of such fragments as ‘alliterating particle/syllable/word’ and ‘intensifying the adjective/concept ...’ or ‘intensifying the adjectives beginning with ...’. Other dictionaries of the period, Budagov 1869–71 and Radloff 1893–1911, contain uncomparably fewer examples but they, too, are inconsistent in their description, hesitate between viewing reduplicated anlauts as particles, syllables or words, and are unclear about what the actual mechanism of their connectivity with specific adjectives is. In the case of *sap*, for example, Radloff suggests in fact a universal applicability (“слогъ, усиливающей прилагательныхъ – eine Eigenschaftswörter verstärkende Vorsilbe”).

In grammars, it was mentioned earlier, the term *reduplication* begins to dominate with the break of the 20th century, which is almost as soon as Turkic grammars start to be published in greater quantities, and to discuss reduplication. In dictionaries, however, the practice of isolating reduplicated anlauts into separate entries continues even today. In itself, such organization of the dictionary is not necessarily a sign of the lexicographer’s preferred interpretation; SKzkP, for example, gives these anlauts as separate entries and defines them consistently as ‘doubled syllable of adjectives and adverbs beginning with ...’. Often, however, the anlauts are given as separate entries without any commentary, which can be only understood as a sign that the author considered them to be independent entities. Explicit mentions of *intensifying particles* can also be found in more than one 20th century dictionary. In EDAL, for example, ПТкс. <*Ap / *Ep> is said to be an “emphatic strenghtening particle”, and is connected with ПМо. <*aba-> ‘huge’, Proto-Japanese <*əpə->, and Proto-Korean <*opi->.

But grammars, too, are frequently imperfect. Most often, they fail to mention the shift of accent to the reduplicated anlaut, and the shortening of the reduplicated vowel. Typically, the process is described as a repetition of the first syllable with an insertion of the closing consonant. Only the more careful authors note that when the initial syllable is closed, its final consonant is dropped. Attempts to establish a distribution rule for closing consonants are hardly ever made, and the possibility of secondary phonetic modifications is almost never mentioned. Müller 2004: 96f reviews several descriptions of Turkish reduplications and mundanely points out the recurring flaws; see also the “Sources” subsections in respective sections in chapter 2. Grammars of the other Turkic languages are generally no different in this regard.²

² Not to stigmatize, but to exemplify, the following two descriptions can be adduced: “Pekıştirme derecesi, sıfatın ilk hecesine /p/ sesinin getirilmesiyle de yapılır: [examples]” (Kasapoğlu Çengel 2005: 188), “Простая форма превосходной степени образуется путем повторения первого слога прилагательного. В конце этого слога прибавляется согласный звук (в большинстве случаев п или очень редко — м); таким образом, открытый слог превращается в закрытый и ударение падает на него: [examples]” (Ahmerov 1958: 766).

20th century – specialized works

In the second half of the 20th century, works devoted specifically to reduplication begin to appear. The great majority are effectively limited to Turkish, further magnifying the disproportion between it and the rest of the Altaic languages. At first, they usually concentrated on one of the two topics: 1. so rephrasing the synchronic description as to make reduplication better fit one or another general linguistic theory, or 2. establishing the rules of distribution of the closing consonant.

Works from the first group are usually ignored in the present book, not least because they typically operate on highly selective data from a wide range of completely unrelated languages, and in effect can hardly be expected to produce trustworthy and substantial conclusions about either.³ Here, it will be only noted that in the course of these and similar studies, the idea appeared that reduplicated anlauts could be interpreted as prefixes (see e.g. Deny 1938). I should like to object. *Prefix* is, of course, a secondary notion, and as effectively all notions in linguistics, it lacks a proper and, at the same time, widely accepted definition. It is therefore perfectly possible to craft one that includes reduplication, but one needs to be aware of the balance of costs of the operation. The already dubious clarity of linguistic terminology would be shed, and I fail to see for what gain. Müller 2004: 209f criticizes the idea in more detail, and see also Frankle 1948: 115.

As for the second group, three authors in particular seem to deserve a mention. All focus on establishing a set of synchronic, phonetic rules of distribution of closing consonants in Turkish. Methodologically, this is a chancy choice. It becomes clear why it is so when one begins to consider words which can have their reduplicated anlauts closed by different closing consonants (seventeen in Turkish, e.g. *jamjaš* ~ *japjaš* ‘completely wet’, see 2.16.4), or the fact that the stock and distribution of closing consonants is different in various languages (e.g. *Az. dimdiri* ~ *Tksh. dipdiri* ~ *Trkm. disdiri* ‘absolutely (a)live, very lively’). The answer, then, to how all the Turkish reduplications actually came to be must be sought in history and comparison with other languages, but, the rest of this work will show, in reality only fragments of it can be found even there (see 3.4.1 and 3.4.2).

³ In the way of an example, let us mention Di Sciullo 2005 who analyses English, Yekhee (Niger-Congo) and Turkish and, among other things, correctly recognizes *ne* in *Tksh. neredede* ‘where’ and *ne zaman* ‘when’ to be an independent word meaning ‘what’, but proceeds undismayed to interpret it as a “wh-affix” and to conclude that reduplication is not the only case of prefixation in Turkish (p. 109). See also p. 108 for a creative use of loanwords for the same purpose (e.g. *isti* in *istifade* ‘utilization’).

Other examples of works in this group might be: Marantz 1982, Alderete et al. 1999, Kelepir 2000, Kim 2009, McCarthy/Prince 1988, or Raimy 2000. Note that not all of them are as blatant as Di Sciullo 2005, but also not all even mention any Turkic language.

Hatiboğlu 1973 was the first extensive study. She gives an overview of the history and the present state of Turkish reduplications, and concludes with a set of four phonetic rules, or rather *tendencies*, that capture the majority of examples (pp. 34, 37, 38, and 42):

1. İlk hecesi ünlüyle başlayan veya biten ya da sürekli ünsüzlerden biriyle kapanan sözcükler, “p” ünsüzüyle pekiştirilir.
2. Tek heceli sözcüklerin çoğu “m” ünsüzü ile pekiştirilir.
3. İlk hecesi, dudak ünsüzlerinden “b, p, m”, dış ünsüzlerinden “c, d, t”, damak ünsüzlerinden “k, y” ile başlayan bazı sözcükler “s” sesiyle pekiştirilir.
4. İlk hecesi “ç” ya da “s” ünsüzüyle başlayan ve ters orantı ilkesine göre “p, m, s” ünsüzleriyle pekiştirilemeyen sözcüklerin bazıları da “r” ünsüzüyle pekiştirilir.

Rules of this kind can be made very accurate by increasing their complexity, and at the same time, dissociating them from any reasonable phonetic motivation. In the extreme case, a separate rule could be devised for every example. On the other end of the scale, a general phonetic motivation can be captured by just one or two rules which, however, will not be without exceptions. Hatiboğlu balances these constraints and delivers an acceptable compromise. According to Demircan 1987: 26 and 1989: 161, her tendencies account for about 70% of the examples.

Demircan 1987 and 1989 aims to kill two birds with one stone, and to establish a set of rules that is both accurate and phonetically motivated. Below is an extract of the results (1987: 36f):

A. Basic processes

2. Close the preceding syllable, (C)V, with /p/.

B. Filtering operations

1. Avoid clusters identical with any of the base consonants [...].
2. Select the closer bearing features in contrast with the base-second consonant [...].
3. Balance and optimise the distribution of features across the emphatic form [...].

To be sure, both the 1987 and the 1989 paper explain in detail how the contrasting, balancing, &c. are supposed to be effectuated. Overall, Demircan’s procedure is more accurate than Hatiboğlu’s and, by referring to phonetic properties rather than specific sounds, it is also based on a stronger phonetic foundation. Naturally, it is not entirely exception-free because no purely phonetic set can be so, if seventeen stems in Turkish have reduplications that can be closed by more than one consonant (e.g. *çimçij* ~ *çipçij* ‘completely raw’, *tamtāze* ~ *taptāze* ‘absolutely fresh’, &c., see 2.16.4).

By far the most extensive work on Turkish reduplication so far, and not only the *C*-type, is Müller 2004. It is very uneven, and therefore difficult to characterize shortly. Apparently, a larger part of it has been assembled from summaries of previous works on reduplication and various other topics, neither of which seems to serve any tangible purpose in the book. Much place is also devoted to general considerations which fall under the first type of works, i.e. those usually ignored here. The relevant and original bits are foremostly a new set of rules of distribution of closing consonants, and an interview of 125 Turkish students. Both are presented in appendix A.

Although more accurate than Hatiboğlu's (they account for about 79% of examples), Müller's rules are also complex to the point of overfitting, or beyond. Like Hatiboğlu's, they refer to specific sounds and therefore entirely obscure the eventual phonetic motivation. This is corrected by his "Kontrast-These" (pp. 156f), not quite unlike that of Demircan's.

Having established the rules, Müller put them to the test and asked 125 Turkish students to reduplicate some real and some non-words. It is not clear to me what results he had expected, and how he had intended to interpret them. See appendix A.2 for a more detailed summary, and below for a similar experiment performed by Sofu 2005 and Sofu/Altan 2009.

All things concerned, Müller 2004 is a convenient source of information on Turkish reduplication and the history of research into it, but the advance it brings into the actual understanding of the phenomenon is disproportionate to the effort. See also 2.16.1 and appendix A for other remarks on the book.

20th–21st century

Theoretical works on reduplication (the group generally ignored here) not only continue to appear in the 21st century, but are even increasingly numerous (see fn. 3). By the end of the 20th century, however, also other aspects of reduplication apart from the two mentioned above, started to attract some attention.

In particular, two papers in the area of language acquisition deserve a mention, Sofu 2005, and Sofu/Altan 2009. Both discuss an experiment similar to Müller's (see above), but interpret it in an adept way. Namely, they conclude that "[w]ords beginning with vowels are rule-governed" while "of the words beginning with consonants, frequently used ones seem to be stored individually in the lexicon", and the infrequent ones are not, and "pose problems in production and [are] more prone to errors" (Sofu/Altan 2009: 72).

1.1.3 Objectives

The great majority of works discussing Turkic reduplications, especially from the last hundred years, are devoted almost exclusively to Turkish, and only ever mention other

languages as if incidentally. The research is also dominated by the synchronic perspective which, I believe, cannot by definition provide certain answers (see 3.4.1 and 3.4.2).

Thus, the secondary objectives of the present work are: 1. to assemble a possibly complete collection of reduplications in various Turkic languages, 2. to introduce the diachronic perspective into the study of Turkic reduplications, and 3. to incorporate quantitative methodology into the research into the history of the Turkic languages.

These three combine into the primary objective, which is to begin to seal the gap that previous research has been circumventing. The current work attempts to draw a general sketch of the past and the present of *C*-type reduplications in the Turkic family as a whole. It collects the oldest available attestations, historic data from Ottoman, and modern data from twenty languages, and then analyses them etymologically, historical-comparatively, and quantitatively.

1.2 Technical

This section explains the technical aspects of the present work. It begins with a description of the sources (1.2.1), then of the transcription (1.2.2), then of the rules observed during the translation of examples (1.2.3), and lastly of the structure of entries in chapter 2 (1.2.4). Finally, the programs used during the writing of this work are listed as a modest acknowledgement (1.2.5).

1.2.1 Sources

This book collects material from twenty modern Turkic languages, Ottoman, and the oldest available attestations. Only the literary varieties have been taken into account because comprehensive collections of dialectal material are effectively only available for Turkish. Including just them would result in a skewed picture. Also excluded are Chaghatai data because they cannot be unequivocally assigned to an earlier stage of any specific modern language and, unlike the oldest attestations, are not a prospective source by which to establish *termini ante quos*. Several languages have not been taken into account because the sources that were available to me were found to only contain very clearly incomplete collections of reduplications: Abdal (Ladstätter/Tietze 1994), Armeno-Kipchak (DAK, Schültz 1968), Chulyim (Birjukovič 1984, Li et al. 2008, Pomorska 2004), Crimean Tatar (Jankowski 1992), Fuyu (Li/Ölmez/Juwon 2007), Khalaj (Doerfer/Tezcan 1980), Salar (Tenišev 1963, 1976a), Tofalar (Rassadin 1978), and Western Yughur (Roos 2000, Tenišev 1976b, Tenišev/Todaeva 1966).

Dictionaries proved to be the richest sources. In total, included are more than 1200 *C*-type reduplications extracted from well above 566 000 entries. Apart from dictionaries, grammars were used for all languages, although their descriptions have generally proven to be imprecise if not plainly wrong, and the examples were scarce (see the

“Sources” subsection in respective sections in chapter 2). They did, however, quite often supply unusual or borderline examples which are difficult to find or altogether missing from dictionaries. Finally, specialized works were used, but since most are devoted almost exclusively to Turkish, so was their contribution to our work. In several cases, however, interesting examples and insights regarding other languages could also be found in them.

Most reduplications, nonetheless, are relatively infrequent words. A search in many small dictionaries often yields no more than a handful of examples; a single big dictionary might well reveal many more, even if the number of entries it contains is lower than the total number of entries in the smaller works. I always used the most extensive dictionary available to me and only supplemented the results with other sources.

Nevertheless, the size and quality of sources varies dramatically between different languages. Turkish, for example, is rather well and comprehensively described, while the South Siberian languages must often settle for one medium-sized or small dictionary and just a general sketch of the grammar. It would not be ungrounded to fear that the collection of reduplications extracted from so uneven sources will be strongly biased in favour of the politically more prominent languages.

Luckily, this is not quite the case. There exists what might be considered a moderate correlation between the number of entries and the number of reduplications in a dictionary (Spearman’s $\rho = 0.615^4$), but note that various dictionaries with the same or almost the same number of entries may still contain very different numbers of reduplications (see e.g. those with 40 000 entries in fig. 1.1). Size of the source is a factor, but not a decisive one. Therefore, I claim that the collections of reduplications presented in this work can be thought of as fairly representative for the respective languages, and not just for the sizes of the sources.

1.2.2 Transcription

For Turkic examples, a version of the Finno-Ugric Transcription is used at the phonological-orthographic level of abstraction (“L5”, see Stachowski K. 2011 for details).

For each source, a short summary of the specific transliteration employed for it is given in the “Sources” subsection of the respective sections in chapter 2. The summaries only contain those characters which actually appear in the examples, and when there is no danger of confusion, omit the graphemes that have been left unchanged,

⁴ Spearman’s rank correlation coefficient is a measure of statistical dependence between two variables. It ranges from -1 to +1, where the extremities denote a perfect negative or positive correlation, and 0 strictly no correlation. Apart from this, there is no fixed way to interpret the result. Unlike the perhaps more popular Pearson’s coefficient, Spearman’s ρ allows for non-linear models so long as they are monotonic, is significantly less sensitive to outliers, and does not assume that the variables are normally distributed, which is not the case here (in the Shapiro-Wilk test, the p -value is 0.22 for the number of reduplications, but only 0.016 for the number of entries).

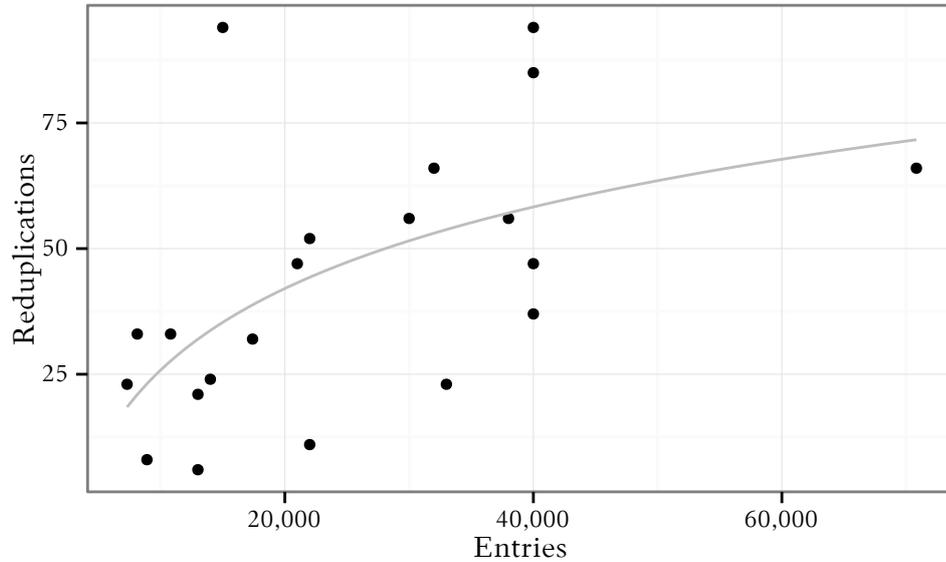


Figure 1.1: The number of entries and the number of *C*-type reduplications in the dictionaries used in this work. The relation can be approximated by the formula $n_{redup.} = 23.41 \cdot \ln(n_{entries}) - 189.78$, resulting in R^2 of only 0.31 (see fn. 54 for the meaning of this index), and represented by the grey line.

or which are provided for by the United Nations system of romanization of Russian (UN 1987, V/18). The character(s) in the source are given in italics, and followed by the character(s) that represent them in this work; different substitutions are separated by a vertical bar (|).

The ordering ignores diacritics for as long as possible, i.e. as in German. Greek letters are ordered as would be their Latin counterparts with diacritics. For example: *ag* < *äg* < *aγ* < *ah*, *sa* < *ša* < *šä* < *se*.

For convenience, the basic notations and most common substitutions are summarized in tab. 1.2.

Non-Turkic examples are usually given in the original transcription of the source or in the official orthography, unless the focus is on the pronunciation or phonology.

1.2.3 Translation

The present work collects data from sources written in a number of languages: most frequently Russian, but also more or less sporadically Belorussian, English, French, German, Italian, Latin, Polish, Spanish, Turkish, Ukrainian, and others. For convenience, the meanings of examples in the lists in chapter 2 are always translated into English, while observing the following rules:

| Manner | Place | | | | | | | |
|---------------|----------|-------------|--------|----------|--------------|-------|--------|---------|
| | Bilabial | Labiodental | Dental | Alveolar | Postalveolar | Velar | Uvular | Glottal |
| Plosive | p b | | | t d | | k g | q̤ | ʔ |
| Affricate | | | | c ʒ | č ʒ̣ | | | |
| Fricative | ɸ β | f v | θ ð | s z | ʃ ʒ̣ | χ ʁ | | h ɦ |
| Nasal | m | | | n | | ŋ | | |
| Lateral/Trill | | | | l ɭ | | | ʀ | |

(a) Consonants

| | | | |
|---------------|-------------------|--|------------------|
| | ̃ | nasalization | |
| | ◌̘, ◌̙, ◌̚ and ◌̜ | slight shift back, forward, up, and down | |
| | ◌̟ | palatalization | |
| | SMALL CAPS | half-voicedness | |
| | rotation* | reduction | α = ǎ |
| ◌̘, ◌̙ and ◌̚ | | overshortness, half-length and length | ◌̘ = j |
| | ◌̘ and ◌̙ | primary and secondary stress | ◌̙ = y |
| ◌̘◌̙ | | syllable boundary | ◌̘ = q |
| ◌̘◌̙ | | morpheme boundary | ◌̙ = w |
| ◌̘◌̙ | | assimilation, elision | ◌̘ = Ẃ |
| | ◌̘, ◌̙ = '◌̘, ◌̙ | | ◌̘, ◌̙ = '◌̘, ◌̙ |

| | | | | | | | |
|---|----|---|---|----|---|---|----|
| i | ü | — | ı | u | — | ı | u |
| e | ö | — | e | o | — | e | o |
| ε | ọ̈ | — | ε | ọ | — | ε | ọ |
| ä | ǎ | — | a | ǎ | — | a | ǎ |

(b) Vowels

(c) Diacritics

(d) Alternates

| From → To | From → To | From → To | From → To | From → To |
|-----------|-----------------------------|------------------------------|-----------|--------------------------|
| c → č | a → a | ж → ž | o → o | x, h → h, χ [†] |
| ç → č | б → b | з → z | ø → ö | ц → c |
| ğ → ɣ | в → v | и → i | π → p | ч → č |
| ĩ, i → y | г, ґ, ɣ → g, ɣ [†] | й → j | ρ → r | ш → š |
| j → ž | д → d | к, қ, қь → k, q [†] | c → s | щ → šč |
| ñ → ŋ | дъ → ž | л, l → l | т → t | ы → y |
| ş → š | e → e | м → m | у → u | э → è |
| q → k | ë → ë | н → n | ү → ü | ю → ju |
| y → j | ə → ä | ң, ҥ, ңь → ŋ | ф → f | я → ja |

(e) Transliteration

Table 1.2: A fragment of the transcription (1.2a–1.2d) and transliteration (1.2e) used in the present work. See Stachowski K. 2011 for details on the former, and the respective sections in chapter 2 for the latter.

* The default is 180° rotation (⟨e⟩ → ⟨ə⟩), but when this is likely to cause confusion, 90° ccw rotation is used instead (⟨u⟩ → ⟨≡⟩ as in most typefaces ⟨n⟩ is too similar to ⟨n⟩).

† Depending on the phonological status in the given language.

- Translations have been standardized and simplified as much as it was possible without falsifying the actual attestations. Examples: Bshk. *takyr* ‘ровный’ (BškRS58), ‘ровный, гладкий’ (BškRS96) → *smooth*; Oir. *žapžaju* ‘1. уер~~е~~ yeni; 2. haber’ (AltTS), ‘brand-new, quite new; novelty’ (Li et al. 2007) → ‘1. brand-new; 2. news, novelty’.
- Where the original meaning was not specified clearly, and a correspondingly ambiguous English word was available, it was used instead of the more precise equivalents. Example: *лэжкуй* → *light* rather than **lightweight* or **easy*. Otherwise, I avoided ambiguity to the best of my abilities.
- Where only the base was translated, and the reduplication listed without any specific meaning (which was fairly often the case), it was assumed that the reduplicated meaning is simply an intensification of the base meanings.
- Distinctions inside nomina have not been specified where they are an obvious consequence of the structure of the Turkic languages (i.e. a part of the translation rather than of the actual meaning of the word itself), and not important for the meaning of the reduplication. Examples: “1. ясный; 2. ясно” → “clear, bright” rather than *“1. clear, bright; 2. clearly, brightly”; “1. холод, мороз; 2. холодный, морозный” → “cold, frost” rather than *“1. cold, frost; 2. cold, frosty”. In less clear cases, they have been listed as one meaning. Example: “1. шар; 2. шарообразный, круглый” → “sphere, spherical, round” rather than *“1. sphere; 2. spherical, round”.
- Inside translations, a colon denotes grouping, brackets denote optionality, and italics denote additional specification. Examples: *very: light, bright* = *very light, very bright*; *(a)live* = *alive, live*; *narrow of eyes*.

Meanings outside the lists in chapter 2 are mostly translated, too, except for where I felt my translation would have to depart too far from the original, would obfuscate the general picture, or where particular precision was required.

Meanings quoted from dictionaries might have qualifiers, numbering of components, and examples of usage omitted from inside them without notification, where they were not important for the matter currently at hand. Also, in some cases the style of numbering has been slightly changed without notification for the purpose of standardization.

1.2.4 Structure of an entry

All entries in chapter 2 follow two fixed schemes, one for the lists of standard cases, and one for the lists of special cases. The former is as follows (square brackets denote optional elements):

base ‘meaning of the base’ [(source of the base)] \diamond **reduplication** [‘meaning of the reduplication’] (source of the reduplication) [**• example**]

‘meaning of the example’ [(source of the example)]
[additional comments]

Remarks (ordered by topic):

- Semantics
 - Only those components of the meaning of the base word are listed which are also present in the meaning of the reduplication. Where they ranked one in the source, the subsequent components are omitted unless they are important for the general picture. Where they did not rank one, they are marked as “i.a.”
 - The meaning of the reduplication is omitted where it is the same as the meaning of the base, or its simple intensification. Where additional meanings appear, the common part is abbreviated to *intens*. Example: Kzk. *tolyk* ‘full, complete’ → *toptolyk* ‘1. absolutely full, to the brim, perfectly complete; 2. puffy, fat’ is abbreviated to ‘1. *intens*.; 2. puffy, fat’.
 - Where the meaning of the reduplication was not given specifically in the source and I could not find it elsewhere, it was assumed to be a simple intensification of all the meanings of the base.
- Sources
 - The source of the attestation of the base is not specified when it is the same as the source of the reduplication, or when the word can be easily found in a dictionary (this refers in particular to examples extracted from grammars and specialized works).
 - Sources are ordered chronologically.
 - The exact location in the source is not specified where it is obvious. For dictionaries, the ‘obvious’ locations are: the base, the reduplicated anlaut, and the reduplication.
- Other
 - Multiple reduplications of a single base word are listed inside a single entry, each preceded by a white lozenge (◊). Reduplications are given in full because changes in the phonetic shape can occasionally occur.

The “Special cases” sections collect unusual words, reduplications with unattested bases, &c., which, most of the times, require a longer and frequently substantial commentary. Short and mostly technical remarks relating to specific parts of the entry, are placed in footnotes. The “special” and “standard” lists are not intended to necessarily be disjoint, but words belonging to both are rare.

Entries in these lists are structured almost the same as the standard ones. The main difference is that the base word and the reduplication are swapped. This is because

in standard cases, a single base can have multiple reduplications but in special cases, reduplications can occur with no base word at all. Unlike in the standard lists, additional comments that follow in a separate paragraph, are not optional.

1.2.5 Technical

During the writing of this book I used rather extensively a number of programs whose authors have been so generous as to make them available for free for everyone. I believe that at least a mention here is due.

The text was written in \LaTeX in Vim on Arch GNU/Linux. Calculations were performed in R (R Core Team 2013). Plots were prepared with ggplot2 (Wickham 2009) and Inkscape (Inkscape Team 2013), except for fig. 3.5 and 3.8 which were created with Circos (Krzywinski et al. 2009). Graphs were composed with TikZ & PGF (Tantau et al. 2011). Maps were made in Quantum GIS (Quantum GIS Development Team 2013) based on the GREG dataset (Weidmann/Rød/Cederman 2010), which is a digitalization of *Атлас народов мира* (Bruk/Apenčenko 1964).⁵

⁵ Maps are in the Lambert azimuthal equal-area projection centred at 55° N, 90° E (near Krasnoyarsk), and have been purposefully simplified to emphasize that they represent the literary languages rather than specific dialects. The area where Karaim is spoken has been significantly enlarged for better visibility. Map 3.12 is an exception in that it is additionally based on the map in Doerfer/Weiers 1985, and has not been simplified.

Chapter 2

Data

This chapter presents *C*-type reduplications as can be found in the oldest available sources (2.1), twenty modern Turkic languages (2.2–2.21), and one historic (Ottoman, 2.13).

All sections follow a fixed scheme. They begin with an outline of *C*-type reduplications in the given language, followed by an enumeration of the sources used, and a brief summary of the descriptions of the phenomenon that can be found in grammar books. Next, standard examples are listed, and after them, special cases together with a commentary. Closing the section are summaries of the structural and semantic features and peculiarities of reduplications in the given language.

Analysis of the data presented here can be found in chapter 3, and the final conclusions together with a summary in chapter 4.

2.1 The oldest attestations

The oldest attestations of reduplications are all post-runic. They belong to a wide array of just generally characterized dialects of the first centuries of the second millennium (Čigil (see Schönig 2004), Kipchak, Oghuz, Uighur, and others) but, it appears, are representative of none. Some interesting bits of information can be extracted from them, but in general they will be of very limited use for the present work.

Al-Kāšyārī's is a particularly precious evidence. He states quite clearly that *p* is the most common closing consonant, *m* is characteristic of Oghuz, and the one example in *s* is exceptional. See 2.1.1.

Four closing consonants are attested. The closers *pp* and *s* only have one example each, *m* has three, and the remaining 26 examples are all closed by *p*. The non-*p*-reduplications are probably all Oghuz (*kömkök*, *sümsüčig* and *tästägirmä*), and Kipchak (apparently *jamjašyl* and possibly *kömkök*). See 2.1.4.

As for semantics and parts of speech, the oldest attested reduplications are very standard with only the exception of *jazi*. See 2.1.5.

Special cases are not unusually numerous, but interesting. See especially the rather mysteriously alloyed *čimjġg*, and also the dialectally unclear *kömkök* in 2.1.3.

2.1.1 Sources

The main sources of both material and grammatical descriptions are: Clauson 1972 (ca. 10 000 entries), Dankoff/Kelly 1982, DTS (ca. 20 000 entries), Erdal 1991: 65f and 2004: 98, 150f, and Karamanlioğlu 1994: 97. Some attestations have also been found in Erdal 1998: 141, von Gabain 1950: 173, Hacıeminoğlu 1996: 61, Houtsma 1894, Malov 1951, Räsänen 1957: 74, and Röhrborn 1977–.

Clauson 1972 is ambiguous about the nature of reduplication. In some entries, he calls the reduplicated anlauts *alliterative prefixes*, but in some other *reduplicative* or *reduplicating prefixes* – which is in fact self-contradictory as prefixes have by definition a fixed phonetic shape – and makes his term *jingle* for ‘echoic compound’ look acceptable in comparison.

A much more useful source is Maḥmūd al-Kāšyarī’s *Compendium* (Dankoff/Kelly 1982: 261) who explains that the “rule about colors and exaggerating the description of things is to take the first letter of the word and join it to *bā*’ in most of the Turkic dialects, but to *mīm* in Oğuz”, and, one can understand, to subsequently also bring forward the original word. While perhaps slightly clumsy from the contemporary perspective, this description is only as imprecise as the greater part of formulations in modern grammars.

One important piece of information is that the Oghuz “change the *bā*’ to *mīm*” as it sets the terminus post quem for *m* as the closing consonant. The entry concludes with the statement “All exaggeratives are according to this rule. But there is no rule for changing *bā*’ to *sān*”. The latter surely refers to *tästāgirmā* which, as is earlier mentioned, “goes against the rule”. Apparently, *p* was the usual closing consonant for al-Kāšyarī, *m* he was familiar with, but *s* surprised him to some degree. This is perfectly in line with the modern distribution, too (see map 3.2).

Another interesting piece of information in Dankoff/Kelly 1982: 162 are the two words: *essiz* ‘alas’ and *arriy* ‘very clean’, both with a doubled consonant which al-Kāšyarī explains is “for exaggeration”. This supports the idea that the double *pp* in the general Tkc. *appak* and in Yak. *üppürüŋ*, both ‘very white’, is just an emphatic form of this kind, rather than a separate type of reduplication. See 3.1.8.

The most exhaustive descriptions can be found in Erdal 1991: 65f and 2004: 98, 150f. They give a precise account of the process complete with examples, but unfortunately without always specifying the exact dialect the words come from. Erdal 1991: 65 notes that colour names can only be intensified through reduplication and never by means of the *-rak* suffix (repeated in 2004: 150), which he believes cannot be a *meaningful complementary distribution* as many *lexemes are expanded both ways*. Rather, he

believes it to be a proof that *-rak* only intensifies in terms of grade and degree, which reduplication does not, and the reason why reduplication never serves comparison. The logical links between these statements are not entirely obvious to me. Nonetheless, the issue has no direct implication for our case.

Karamanlioğlu 1994: 97 refers to a later period. He gives a succinct description of the process of reduplication and some examples, and enumerates *p* and *m* as the only possible closing consonants.

To sum up, although a relatively high number of reduplications has been attested in the first centuries of the second millennium, their exact dialectal affiliation is often unclear and – unless we assume a late or post-mediaeval explosion of their popularity in the entire Turkic world – they form still too small a set to be considered representative. The fact that they are missing from the runic monuments and are far from being very common in the later ones, is probably to be explained by their intensive semantics which naturally limits the number of opportunities to use them, and can easily make them seem inappropriate in an official document.

Transcription:

| | |
|--|--|
| Clauson 1972: $\zeta \rightarrow \check{c} \mid d \rightarrow \delta \mid \check{g} \rightarrow \gamma \mid$ $\iota \rightarrow y \mid V: \rightarrow V,$ | Erdal 1998: $\gamma \rightarrow g \mid \check{i} \rightarrow y,$ |
| Dankoff/Kelly 1982: $d \rightarrow \delta \mid \eta \rightarrow \eta,$ | Erdal 2004: $\check{i} \rightarrow y,$ |
| DTS: $\rho \rightarrow \check{a} \mid \gamma \rightarrow g \mid \check{i} \rightarrow y \mid q \rightarrow k \mid$ $\check{z} \rightarrow z,$ | Hacıeminoğlu 1996: $\iota \rightarrow y \mid \check{s} \rightarrow \check{s},$ |
| Erdal 1991: $\zeta \rightarrow \check{c} \mid \iota \rightarrow y \mid \check{n}, \eta \rightarrow \eta \mid$ $\check{s} \rightarrow \check{s} \mid y \rightarrow j,$ | Karamanlioğlu 1994: $\iota \rightarrow y \mid k \rightarrow k \mid$ $\check{s} \rightarrow \check{s} \mid y \rightarrow j,$ |
| | Kuryšzanov 1970: $\theta \rightarrow \ddot{o},$ |
| | Röhrborn 1977–: $\iota \rightarrow y.$ |

2.1.2 Standard cases

ädgü (Erdal 1991: 66) ~ **ädgü** (DTS, Clauson 1972: 3, Dankoff/Kelly 1982: 87) ‘good’ \diamond **äpädgü** (Uighur; Erdal 1991: 66) ~ **äpädgü** (Khakani, Uighur; DTS, Clauson 1972: 3, Dankoff/Kelly 1982: 87)

ädgü see *ädgü*

ak ‘white’ \diamond **apak** (Chaghatai, Khwarezmian, Kipchak, Oghuz; DTS, Clauson 1972: 3, 75, Dankoff/Kelly 1982: 87, Houtsma 1894: 50) \diamond **appak** (Kipchak; Clauson 1972: 3, 75, Räsänen 1957: 74⁶)

alčak ‘friendly, mild’ \diamond **apalčak** (Uighur; Röhrborn 1977–)

aryg ‘clean’ \diamond **aparyg** (Uighur; DTS s.v. *ap* I and *süp*, Röhrborn 1977–, Erdal 1991: 65f, 1998: 141, 2004: 151)

⁶ Räsänen 1957: 74 derives *appak* from **ap-ak*.

äsän ‘healthy’ (Erdal 1991: 66, 2004: 151) ~ **esän** (DTS) ~ **esen** (Erdal 1998: 141)
 ◇ **äpäsän** ‘quite healthy’ (Uighur; Erdal 1991: 66, 2004: 151) ~ **epesän** (DTS)
 ~ **epesen** (Erdal 1998: 141)

esän see *äsän*

esen see *äsän*

jašyl ‘green’ ◇ **jamjašyl** (Azeri (?), Kipchak; Clauson 1972: 978, Houtsma 1894: 103, Karamanlioğlu 1994: 97) ◇ **japjašyl** (Khakani; Clauson 1972: 978, Erdal 1991: 65, Hacıeminoğlu 1996: 61, Karamanlioğlu 1994: 97)

See 2.1.4 below.

jazi see *jazy*

jazy (DTS: ‘1. steppe, plain; 2. outspread, open *of heart*’, Erdal 1991: 65: ‘flat’) ~ **jazi** (Dankoff/Kelly 1982: 261: ‘steppe, plain, open space’) ◇ **japjazy** (Uighur) ‘very smooth, very even (DTS); quite level (Erdal 1991: 65f)’ ~ **japjazi** ‘wide open space’ (Dankoff/Kelly 1982: 261)

jumšak ‘soft’ ◇ ***jupjumšak**⁷ (Uighur; Erdal 1991: 65f)

(j)ürün ‘white’ ◇ **(j)üp(j)ürün** (*j-*: Uighur; DTS, Erdal 1991: 65; *ü-*: Čigil; DTS, Clauson 1972: 3, Dankoff/Kelly 1982: 87, Erdal 1991: 65)

kara ‘black’ ◇ **kapkara** (Chaghatai, Khakani, Khwarezmian, Kipchak, Uighur; von Gabain 1950: 173, Malov 1951, DTS, Clauson 1972: 3, Erdal 1991: 65, 1998: 141, 2004: 151, Hacıeminoğlu 1996: 61, Houtsma 1894: 88, Karamanlioğlu 1994: 97)

karaŋu ‘dark’ ◇ ***kapkaraŋu**⁷ (Uighur; Erdal 1991: 65)

kök ‘blue (DTS), grey (Dankoff/Kelly 1982: 261)’ ◇ **kömkök** (Kipchak, Oghuz) ‘very blue (DTS, Erdal 1991: 65, Hacıeminoğlu 1996: 61), deep grey (Dankoff/Kelly 1982: 261, 267, Clauson 1972: 709, 721)’, Karamanlioğlu 1994: 97: no meaning given ~ **kün kök** (Houtsma 1894: 97) ◇ **köpkök** (Kipchak, Uighur) ‘very blue (DTS, Erdal 1991: 65, Hacıeminoğlu 1996: 61), deep grey (Dankoff/Kelly 1982: 261, Clauson 1972: 687, 709)’, Karamanlioğlu 1994: 97: no meaning given
 See *kömkök* in 2.1.3 below.

köni ‘straight’ ◇ **köpköni** (Uighur; Erdal 1991: 65f)

kötgi ‘protruding’ ◇ ***köpkötgi**⁷ (Uighur; Erdal 1991: 65)

kyrmyzy ‘red’ ◇ **kypkyrmyzy** (Chaghatai; Clauson 1972: 578)

kyzyl ‘red’ ◇ **kypkyzyl** (Khakani, Khwarezmian, Kipchak; Brockelmann 1954: 373: ⟨qip qizil⟩, Clauson 1972: 578, Erdal 1991: 65, Hacıeminoğlu 1996: 61, Houtsma 1894: 90, Karamanlioğlu 1994: 97)

saru see *saryg*

saryg (DTS, Clauson 1972: 848, Dankoff/Kelly 1982: 261) ~ **saru** (Clauson 1972: 848, Karamanlioğlu 1994: 97) ‘yellow’ ◇ **sapsaryg** (Chaghatai, Khakani; DTS,

⁷ The final shape of the reduplication is not given explicitly, and only deduced here from the description.

- Clauson 1972: 848, Dankoff/Kelly 1982: 261, Erdal 1991: 65) ~ **sapsaru** (Kipchak; Clauson 1972: 848, Karamanlioğlu 1994: 97)
- savuk** ‘cold’ (Toparlı 1993) ◇ **sapsavuk** (Kipchak; Karamanlioğlu 1994: 97)
- süçig** ‘sweet’ ◇ **sümsüçig** (Oghuz; DTS, Clauson 1972: 819, Dankoff/Kelly 1982: 267, Erdal 1991: 65)
- süzök** (Erdal 1991: 65) ~ **süzük** (DTS) ‘transparent’ ◇ **süpsüzök** (Uighur; Erdal 1991: 65f) ~ **süpsüzük** (DTS s.v. *ap* I and *süp*)
- süzük** see *süzök*
- tägirmä** (Dankoff/Kelly 1982: 261) ~ **tägirme** (Erdal 1991: 65) ~ **tägirmi** (Erdal 1991: 65f) ~ **tegirmä** (DTS) ‘round’ ◇ **täptägirme** ~ **täptägirmi** (Uighur; Erdal 1991: 65) ◇ **tästägirmä** (Oghuz; Dankoff/Kelly 1982: 261) ~ ***tästägirmi**⁷ (Oghuz; Erdal 1991: 65) ~ **testegirmä** (Oghuz; DTS)
- tägirme** see *tägirmä*
- tägirmi** see *tägirmä*
- tegirmä** see *tägirmä*
- tirig** ‘alive’ ◇ **tiptirig** (Uighur; DTS, Erdal 1991: 66)
- tolu** ‘full’ ◇ **toptolu** (Uighur; DTS, Erdal 1991: 65)
- tolun** ‘full’ ◇ ***toptolun**⁷ (Erdal 1991: 66)
- tutčy** ‘1. contiguous, osculant; 2. permanent, continuous’ ◇ **tuptutčy** ‘quite uninterruptedly, always’ (Uighur; Zieme 1985: 36, 238, Erdal 1991: 66, 1998: 141, 2004: 151)
- tüz** ‘straight, even, level’ ◇ **tüptüz** (Uighur; DTS s.v. *ap* I, *jap* III, *süp* and *tüp*, Erdal 1991: 65f, 1998: 141, 2004: 151)
- ürün** see (*j*)*ürün*
- uzun** ‘long’ ◇ **upuzun** (Uighur; Erdal 1991: 65f)

2.1.3 Special cases

- abam** (Uighur; Röhrborn 1977–) ~ **apam** (von Gabain 1950: 173) ~ **apan** (Uighur; Röhrborn 1977–, Erdal 2004: 341) ‘in case’ ◇ ? ***am** or ***aŋ** ‘now’
- The etymology deriving *abam* &c. from the word for ‘now’ was accepted by Röhrborn 1977– and Erdal 2004: 341; see both for the earlier history of the idea. Röhrborn believes that the base had the shape ***aŋ**, and Erdal that it was ***am**.
- This issue may be kept beyond the scope of the present work. The etymology is in fact no more than an unfinished idea. Even if it is true, the word *abam* &c. seems to be missing from the contemporary languages, which suggests that it was but a short-lived innovation within the Karakhanid group and as such, of very little importance for the general history of Turkic reduplication.
- čimjig** (DTS, Dankoff/Kelly 1982: 267) ◇ ***jig** ‘raw’ (Ölmez 1991: 181, Stachowski M. [in print])

This word is quite mysterious. It is attested by al-Kāšyarī who defines *čim* as ‘an exaggerative particle of dampness or rawness’ and gives two examples of use: *čim jīg at* ‘very raw meat’, and *čim ol tōn* ‘a very damp garment’ (DTS, Dankoff/Kelly 1982: 267). Unfortunately, the exact dialect is not specified; further examples can be found in Clauson 1972, see below.

It is not the case that al-Kāšyarī failed to recognize the mechanism of reduplication, for he explains it quite adequately s.v. *tās*, see 2.1.1 above. It is possible that his diagnosis is correct, though, even if the origin of *čim* remains unknown, because how a form such as *čimjīg* could have arisen through reduplication is not clear at all.

The initial *č-* is a Kipchak trait. The closing *m-* is more characteristic of Oghuz than of any other group; see 2.1.4 below. Modern Turkish and Turkmen have *čij* and *čīg*, respectively, which Stachowski M. [in print] suggests to be an Old Kipchak loanword or loanwords.

It seems that no scenario can plausibly explain *čimjīg* as long as Old Kipchak and Old Oghuz are considered separate, disjoint, languages. Probably, it would have to be attributed to some transitive or intermediary idiolect between Old Kipchak and Old Oghuz, perhaps one akin to that attested in Houtsma 1894 (see 2.1.4; only contains *چيك* ‘raw’). But even in this case, a somewhat acrobatic reconstruction would apparently be necessary, which could only be accepted if supported by a sizeable number of solid proofs.

The *čim* part might have been simply an independent word, as yet unidentified, but there is also another interesting theoretical possibility. In more than one language, one syllable words are observed which do not seem to possess a meaning of their own, and only serve to intensify a very limited number of adjectives, in particular Trkm. *čym* in *čym āk* ‘snow-white’ and *čym gyzyl* ‘bright red’. At least some of them are probably severed reduplicated anlauts reinterpreted as separate words (see 3.1.10). Our *čim* might be one such word, perhaps even extracted from **čimčīg?* – and/or related to Uigh. *žim* ‘quietly’ &c. (see 2.19.3), and possibly also to Bshk. *šym* ‘quiet’ (see 2.3.3)?

Clauson 1972: 424, 804 reports *čym* ~ *šym* in *čym ak* ‘plain white; snow-white’ and *čym* ~ *šym kara* ‘pure black’ (as opposed to *kap kara* ‘intensely black’), *čyŋ* in *čyŋ tolu* ‘full’ and *suk* in *suk jalŋuz* ~ *jalyuz* ‘lonely, isolated’. They all bear some resemblance to *čym*: they are one syllable long, end in a consonant that could be the closer in a reduplication, and are intensifiers which apparently can only be used with a limited group of words.

The list might be completed in a perhaps overly imaginative way with *tüg* ‘several’ in *tüg tümen* ‘several thousand’ but also ‘many myriads’ (Clauson 1972: 476, and explicitly considered a reduplication in Clauson 2002: 227), in order

to allow the observation that *čim* and *tüg* are combined with front words, while *čym* ~ *čyŋ* and *suk* – with back ones.

See 3.4.4 for a continuation of these considerations.

japjavyšgu (Uighur; Erdal 1991: 66) ◊ *javyšgu* ‘leaves’

Erdal 1991: 66 mentions this case in the section on reduplication but eventually discards it accepting Ş. Tekin’s idea that the phrase should be completed as *japyrgak javyšgu*, “since *yap* [= *jap*] is written at the end of the line, and the rest of the word may have been omitted inadvertently in writing on in the next one”.

A much more solid ground for rejection of this example is found in DTS which attests the phrase with a Px3Sg as *japy javyšgusy* ‘leaves’ and thus proves it is simply a binomial pair.

kömkök (Kipchak, Oghuz) ‘very blue (DTS, Erdal 1991: 65, Hacıeminoğlu 1996: 61), deep grey (Dankoff/Kelly 1982: 261, 267, Clauson 1972: 709, 721)’, Kara~~manlioğlu~~ 1994: 97: no meaning given ~ **kün kök** (Houtsma 1894: 97) ◊ **kök** ‘blue (DTS), grey (Dankoff/Kelly 1982: 261)’

Kömkök, unlike its sister reduplication *köpkök*, is somewhat moot.

Al-Kāšyarī ascribes the form to Oghuz (Dankoff/Kelly 1982: 261, 267), and does not mention Kipchak at all. But the shape with *m* appears also in Houtsma 1894 which primarily contains Kipchak words with only an admixture of Turk~~men~~ (Stachowski M. 2010: 130). The spelling is also unusual: *kün kök* (Houtsma 1894: 97).

However, Kuryšzanov 1970 notices that in the Arabic version of the manuscript published by Houtsma, the word is written كوز كوك (Houtsma 1894: ٣١), which Kuryšzanov reads *közkök* and considers the *m* ~ *n* form to be a misprint. Also Clauson 1972: 709 gives the *z* variant, but with a “sic” and no more commentary.

Both *közkök* and *künkök* would have been unique shapes. *Könkök* or *könkök* appear to be more plausible as assimilated variants of *kömkök*. But beside the reading, also the question of affiliation remains open, for Houtsma 1894 is not dialectally uniform. For the present purpose, I will include *kömkök* because it is also clearly attested elsewhere, and disregard *közkök*.

Karamanlioğlu 1994: 97 notes that *köpkök* is the ‘original’ form, but puts the word *aslı* in quotation marks. Unfortunately, he does not elaborate on this remark. Clauson 1972: 978 makes a parallelly ambiguous remark on *jamjašyl* : *japjašyl*, see 2.1.4 below.

oposalkyja ‘without care’ (Uighur; Erdal 1991: 65f) ◊ *osal* ‘negligent, idle; negligence, idleness’ (Clauson 1972: 247)

Erdal 1991: 65f notes that *oposalkyja* “heaps up two emotive elements”. One is the reduplicated anlaut (*op-*), and the other, I gather, must be *-kyja* DIMIN., whose usual form in Uighur is *-kyja* (Erdal 1991: 48). However, neither **osalkyja* nor **oposal* seem to be attested. Lacking parallel examples from Uighur, it is

not possible to determine whether it was the two intensifiers that were attached simultaneously, or simply an attestation of the singly intensified form that does not seem to have survived.

See 3.1.13 for more examples of multiple intensification.

symsymrak ‘dish of meat cut up small’ (Čigil; DTS, Clauson 1972: 830)

The word is unclear and perhaps not a reduplication at all. However, consider~~ing~~ *simür-* ‘to swallow in a single gulp’ (Clauson 1972: 829) \sim *sümür-* ‘to gulp down’ (Dankoff/Kelly 1982: 171), the form **simür.ak* \sim **sümür.ak* might be assumed with the meaning ‘bite-sized piece’. To it, three processes would have been applied in a probably unreconstructable order: 1. dropping of the middle high vowel (*u* or *ü*), 2. shifting to the back harmony (for a possible parallel, see Stachowski M. [in print] s.v. *biçmek* and *bıçak*), and 3. reduplication. The se~~quence~~ would eventually yield *symsymrak*.

Reduplication of nouns is very rare but not impossible, see 3.1.14, and es~~pecially~~ Az. *bärbäzäk* ‘decorations, ...’ < *bäzäk* ‘decoration, ...’ and *sörsöküntü* ‘chips, splinters’ < *söküntü* ‘chip, splinter’ (2.2.2 and 2.2.5) where, importantly, the meaning of reduplication is in fact pluralization. This fits our case perfectly as it would enable the then-trivial shift ‘bite-sized pieces’ \rightarrow ‘dish of meat cut up small’.

Finally, it is not in fact very unusual for the closing consonant to be identical to C_2 . See 3.1.6 for more examples.

2.1.4 Structure

Four closing consonants of C -type are attested in a total of 31 examples derived from 27 unique bases, in a quite even distribution:

m: 3 examples: *jašyl*, *kök*, and *süčig*,

p: 26 examples: *ädgü*, *ak*, *alčak*, *aryg*, *äsän*, *jašyl*, *jazy*, *jumšak*, *(j)ürüñ*, *kara*, *karañu*, *kök*, *köni*, *köti*, *kyrmyzy*, *kyzyl*, *saryg*, *savuk*, *süzök*, *tägirmä*, *tirig*, *tolu*, *tolun*, *tutčy*, *tüz*, and *uzun*,

pp: 1 example: *ak*, and

s: 1 example: *tägirmä*.

Four words have more than one closing consonant possible: *ak* (p and pp), *jašyl* (m and p), *kök* (m and p) and *tägirmä* \sim *tägirmi* \sim *tegirmä* (p and s), i.e. all have p as one of the possibilities. The only word whose reduplication cannot be closed by p is *süčig*.

Two of three examples with m (*kök* and *süčig*), and the only one with s (*tägirmä*), are all marked by al-Käšyarī as Oghuz. Some other sources adduce *kömkök* as a Kipchak shape, but this is not surprising in light of the modern Bshk. and Tat. *kümkük*. The one remaining example with m , *jašyl*, Clauson 1972 ascribes to Kipchak but immediately

noting that *japjašyl* is ‘more correct’. It is not clear to me, what this observation was intended to mean. Karamanlioğlu gives a parallelly ambiguous remark on *kömkök* : *köpkök*; see 2.1.3 above.

It appears that as far as reduplication is concerned, it could be beneficial to assume the existence of some intermediary idiolect between Old Kipchak and Old Oghuz, perhaps one similar to that attested in Houtsma 1894. See *čimjīg* and *kömkök* in 2.1.3 above.

2.1.5 Semantics

In all cases the reduplicated meaning is a simple intensification of the base meaning. With eight examples out of 26, colour names are the most numerous group, and perhaps more numerous than elsewhere, but they certainly do not monopolize the stock.

Almost all examples are of a primarily adjectival character. The only exception is *jazy* which, together with its reduplication *japjazy*, can act as both an adjective and a noun: ‘flat, level’ → ‘intens.’, and ‘steppe, plain, open space’ → ‘wide open space’.

2.2 Azeri

C-type reduplications have flourished in Azeri. While less numerous, they are clearly more diversified than in the neighbouring languages.

There are as many as six closing consonants and although, admittedly, three of them have very few examples (eight in total), among the other three, the domination of *p* is much less overwhelming than in some other languages, as *p*, *m* and *r* have 30, 19 and 12 examples, respectively. See 2.2.4.

Reduplications of what are more nouns than adjectives are not uncommon. Indeed, in two cases, reduplication has expanded its meaning from the original intensification onto pluralization. See 2.2.5.

Also, unusually numerous among Azeri reduplications, are derivatives. The ‘order of operations’ (of derivation and reduplication) is generally impossible to reconstruct. See 2.2.5.

Finally, noteworthy among the special cases are the words *ayappag* (probably a reduplication with the base prepended to it: *ay.a.pp.ag*), and *garyš* as a representative of a larger, and quite characteristic of Azeri, family of words. See 2.2.3.

2.2.1 Sources

The main source of the Azeri material is AzRS which contains ca. 70 830 entries. Some attestations have also been found in Simpson 1957: 15, Širaliev/Sevortjan 1971: 61, 68f, Ščerbak 1977: 120, Schönig 1998a: 251, Tenišev 1988: 155, and Žäfärov 1984: 49.

For grammatical descriptions, Širaliev/Sevortjan 1971, Budagov 1987: 51, and Zeynalov 1993: 149f have been used.

In Širaliev/Sevortjan 1971: 61, 68f, all description is in fact reduced to listings of examples.

Budagov 1987: 51 lists the possible closing consonants: *g* [sic], *m*, *p*, *r*, and *s*, but unfortunately gives no examples. I was not able to find any form that could be interpreted as a reduplication closed with *g*.

Zeynalov 1993: 149f only states that the intensive form is created with the help of *m*, *p*, *r*, and *s*, and gives some examples.

Transcription:

AzRS, Simpson 1957: *c* → *č* | *ç* → *č* | *a* → *ä* | *ğ* → *γ* | *x* → *χ* | *ι* → *y* | *q* → *g* | *ş* → *š* | *y* → *j*,
 Ščerbak 1977: *ɣ* → *g* | *i* → *y*,
 Schönig 1998a: *i* → *y* | *e* → *ä* | *é* → *e* | *y* → *j*,
 Širaliev/Sevortjan 1971: *a* → *ä* | *ø* → *ö*,
 Zeynalov 1993: *ğ* → *γ* | *ι* → *y* | *ş* → *š* | *y* → *j*,
 Žäfärov 1984: *z* → *γ* | *ø* → *ö*.

2.2.2 Standard cases

ačyγ i.a. ‘clear, bright’ ◊ **apačyγ** (AzRS)

aγ ‘white’ ◊ **apaγ** (AzRS)

See *aγappag* in 2.2.3 below.

ajdyn ‘bright, clear, clean’ ◊ **apajdyn** (AzRS)

balača see *balaža*

balaža ‘small, tiny’ ◊ **bambalaža** (AzRS) ◊ **bapbalaža** (AzRS) ~ **bapbalača** ‘very small, very young’ (Simpson 1957: 15)

bašga ‘different, other’ ◊ **bambašga** (AzRS)

bäzäk ‘decoration, decorative’ ◊ **bärbäzäk** ‘1. decorations *with a large number of knick-knacks*; 2. luxury’ (AzRS)

See *bärbäzaklik* in 2.2.3 below, and also 2.2.5 below.

bäzäkli ‘decorated, fancy, chic’ ◊ **bärbäzäkli** ‘1. *intens.*; 2. luxurious’ (AzRS)

See *bärbäzaklik* in 2.2.3 below, and also 2.2.5 below.

betär ‘worse’ ◊ **bešbetär** (AzRS)

See also 2.2.4 below.

biz ‘awl’ ◊ **bimbiz** ‘sharp, pointed, spiky’ (AzRS)

boš ‘empty’ ◊ **bomboš** (Simpson 1957: 15, Zeynalov 1993: 150, AzRS)

boz ‘grey, ashen’ ◊ **bomboz** (Širaliev/Sevortjan 1971: 68f, Žäfärov 1984: 49, AzRS)

bulašyγ ‘stained, dirty’ ◊ **bumbulašyγ** (AzRS)

bütün ‘all, whole’ ◊ **büsbütün** ‘completely, entirely’ (Širaliev/Sevortjan 1971: 69, Ščerbak 1977: 120, Tenišev 1988: 155, Zeynalov 1993: 150, AzRS)

- buz** ‘cold, frosty, chilly’ ◇ *bumbuz* (Zeynalov 1993: 150, AzRS)
See *bumbuzlug* in 2.2.3 below.
- däm** i.a. ‘time, moment’ ◇ *därdäm* ‘1. immediately; 2. suddenly’ (AzRS)
- dik** ‘vertical, steep’ ◇ *dimdik* ‘1. apeak, on end; 2. in full length; 3. beak’ (AzRS)
- diri** ‘(a)live, lively’ ◇ *dimdiri* (AzRS)
- doγru** ‘1. true, faithful, accurate; 2. straight, direct’ ◇ *dopdoγru* (AzRS) ◇ *dosdoγru* (AzRS)
- dolu** ‘full’ ◇ *dopdolu* (AzRS)
- duru** ‘1. liquid; 2. clear, clean’ ◇ *dumduru* (AzRS)
- durulug** i.a. ‘transparency’ ◇ *dumdurulug* (AzRS)
- düz** ‘smooth, even, straight’ ◇ *dümdüz* (AzRS)
- farayat** ‘calm(ly), meek(ly)’ ◇ *fasfarayat* ‘calmly, meekly’ (AzRS)
- gara** ‘black, dark’ ◇ *gapgara* (Širaliev/Sevortjan 1971: 61, 69, Žäfärov 1984: 49, Teñišeŷ 1988: 155, Zeynalov 1993: 150, AzRS)
See also 2.2.4 below.
- garalyg** ‘black(ness)’ ◇ *gapgaralyg* (AzRS)
- garanlyg** ‘dark(ness)’ ◇ *gapgaranlyg* (AzRS)
- garyšyg** ‘mixed, varied, mingled, messy’ ◇ *gatgaryšyg* ‘1. mixed, complicated, entangled; 2. scattered, disordered; 3. uncombed’ (AzRS)
See *gatgaryš* in 2.2.3 below.
- garyšyglyg** ‘1. entanglement, confusion; 2. disorder, unrest, stir’ ◇ *gapgaryšyglyg* ‘1. entanglement, confusion; 2. dispersion, disorder’ (AzRS)
See *gatgaryš* in 2.2.3 below.
- girdä** ‘round’ ◇ *gipgirdä* (AzRS)
- göj** ‘1. blue; 2. green’ ◇ *gömgöj* ‘1. intens.; 2. dark blue; 3. dark green’ (Ščerbak 1977: 120, Zeynalov 1993: 150) ~ *kömköj* ‘very blue’ (Širaliev/Sevortjan 1971: 69, Žäfärov 1984: 49) ~ *kömkök* (Tenišeŷ 1988: 155)
- göjlük** ‘1. azure, colour blue; 2. a (very) green area’ ◇ *gömgöjlük* ‘1. azure, the blue colour of something; 3. greenness, colour green; 4. a very green area’ (AzRS)
- guru** ‘dry’ ◇ *gupguru* (Simpson 1957: 15, Širaliev/Sevortjan 1971: 69, Tenišeŷ 1988: 155, AzRS)
See also 2.2.4 below.
- gyrmyzy** ‘red’ ◇ *gyppgyrmyzy* ‘1. intens.; 2. light red; 3. boldly, shamelessly’ (Simpson 1957: 15, Širaliev/Sevortjan 1971: 61, 68f, Ščerbak 1977: 120, Žäfärov 1984: 49, Tenišeŷ 1988: 155, Zeynalov 1993: 150, AzRS)
- gyvrag** ‘1. clever, glib, jaunty; 2. agile, nimble; 3. cheerful, buoyant, healthy; 4. taut, natty’ ◇ *gysgyorag* ‘1. taut, natty, neat, trim; 2. calm’ (AzRS)
- gyvraglyg** ‘1. glibness; 2. freshness, cheerfulness; 3. agility; 4. restlessness’ ◇ *gysgyvraglyg* ‘smartness, tautness’ (AzRS)

jalgyz ‘lone(ly)’ ◇ **japjalgyz** (AzRS)

See *jalnyz* below, and 3.1.11 on families of reduplications.

jalnyz ‘lone(ly)’ ◇ **japjalnyz** (AzRS)

See *jalnyz* below, and 3.1.11 on families of reduplications.

jalnyzža ‘completely alone, completely lone(ly)’ ◇ **japjalnyzža** (AzRS)

jasty ‘flat’ ◇ **jamjasty** (AzRS)

jašyl ‘green’ ◇ **jamjašyl** (Širaliev/Sevortjan 1971: 61, 69, Žäfärov 1984: 49, Tenišev 1988: 155, Zeynalov 1993: 150, Schönig 1998a: 251, AzRS)

jekä ‘1. big, large, great; 2. adult’ ◇ **jesjekä** (AzRS)

jeni ‘new’ ◇ **jepjeni** (AzRS)

joxsul ‘poor, destitute’ ◇ **jorjoxsul** ‘the poor, the squalid, beggars’ (AzRS)

joxsullug ‘poverty, destitution’ ◇ **jorjoxsullug** (AzRS)

jumru ‘round’ ◇ **jupjumru** (AzRS)

jumruža ‘very round’ ◇ **jupjumruža** (AzRS)

jumšag ‘soft’ ◇ **jumjumšag** (AzRS)

See 3.1.6 on the closer being identical to C_2 .

kobud ‘rough, coarse, crude’ ◇ **korkobud** (AzRS)

köj see *göj*

kök see *göj*

lüt ‘naked, bare’ ◇ **lümlüt** ‘1. *intens.*; 2. poor, destitute, beggar’ (AzRS)

say i.a. ‘1. healthy; 2. unscathed, sound’ ◇ **sapsay** (AzRS)

saylam ‘healthy’ ◇ **sapsaylam** (AzRS)

saylamlyg ‘health’ ◇ **sapsaylamlyg** ‘1. the state of being in good health; 2. vim, good health’ (AzRS)

saylyg ‘health’ ◇ **sapsaylyg** ‘1. the state of being in good health; 2. vim, good health’ (AzRS)

sary ‘yellow’ ◇ **sapsary** ‘1. *intens.*; 2. bright yellow’ (Širaliev/Sevortjan 1971: 61, 69, Žäfärov 1984: 49, Tenišev 1988: 155, Zeynalov 1993: 150, AzRS)

See also 2.2.4 below.

sarylyg ‘yellowness, yellow colour’ ◇ **sapsarylyg** ‘1. very yellow colour; 2. bright yellow colour; 3. paleness; 4. soreness’ (AzRS)

širin ‘sweet’ ◇ **šipširin** (AzRS)

širinlik ‘sweetness’ ◇ **šipširinlik** (AzRS)

šit ‘1. sweet, not savoury; 2. not salted enough’ ◇ **šipšit** ‘(completely) unsalted of *butter and dishes*’ (AzRS)

šitlik ‘insufficient salting’ ◇ **šipšitlik** (AzRS)

söküntü ‘chip, splinter’ ◇ **sörsöküntü** ‘chips, splinters’ (AzRS)

See also 2.2.5 below.

täläsik ‘hastily, hurriedly’ ◇ **tärtäläsik** (AzRS)

See *tez-täläsik* in 2.2.3 below.

tämiz ‘clean’ ◊ **tärtämiz** (Širaliev/Sevortjan 1971: 69, Ščerbak 1977: 120, Tenišev 1988: 155, AzRS)

tämizlik ‘cleanness’ ◊ **tärtämizlik** (AzRS)

tüzä ‘1. new; 2. fresh’ ◊ **täptüzä** ‘brand-new’ (Širaliev/Sevortjan 1971: 61, Zeynalov 1993: 150, Schönig 1998a: 251, AzRS) ◊ **tärtüzä** (AzRS)

tüzäzä ‘1. brand-new; 2. very fresh’ ◊ **tärtüzäzä** (AzRS)

turş ‘sour’ ◊ **tumturş** (Širaliev/Sevortjan 1971: 69, AzRS)

turşlug ‘sourness’ ◊ **tumturşlug** (AzRS)

2.2.3 Special cases

aγappag (Širaliev/Sevortjan 1971: 61, 69, Žäfärov 1984: 49, Zeynalov 1993: 150, AzRS) ◊ **aγ** ‘white’

This word appears to be a petrified compound **aγ* + *a.pp.ag*, of a structure rather similar to that of Russ. *белый-пребелый*. It is not clear, however, why it has -*g* rather than -*γ* in auslaut, and why the reduplication is closed with a double rather than a single *p* (in itself, **appag* does not seem to be attested in Azeri). Possibly a borrowing? See 3.1.16 on reduplications with a prepended base.

bärbäzäklik ‘1. excessive, kitsch decoration; 2. taste for luxury’ (AzRS)

Although **bäzäklik* does not appear to be attested, the form seems to be very likely, given the number of derivative reduplications in Azeri (see 2.2.4). The root *bäzäk* is attested together with its derivatives in -*čy*(*lyg*), -*länmäg*, -*ly*(*lyg*), and -*syzlyg*, and two reduplications (*bärbäzäk* and *bärbäzäkli*). Meanings seem to match, too.

However, reduplications with unattested base are in the present work listed among other examples but excluded from further considerations (see 1.2.4).

bumbuzlug ‘excessive cold’ (AzRS)

The word seems to be one of the many derivatives from reduplications that can be found in Azeri (see 2.2.4). However, the base *buzlug* seems to only be attested in the meaning ‘icehouse, icebox’, which makes it clear that the -*lyg* suffix was added to the reduplicated form, not the other way round, and thus eliminates this word from further considerations as a derivative from a reduplication, not a reduplication of a derivative.

gatgaryş ‘= *gatgaryşyg*’ (AzRS)

The phonetic shape *garyş* is only attested with the meaning ‘span (length unit)’, which is a different word altogether; see e.g. ÈSTJa. The **garyş* here appears to be a derivative in -*yş* (see Sevortjan 1966: 140–52) from *kar-* ‘to mix, to stir’ with the meaning *‘mixing, mix-up, confusion’.

Related to it are two non-reduplicated forms: *garyşyg* (diminutive, see Sevortjan 1966: 166–168) and *garyşglyg* (both see 2.2.2 above), and four reduplications: *gatgaryşyg*, *gatgaryşglyg*, *gatgaryş*, and *gatgaryşlyg*.

In the case of the first two reduplications, the bases are attested independently, and so it was assumed here that they were derived first, and reduplicated later.

In the other two cases, the bases seem to be unattested. The form *gatgaryšlyg* can be explained as a derivative of *gatgaryš*, which must be a reduplication of **garyš* which, in turn, must be the missing link between *kar* and *garyšyg*.

Alternately, *gatgaryšlyg* can be viewed as a reduplication of the unattested **garyšlyg*. In either case, it will have to be excluded from further considerations here.

The family can be presented schematically as in fig. 2.1.

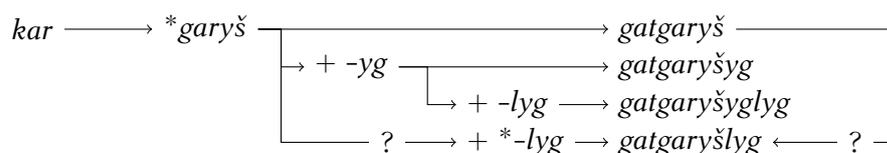


Figure 2.1: Reduplications of Az. **garyš* i.a. ‘mixed, scattered’ &c.

Apparently, the original base has worn out over time and been replaced by its more emphatic derivatives, a reduplication and a diminutive. See 3.1.13.

See *garyšyg* and *garyšyglyg* in 2.2.2 above, *gatgaryšlyg* below, and also 3.1.11 for other families of related reduplications.

gatgaryšlyg ‘turmoil, chaos’

See *gatgaryš* above.

tez-täläsik ‘hurriedly, hastily, urgently, slapdash’ (AzRS)

At first sight, this form might be taken for a reduplication. The strange lowering of the alleged reduplicated vowel, however, and the use of *z* for the closer, especially before a voiceless consonant, suggest that it is much more likely to be in fact a simple composition of *tez* ‘quickly, soon, precipitously’ + *täläsik* ‘hastily, speedily, slapdash’.

2.2.4 Structure

Six closing consonants of *C*-type are attested in a total of 70 examples derived from 67 unique bases, in a quite uneven distribution:

m: 20 examples: *balaža*, *bašga*, *biz*, *boš*, *boz*, *bulašyg*, *buz*, *dik*, *diri*, *duru*, *durulug*, *düz*, *göj*, *göjlük*, *jasty*, *jašyl*, *jumšag*, *lüt*, *turš*, and *turšlug*,

p: 30 examples: *ačyg*, *aγ*, *ajdyn*, *balaža*, *doyru*, *dolu*, *gara*, *garalyg*, *garanlyg*, *garyšyglyg*, *girdä*, *guru*, *gyrmyzy*, *jalgyz*, *jalnyz*, *jalnyzža*, *jeni*, *jumru*, *jumruža*, *say*, *saylam*, *saylamlyg*, *saylyg*, *sary*, *sarylyg*, *širin*, *širinlik*, *šit*, *šitlik*, and *täzä*,

r: 12 examples: *bäzäk*, *bäzäkli*, *däm*, *joχsul*, *joχsullug*, *kobud*, *söküntü*, *täläsik*, *tämiz*, *tämizlik*, *täzä*, and *täzäžä*,

s: 6 examples: *bütün*, *doγru*, *farayət*, *gyorag*, *gyvraglyg*, and *jekä*,

š: 1 example: *betär*, and

t: 1 example: *garyšyg*.

Three words have more than one closing consonant possible: *balaža* (*m* and *p*), *doγru* (*p* and *s*) and *täzä* (*p* and *r*). The only regularity is that all have *p* and some other consonant.

The closer *š* in *bešbetär* is highly unusual. The only other languages to allow a reduplication closed with *š* are Kumyk and Turkish, and it is in the exact same word; see *beter* in 2.11.2 and 2.16.2.

Also *jumjumšag* is a peculiar form; see 3.1.6 for other reduplications with the closing consonant identical to C_2 .

Dialectally, *p* can be replaced by *f*: *gafgara*, *gufguru* and *safsary* ‘completely: black, dry and yellow, respectively’ (Ščerbak 1977: 120). See 3.1.21 for other examples of spirantized closing consonants.

2.2.5 Semantics

In the great majority of cases, the reduplicated meaning is a simple intensification or apparently the same as the base meaning. The latter case should probably be attributed to dictionary definitions not being sufficiently exhaustive.

The majority of base words are closer to being an adjective, if one agrees to regard Turkic nomina as a continuous scale. Some exhibit a more dual character (e.g. *gara*, *joχsul*), and a few lean closer to nouns (e.g. *biz*, *bütün*). The meanings of their reduplications, however, are not typologically surprising, with two clear exceptions (see also *joχsul* in 2.2.2 for a similar but less distinct example):

Bäzäk has both a substantival and an adjectival meaning. We will ignore the latter because it is not present in the reduplicated meaning. *Söküntü* is purely substantival. In both cases, the reduplicated meaning appears to be a plural of the meaning of the base (AzRS):

bäzäk

1. украшение: 1.1. предметы украшения, драгоценности; 1.2. тот, кто своей деятельностью, участием придаёт особую ценность чему-л.; 2. узор, рисунок, орнамент; 3. отделка; 4. убранство

bärbäzäk

1. украшения (о большом количестве разнообразных безделушек); 2. роскошь

söküntü

обламок (остаток чего-л. прежде существовавшего, исчезнувшего)

sörsöküntü

обломки (остатки чего-л. существовавшего, разрушенного)

Also characteristic of Azeri is that unusually many, seventeen, words are derivatives: *büzäkli*, *durulug*, *garalyg*, *garyşyglyg*, *göjlük*, *gyoraglyg*, *jalnyzža*, *joxsullug*, *jumruža*, *saylamlyg*, *saylyg*, *sarylyg*, *şirinlik*, *şitlik*, *tämizlik*, *tüzäžä* and *turşlug*. Thirteen of them are in *-lyg*, three are in *-ža* and one is in *-ly*.

While in some cases a (shaky) conjecture could be ventured based on semantics, it is generally not possible to determine whether the derivation happened before or after the reduplication. All suffixes are still very productive today and new formations with them are created with great ease. It is doubtful that even the most meticulous examination of historical sources could reveal the truth.

Here, they are all considered reduplications of derivatives because all the bases are attested independently. See in particular *gatgaryş* in 2.2.3.

2.3 Bashkir

Bashkir *C*-type reduplications are very numerous but hardly diversified. This is consistent with the state in the Kipchak languages in general.

Against 91 examples closed with *p* or *pp*, there are only two closed with *m*, and no other closing consonants appear to be attested. Moreover, one of these two examples can alternately be closed with *p*, too. See 2.3.4.

Also semantically are Bashkir reduplications very standard. Only two words are not obviously adjectives, and just one is slightly special in that its base is already an intensification itself. See 2.3.5.

2.3.1 Sources

The main sources of the material are BškRS58 and BškRS96 which contain ca. 22 000 and 32 000 entries, respectively. Some attestations have also been found in the grammars listed below.

For grammatical descriptions, Ahmerov 1958: 766 and Juldašev 1981: 196 have been used.

Ahmerov 1958: 766 gives a description of the formation of reduplications but the question of the closing consonant is reduced to the statement that in most cases, the reduplicated syllable is closed by *p* and only very rarely by *m*. This, however, is followed by an affirmation that *исторически эти формы возникли из полного повторения* which, he believes, is confirmed by the fact that full reduplications are used in modern Bashkir (*jakšy-jakšy jorttar* ‘very beautiful houses’, *maturdarduŋ-matury* ‘very beautiful’, &c.).

In Juldašev 1981: 196, the description of the closing consonant is the same, and supplemented by the information that the stress falls on the reduplicated syllable. Also,

a few interesting examples are provided. Surprisingly, *örjanġy* is included on par with its actual reduplication, *öpörjanġy*; see 2.3.5.

Transcription:

Ahmerov 1958, BškRS58, BškRS96, Juldašev 1981: $a \rightarrow ä$ | $\varepsilon \rightarrow \gamma$ | $h \rightarrow h$ | $\kappa \rightarrow k$ | $\eta \rightarrow \eta$ |
 $e \rightarrow ö$ | $\zeta \rightarrow \theta$ | $Vy \rightarrow Vw$ | $\gamma \rightarrow ü$ | $V\gamma \rightarrow V\ddot{w}$ | $\mathfrak{z} \rightarrow \delta$

2.3.2 Standard cases

äd ‘little, few’ \diamond **äpäð** (BškRS96)

ak i.a. ‘1. white; 2. good, kind’ \diamond **apak** ‘snow-white’ (Ahmerov 1958: 766, BškRS58)
 \diamond ***appak** (BškRS96: only attested in *appaġym* ‘my dear, my darling’)

See *appaġym* and also *apakaj* in 2.3.3 below.

äkren i.a. ‘quietly’ \diamond **äpäkren** (BškRS96)

aryw ‘good, passable’ \diamond **aparyw** ‘best, good (s.v. *an-*); quite good, not bad, decent, passable (s.v. *an-apby*)’ (BškRS96)

äse ‘1. bitter; 2. sour; 3. pungent’ (BškRS58) \diamond **äpäse** ‘1. very bitter; 2. very sour’ (BškRS58, BškRS96)

asyk i.a. ‘clear, bright, precise’ \diamond **apasyk** (BškRS58, BškRS96)

bäläkäj ‘very small’ \diamond **bäpbäläkäj** (Ahmerov 1958: 766, BškRS58, BškRS96)

bojok ‘dull’ \diamond **bopbojok** (BškRS96)

bökrö ‘slouching’ \diamond **böpbökrö** (BškRS96)

bötön ‘whole’ \diamond **böpbötön** (BškRS96)

buš ‘empty’ \diamond **bupbuš** (BškRS58, BškRS96)

byrsak ‘dirt(y), slush(y)’ \diamond **bypbyrsak** ‘very dirty’ (BškRS96)

haj ‘1. small, fine; 2. shallow’ \diamond **haphaj** (Ahmerov 1958: 766, BškRS96)

halkyn ‘cold, frosty, chilly’ \diamond **haphalkyn** (BškRS58)

hary ‘yellow’ \diamond **haphary** (Ahmerov 1958: 766, BškRS58)

haw ‘healthy’ \diamond **haphaw** (BškRS96)

hiräk ‘rare, sparse’ \diamond **hiphiräk** (BškRS96)

horo ‘(dark) brown (BškRS58), (light) brown (BškRS96)’ \diamond **hophoro** (BškRS58, BškRS96)

jakšy ‘good’ \diamond **japjakšy** (BškRS96)

jakty ‘light(ing), bright’ \diamond **japjakty** ‘very light, very bright’ (BškRS96)

jakyn ‘close, near’ \diamond **japjakyn** (BškRS96)

jaltyr ‘shine, gloss, glaze’ \diamond **japjaltyr** (BškRS96)

janġy ‘new, fresh’ \diamond **japjanġy** (BškRS96)

janġyð ‘lone(ly)’ \diamond **japjanġyð** (Juldašev 1981: 196)

jäš ‘young’ \diamond **jäpjäš** (BškRS58, BškRS96)

- jäšel** ‘(light) green’ (in BškRS96 only attested in derivatives) ◇ **jämjäšel** (BškRS58, Juldašev 1981: 196, BškRS96) ◇ **jäpjäšel** (BškRS96)
- jeŋel** ‘light, easy’ ◇ **jepjeŋel** (BškRS58, Juldašev 1981: 196)
- jeŋeš** i.a. ‘wet’ ◇ **jepjeŋeš** (BškRS96)
- joka** ‘thin’ ◇ **jopjoka** (BškRS96)
- jomro** ‘round’ ◇ **jopjomro** (BškRS96)
- jomšak** ‘soft’ ◇ **jopjomšak** (BškRS58)
- jyltyr** ‘brilliant, sparkling’ ◇ **jypjyltyr** (BškRS96)
- lyly** ‘tepid, warm’ ◇ **jyplyly** (BškRS58, BškRS96)
- kak** ‘1. naked, bare; 2. very slim, skinny’ ◇ **kapkak** (BškRS58)
- kara** ‘black’ ◇ **kapkara** (Ahmerov 1958: 766, BškRS58, Juldašev 1981: 196, BškRS96)
- karaŋgy** ‘dark’ ◇ **kapkaraŋgy** (BškRS96)
- katy** ‘hard, solid’ ◇ **kapkaty** (BškRS96)
- körän** ‘brown’ ◇ **köpkörän** (BškRS96)
- koro** ‘dry’ ◇ **kopkoro** (BškRS58, Juldašev 1981: 196, BškRS96)
- kujy** ‘thick, dense’ ◇ **kupkujy** (BškRS96: only attested in ~ *aš* ‘very thick soup’, and ~ *itep bešereü* ‘to cook something very thick’)
- kük** ‘(dark) blue’ ◇ **kümkük** (Ahmerov 1958: 766, BškRS58, Juldašev 1981: 196, BškRS96)
- kyθka** ‘short’ ◇ **kypkyθka** (BškRS58)
- kyzyl** ‘red, ruddy’ ◇ **kypkyzyl** ‘scarlet, crimson, red’⁸ (Ahmerov 1958: 766, BškRS58)
- matur** ‘beautiful’ ◇ **mapmatur** (BškRS96)
- nädek** ‘thin’ ◇ **näpnädek** (BškRS58, BškRS96)
- nakyθ** ‘1. small, short (clothes); 2. shortage’ ◇ **napnakyθ** ‘very little’ (BškRS58, where marked as *superlative*)
- nasar** ‘bad, nasty’ ◇ **napnasar** (BškRS58, BškRS96)
- neskä** ‘thin’ ◇ **nepneskä** (BškRS58, BškRS96: only attested in ~ *bil* ‘very thin waist’, and ~ *kyrkyw* ‘to chop finely’)
- ör-jaŋgy** ‘brand-new’ ◇ **öpörjaŋgy** (Juldašev 1981: 196) See *öpörjaŋgy* in 2.3.3 below.
- šaktaj** ‘quite, significantly, very’ ◇ **šapšaktaj** (BškRS96: only attested in ~ *bala* ‘big, healthy child’)
- salyš** ‘oblique, slanted, crooked; sideways’ ◇ **sapsalyš** ‘ramose, forked’ (BškRS96)
- šärä** ‘naked’ ◇ **šäpšärä** (BškRS58, BškRS96)
- sej** ‘crude, raw’ ◇ **sepsej** (BškRS58, BškRS96)
- serek** ‘rot(ten)’ ◇ **sepserek** ‘rotten’ (BškRS96)
- sösö** ‘unleavened’ ◇ **söpsösö** (Juldašev 1981: 196)
- sybar** ‘motley’ ◇ **sypsybar** (Juldašev 1981: 196, BškRS96)

⁸ The original meanings in BškRS58 are: *кызыл* ‘красный, румяный’ : *кып-кызыл* ‘ярко-красный, багряный, совершенно красный’.

- šyjk** ‘1. elastic, flexible; 2. thin; with smooth hair *of an animal*’ ◊ **šypšyjk** ‘liquid, thin’ (BškRS96)
- šym** ‘quiet’ ◊ **šypšym** (BškRS96)
See 2.3.3.
- šyma** ‘smooth’ ◊ **šypšyma** (BškRS58)
- taða** ‘clean’ ◊ **taptaða** (BškRS58, BškRS96)
- takyr** i.a. ‘smooth, groomed, footworn’ ◊ **taptakyr** ‘very smooth, very even’ (Ahmerov 1958: 766, BškRS58, BškRS96)
- täläšäk** ‘low’ ◊ **täptäläšäk** (BškRS96)
- taman** ‘exactly, just’ ◊ **taptaman** (BškRS58, BškRS96)
- tämheð** ‘unpalatable’ ◊ **täptämheð** (BškRS96)
- tängäl** i.a. ‘straight’ ◊ **täptängäl** (BškRS58)
- täpäš** ‘low’ ◊ **täptäpäš** (Ahmerov 1958: 766)
See 3.1.6 on the closer being identical to C_2 .
- täpäšäk** ‘low’ ◊ **täptäpäšäk** (BškRS58)
See 3.1.6 on the closer being identical to C_2 .
- tärän** ‘deep’ ◊ **täptärän** (BškRS58)
- tar** ‘tight, narrow’ ◊ **taptar** (BškRS58, BškRS96)
- tekä** ‘steep’ ◊ **teptekä** (BškRS58, BškRS96)
- tewäl** ‘exactly, precisely’ ◊ **teptewäl** (BškRS58, BškRS96)
- tið** ‘quickly, soon’ ◊ **tiptið** (BškRS96)
- tigeð** ‘smooth’ ◊ **tiptigeð** (BškRS58, BškRS96)
- tiŋ** ‘equal, similar’ ◊ **tiptiŋ** (BškRS58, BškRS96: only attested in \sim *bulyw* ‘to be exactly the same’)
- töð** ‘straight, direct’ ◊ **töptöð** (BškRS96)
- tokon** ‘*dial.* lack of exit, holes, runoff’ ◊ **toptokon** (BškRS58)
- tokor** i.a. ‘short’ ◊ **toptokor** (BškRS58)
- tölöš** ‘short, undersized’ ◊ **töptölöš** ‘very low’ (BškRS96)
- tonok** ‘settled, still *of water*’ ◊ **toptonok** ‘completely still (water)’ (BškRS58)
- tübän** ‘low’ ◊ **tuptübän** (BškRS58)
- tuly** ‘full’ ◊ **tuptuly** (BškRS58, BškRS96)
- tumalak** ‘sphere, spherical, round’ ◊ **tuptumalak** ‘completely round, completely spherical’ (BškRS58, BškRS96)
- tünäräk** ‘round’ ◊ **tüptünäräk** (BškRS58)
- tuŋ** ‘frozen’ ◊ **tuptuŋ** (BškRS58)
- tura** ‘straight, direct, even’ ◊ **tuptura** (BškRS58)
- tütä** ‘straight, direct *of road*’ ◊ **tüptütä** (BškRS58)
- tyyyð** ‘tight, narrow’ ◊ **typtyyyð** (BškRS58, BškRS96)
- tymyk** ‘quiet’ ◊ **typtymyk** (BškRS58, BškRS96)
See 3.1.11 on families of reduplications.

tyn ‘quiet’ ◊ **typtyn** (BškRS58)

See 3.1.11 on families of reduplications.

tynys ‘quiet, peaceful’ ◊ **typtynys** (BškRS58, BškRS96)

See 3.1.11 on families of reduplications.

zängär ‘(light) blue’ ◊ **zäpzängär** (BškRS58, BškRS96)

2.3.3 Special cases

appaŷym ‘my dear, my darling’ (BškRS96) ◊ **ak** ‘1. white; 2. *i.a.* good, kind’

The meaning of *ak* is made up of a considerable number of components, including ‘white’ in the first place and ‘good, kind’ in the fifth (BškRS96). If we consider *p* and *pp* to be different closers, then *ak* is one of two words in Bashkir which have more than one closer possible (the other is *jäšel* ‘(light) green’). It would seem that both *p* and *pp* are bound to only a single component of the complex meaning of the base: *p* intensifies ‘white’, while *pp* intensifies ‘good, kind’. The latter is only attested with a Px. See 3.1.2 for similar examples in other languages.

öpörjanġy (Juldašev 1981: 196) ◊ **ör-janġy** ‘brand-new’

This word is special in that its base meaning is already an intensification and probably a binom. Apparently, it must have after a time become trite to a certain degree, which made it possible for a reduplication to be formed. See 3.1.13 for more examples of such erosion.

The final element (*janġy*) is clear but the initial *ör-* is less so. Perhaps, it is related to Yak. *ür.üŋ* ‘white’, and the original meaning of the whole was *‘white-new’ (= ‘brand-new’). The same composition is also present in Tatar, see *ör-jaŋa* in 2.15.1.

Note that *janġy* alone can also be reduplicated, but **ör* does not seem to exist as a separate word. Juldašev 1981: 196 lists *ör-janġy* among reduplications, on par with *öpörjanġy*.

šypšym (BškRS96) ◊ **šym** ‘quiet’

The reduplication *šym* → *šypšym* is clear, but the base *šym* is much less so. Quite probably, it is related to Uigh. *žim* ‘quietly, calmly; silently, tacitly’, attested in particular in *žimžit* ‘completely silent, completely mute’. UjgRS explains *žit* as “парное к **جم** [жим]”, but this interpretation has a weak point, see 2.19.3.

The Uighur word may also be a reduplicated anlaut of **čymčyrt* ‘complete silence’, which had been severed and made an independent word; see 3.1.10 for similar examples. In such case, the Bashkir reduplication *šypšym* would most likely be a peculiar case of double reduplication; see 3.1.9 on those.

Possibly, *šym* can also be related to two unclear intensifiers with indefinite semantics: Trkm. *čim* and al-Kāšġarī’s *čim* [?] *čym*; see 3.1.10.

2.3.4 Structure

Three closing consonants of *C*-type are attested in a total of 93 examples derived from 91 unique bases, in a rather skewed distribution:

m: 2 examples: *jäšel* and *kük*,

p: 90 examples: *äd, ak, äkren, aryw, äse, asyk, bäläkäj, bojok, bökrö, bötön, buš, byrsak, haj, halkyn, hary, haw, hiräk, horo, jakšy, jakty, jakyn, jaltyr, jangy, janγγδ, jäš, jäšel, jejel, jeüeš, joka, jomro, jomšak, jylytyr, jyly, kak, kara, karanyγy, katy, körän, koro, kujy, kyθka, kyzyl, matur, nädek, nakyθ, nasar, neskä, ör-jangy, šaktaj, salyš, šärä, sej, serek, söso, sybar, šyjk, šym, šyma, taδa, takyr, täläšäk, taman, tämhed, tängäl, täpäš, täpäšäk, tar, tärän, tekä, tewäl, tiδ, tigeδ, tiη, töδ, tokon, tokor, tölöš, tonok, tübän, tuly, tumalak, tuη, tünjäräk, tura, tütä, tyγγδ, tymyk, tyn, tynys, zängär*, and

pp: 1 example: *ak*.

In only one case (*kük*) is *p* not a possible closing consonant. Two words have more than one closer possible: *ak* (*p* and *pp*) and *jäšel* (*m* and *p*). That is, all words except for *kük* have *p*, and together with it, two words also have other closers possible. Noteworthy is the semantic differentiation of the two reduplications of *ak* (see 2.3.3 above), and the forms *täptäpäš* and *täptäpäšäk* (see 3.1.6 on the closer being identical to *C*₂).

2.3.5 Semantics

The meanings of all reduplications are, unsurprisingly, the same as the base meanings or their simple intensifications. The only diversions are provided by *ak*, where the meaning can be reduplicated in two ways, and *ör-jangy*, where already the base is an intensification (see 2.3.3 above).

With, again, two exceptions, all words are fairly definite adjectives, to the degree allowed for by the general build of the Turkic languages. It is only *tumalak* that has – together with a purely adjectival part – a more distinct substantival component in its meaning (‘sphere’ apart from ‘spherical, round’), and *tokon* that is a pure noun (dialectally).

2.4 Dolgan

Unlike in the neighbouring Yakut, *C*-type reduplication does not appear to be the favourite method of word formation in Dolgan. Very few, and barely diversified examples are available, a fact that cannot be easily blamed on the shortage of lexicographic descriptions. Also, hardly any word in this tiny set appears to be calling for greater attention.

The large disproportion between the Dolgan and Yakut collections is most likely due to the Tungusic substrate in the former. It is not entirely clear how reduplications appeared in the Tungusic languages, but there is little doubt that the method never gained much popularity in them, see 3.4.3. Dolgan and Yakut are known to have diverged not later than in the beginning of the 17th century (Stachowski M. 1996), but there are no old attestations with which to determine whether it was the Yakut collection that expanded very rapidly after that date, or the Dolgan one that shrank. In theory, both processes could also occur simultaneously. It seems that only guesses can be made until a detailed investigation of the Mongolic and Tungusic reduplications is undertaken.

2.4.1 Sources

The main sources of the material are DW and DWS containing ca. 8900 entries in total. Some attestations have also been found in Pomorska 2004: 143f, and in the two grammatical works listed below. Altogether, only eight examples are available. As it was mentioned above, this low number is not merely due to incompleteness of the lexicographic material as even in larger collections reduplications are scarce. Li 2011: 97 states directly that in his materials, there were no intensified adjectives whatsoever.

For grammatical descriptions, Ubrjatova 1985b: 131 and Stachowski M. 1997: 89 have been used.

Both are quite laconic, and effectively limited to a list of examples, and the statement that the reduplicated syllable can be closed by *p* or *bys*.

Transcription:

DW, DWS, Pomorska 2004: 143f: $\ddot{a} \rightarrow e \mid x \rightarrow \chi$,

Ubrjatova 1985b: $\# \rightarrow \eta \mid \gamma \rightarrow \ddot{u}$.

2.4.2 Standard cases

kara ‘black’ \diamond **kapkara** (Ubrjatova 1985b: 131, DW, Pomorska 2004: 144)

karana ‘dark’ \diamond **kapkarana** (DW)

küök ‘blue’ \diamond **küpküök** (DW)

kyhyl ‘red’ \diamond **kypkyhyl** ‘scarlet, purple, intensive red’ (Stachowski M. 1997: 89, DWS, Pomorska 2004: 144)

kyra ‘small’ \diamond **kypkyra** (Ubrjatova 1985b: 131, DW, Stachowski M. 1997: 89)

kytarkaj ‘red’ \diamond **kypkytarkaj** ‘pretty as a picture’ (DW, Stachowski M. 1997: 89)

tögürük ‘round’ \diamond **töptögürük** (Stachowski M. 1997: 89, DWS, Pomorska 2004: 144)

ürün ‘white’ \diamond **üprün** (Ubrjatova 1985b: 131, DW, Stachowski M. 1997: 89)

2.4.3 Special cases

suotčogotok ‘completely alone’ (Stachowski M. 1997: 89) \diamond **č|h|sogotok** ‘alone’

In Stachowski M. 1997: 89, this word is considered a special case because of its unclear phonetics, and discussed separately. It is proposed that it might be a composition rather than a reduplication, and that eventually it could be traced back to **jalgōz-jalgōz.ak*.

Despite its scantiness which rules out any categorical statements, the Dolgan material rather noticeably only features *p* and *bys* as the closing segments – but with the exception of *če.bit.čēlkē* ‘snow-white’. In theory, it is possible that *suotčogotok* is another exceptional form. Nevertheless, the long vowel in the reduplicated anlaut (-*uo-*) is suspicious. There is only one Dolgan example with a long vowel in the head. It is *küök* ‘blue’, and it has its vowel shortened in the reduplication – which would be unlike the hypothetical reduplication of *suotčogotok*.

The word is also present in Yakut as *soyotoχ* (see 2.21.2). Its reduplications are *so.bus-*, *so.s-* and *suo.s-*, and an array of seemingly irregular forms akin to Dolg. *suotčogotok*, see 2.21.3. Apparently, *t* is not allowed as a closing consonant in Yakut at all. The reduplication of Yak. *küöχ* ‘blue’ has its vowel shortened, too, but in other words the vowel might (rarely) preserve – or apparently even acquire – length, see e.g. Yak. *būsbüün* ‘absolutely all’, *čuoṗčuoṗur* ‘very motley’, *kiebiskiej* ‘very wide’, and a more general discussion of shortening in 3.1.20.

One more factor that has to be taken into account here, is the Siberia-wide alternation of *s* ~ *t*. An influence of this phenomenon can also be suspected in the already mentioned *čebitčēlkē* ‘snow-white’. However, the mechanism and scope of this alternation remain for now undeciphered, and an attempt to use it to explain *suotčogotok* can be no more than a speculation.

Neither any of the facts above, and nor their sum, are sufficient to determine with certitude what the nature of *suotčogotok* is. In light of the Yakut forms with long vowels in the reduplicated anlauts, that would be difficult to be all explained by an ancient composition, it is perhaps the possibility of secondary emphatic lengthening that should be viewed as the most likely one.

See also 3.1.11 for other big families of related reduplications.

2.4.4 Structure

Only one closing consonant of *C*-type is attested in a total of eight examples derived from eight unique bases:

p: 8 examples: *karaṇa*, *küök*, *kyhyl*, *kyra*, *kytarkaj*, *kara*, *tögürük* and *ürün*.

One word has a long vowel or a diphthong, *küök* ‘blue’, and it has been shortened in the reduplication. See 3.1.20 on shortening of the reduplicated vowel.

2.4.5 Semantics

From the point of view of semantics, the only worthwhile case is that of *kytarkaj* ‘red’ → *kypkytarkaj* ‘pretty as a picture’. Stachowski M. 1997: 89 sees here an influence of Russ. *красный* ‘1. red; 2. arch. beautiful’, *красивый* ‘beautiful’. Interestingly, the same development did not happen in the case of the synonymous *kyhyl* ‘red’ which only reduplicates to a more predictable ‘scarlet, purple, intensive red’.

2.5 Gagauz

Compared to its most closely related languages, Gagauz *C*-type reduplications do not appear to be particularly numerous. It is not clear, however, whether it is an accurate picture of the state in Gagauz, or just a reflection of the scarcity of available data. As far as diversification is concerned, Gagauz is visibly less rich than Azeri or Turkish, but it still certainly counts among the most diversified ones.

In 43 examples, four closing consonants are attested (*m*, *p*, *r* and *s*), and possibly this number should be increased by *t* in *jat-jaban* and *z* in *bezbelli* and *dozdolaj*. See 2.5.4 and 2.5.3.

Another structural feature to be noted is the consistent shortening of the reduplicated vowel, be it primarily or secondarily long. See 2.5.4.

Semantically, Gagauz reduplications are quite standard. Perhaps only the case of *düz* might be more interesting, provided that it is not merely inadequate translation that makes it appear so. See 2.5.5.

Although beyond the primary scope of this work, the Gagauz word *je.piz.jeni* ‘brand new’ should be mentioned as the only example that I am aware of of a *CVC*-type reduplication outside of the North Siberian group (see also 3.4.4 on the origin of reduplication).

2.5.1 Sources

The main source of the Gagauz material is GagTS which contains ca. 10 800 entries. Some attestations have also been found Ščerbak 1977: 120 and in the grammars listed below.

For grammatical descriptions, Pokrovskaja 1964: 106 and Özkan 1996: 197f have been used.

Pokrovskaja 1964: 106 limits herself to some examples and the statement that the reduplicated syllable is closed by *p* or *s* and, more rarely, *m*, *r* or *t*. For *t*, only one example is given, namely *jat-jaban* ‘1. wild, savage; 2. foreign, alien’. This is most probably a mistake; see 2.5.3 below.

Özkan 1996: 197f is more laconic but he, too, gives some examples and a list of possible closers: *m*, *p*, *r*, *s* and *z*. The latter two appear to be merely phonetic variants but see 2.5.4 below. The form *jat-jaban* or *t* as a closing consonant are not mentioned.

Transcription:

GagTS, Özkan 1996: 197f: $\zeta \rightarrow \check{c}$ | $e- \rightarrow je$ ⁹ | $\iota \rightarrow y$ | $ia \rightarrow ija$ ¹⁰ | $\varsigma \rightarrow \check{s}$ | $VV \rightarrow \bar{V}$ | $y \rightarrow j$,
 Pokrovskaja 1964: $\ddot{y} \rightarrow \ddot{u}$ | $\varkappa \rightarrow \check{z}$ | $VV \rightarrow \bar{V}$,
 Ščerbak 1977: $\ddot{i} \rightarrow y$.

2.5.2 Standard cases

ačyk ‘open’ ◇ **apačyk** (GagTS)

ajdynnyk ‘light, glitter’ ◇ **apajdynnyk** (Özkan 1996: 197)

ak ‘1. white; 2. clean; 3. white of the eye; 4. white speck in the eye’ (GagTS) ◇ **apak** ‘completely white’ (Pokrovskaja 1964: 106, Özkan 1996: 197)

According to Pokrovskaja 1964: 106, *ak* is not used independently. It is, however, attested in GagTS with the meanings as listed here. See also 2.5.3 below.

ansyz ‘sudden(ly)’ ◇ **apansyz** (Özkan 1996: 197)

ansyzdan ‘sudden(ly)’ ◇ **apansyzdan** (GagTS)

ansyzyn ‘sudden(ly)’ ◇ **apansyzyn** (GagTS)

belli ‘clear, obvious’ ◇ **bezbelli** (GagTS, Özkan 1996: 198)

bijaz ‘white’ ◇ **bimbijaz** (Pokrovskaja 1964: 106, Ščerbak 1977: 120, GagTS)

čevre *‘surrounding, around’ ◇ ***česčevre** ‘all around’ only in *česčevreye* (no meaning given; Özkan 1996: 198)

In GagTS, *čevre* is only attested with the meaning ‘kenarları işlenmiş mendil’, which would have made a semantically very unlikely base for reduplication. Possibly, the word is almost congruent with its Turkish equivalent *čevre* ‘surrounding, around, &c.’, also ‘handkerchief, headscarf’, only not attested as such in GagTS. See 2.5.3 below.

čyplak ‘naked’ ◇ **čyrčyplak** (Pokrovskaja 1964: 106, GagTS, Özkan 1996: 198)

See *jat-jaban* in 2.5.3 below.

diri ‘alive, lively’ ◇ **dipdiri** (GagTS)

dolaj ‘vicinity, surrounding, around’ ◇ **dozdolaj** ‘all around’ (GagTS, Özkan 1996: 198)

See *dozdolajanyndan* in 2.5.3 below.

dolu ‘full’ ◇ **dopdolu** (GagTS, Özkan 1996: 197)

dōru ‘1. straight, direct; 2. true, faithful, accurate’ ◇ **dopdōru** (GagTS, Pokrovskaja 1964: 106) ◇ **dosdōru** (Pokrovskaja 1964: 106)

duruk ‘clear, limpid’ ◇ **dupduruk** (GagTS)

⁹ In *jeni*, *ješil*, and their reduplications.

¹⁰ In *bijaz*, after «бийаз» in Pokrovskaja 1964: 106.

düz ‘smooth, even, straight’ ◇ **dümdüz** (Pokrovskaja 1964: 106) ‘1. *intens.* (Pokrovskaja 1964: 106); 2. openly, frankly (Pokrovskaja 1964: 106, GagTS)’ ◇ **düpüz** ‘openly, frankly’ (GagTS)

See 2.5.5 below.

jalabyk ‘brilliant, sparkling’ ◇ **japjalabyk** (Özkan 1996: 197)

jaš *‘wet’ ◇ **jamjaš** (no meaning given; Özkan 1996: 198)

In GagTS, *jaš* is only attested with the meanings ‘1. tear(drop); 2. age’, which would have made a semantically very unlikely base for reduplication. Possibly, the word is almost parallel to its Turkish equivalent *jaš* ‘1. age; 2. wet; 3. tear(drop)’, only not attested in full in GagTS. See 2.5.3 below.

jeni ‘new’ ◇ **jepjeni** (GagTS)

ješil ‘green’ ◇ **jemješil** (GagTS)

kajyl ‘consentient’ ◇ **kamkajyl** (Özkan 1996: 198)

kara ‘black’ ◇ **kapkara** (GagTS, Özkan 1996: 197)

katy ‘strong, sturdy’ ◇ **kaskaty** ‘1. *intens.*; 2. caked, solidified, inflexible, numb’ (Pokrovskaja 1964: 106, Ščerbak 1977: 120, Özkan 1996: 198)

kirli ‘dirty’ ◇ **kipkirli** (GagTS)

koža ‘1. old; 2. a married man; 3. (well-)known’ ◇ **koskoža** (Pokrovskaja 1964: 106: ‘completely grown up’, Özkan 1996: 198: no meaning given)

kožamiti ‘1. old; 2. (well-)known’ ◇ **koskožamiti** ‘very big, huge’ (GagTS)

kuru ‘dry’ ◇ **kupkuru** (Pokrovskaja 1964: 106, GagTS)

kyrmyzy ‘red’ ◇ **kypkyrmyzy** (Pokrovskaja 1964: 106, GagTS)

kyzgyn ‘angry, furious, passionate’ ◇ **kypkzygyn** (Özkan 1996: 197)

māvi ‘blue’ ◇ **masmāvi** (Pokrovskaja 1964: 106, GagTS)

mor ‘purple, violet’ ◇ **mosmor** (Pokrovskaja 1964: 106, GagTS)

pak ‘clean, clear’ ◇ **pampak** (Özkan 1996: 198)

sā ‘healthy’ ◇ **sapsā** (GagTS)

sary ‘yellow’ ◇ **sapsary** (Pokrovskaja 1964: 106, GagTS, Özkan 1996: 197)

sert ‘hard, harsh, rough’ ◇ **sepsert** (GagTS)

silme *‘deletion’ ◇ **sipsilme** (GagTS: no meaning given)

GagTS does not attest **silme*, but it does attest *silmē* with the meaning ‘to delete, wipe, erase, cancel’. The reduplication is unlikely to be a loanword because a similar form does not seem to be attested for Ottoman, Turkish or Azeri. See 2.5.3 below.

sökük ‘rip’ ◇ **söpsökük** (GagTS)

sūk ‘cold’ ◇ **supsūk** (GagTS)

tülü ‘hairy, feathery’ ◇ **tüptülü** (Pokrovskaja 1964: 106)

ülen ‘(at) noon’ ◇ **üpülen** ‘meridian, daily’ (Pokrovskaja 1964: 106)

uzun ‘long, tall’ ◇ **upuzun** (GagTS)

2.5.3 Special cases

In seven cases, GagTS failed to attest a base with a meaning suitable for reduplication. For four (**ač*, **olmuš*, **soluk* and **syky* below), I could not find any additional support, and so they have been classified as special cases and will be excluded from further considerations. For the other three (*čevre*, *jaš* and *silme* in 2.5.2 above), it was only a specific meaning that was missing, or the exact morphological form. These will be treated on par with the entirely clear forms.

apač (Özkan 1996: 197: no meaning given)

It seems unlikely that **ač* ‘hungry’ should be missing from Gagauz. A borrowing is no more plausible as **apač* does not appear to be attested for Ottoman, Turkish or Azeri. However, it is missing from GagTS and, be it for methodological consistency alone, it will be considered a special case here. See the comment at the beginning of 2.5.3 above.

dozdolajanyndan (Özkan 1996: 198: no meaning given)

This form is not clear. Perhaps a misprint for **dozdolajyndan* ‘from all around’?

jat-jaban (Pokrovskaja 1964: 106, Pomorska 2004: 144) ◊ **jaban** ‘wild, savage, alien’

This word is not clear. Eventually, it can be a reduplication of *jaban*, or a composition of **jat* ‘foreign’ + *jaban* id., and either could have possibly happened outside of Gagauz.

If it is a reduplication, then it is a highly unusual one, and the only example with *t* in the role of the closing consonant. Note that also *r* can only be found in one example in this function (*čyrčyplak* ‘stark naked’, see 2.5.2 above), but unlike *t*, *r* is well rooted in Turkish, Azeri and others. Apart from *jat-jaban*, *t* appears in one Azeri reduplication (*gatgaryšyg* ‘entanglement, confusion, ...’, see 2.2.2 above) and in one Crimean Tatar (*četčešit* ‘completely different; very diversified’, see Jankowski 1992: 129). These provide a modest support.

On the other hand, if it is a composition, then it can barely be attributed to Gagauz as the word **jat* does not seem to be attested in it. Note that this does not exclude the possibility of inheritance.

It appears, then, that the available options are: 1. Gagauz or earlier reduplication, which seems unlikely; 2. Gagauz or earlier composition, which seems more likely; and 3. borrowing from Turkish, Karaim, Urum or perhaps yet another (see 2.6.3 and Stachowski K. 2010: 153), which seems equally likely.

The presently available data do not seem sufficient to allow to continue these considerations.

This example might, however, prove valuable for more distant musings, as in fact, it could be also interpreted as a composition which is at the same time a reduplication in statu nascendi; see 3.4.4 on the origin of reduplication.

opolmuštular (Özkan 1996: 197: no meaning given)

This reduplication is most probably native Gagauz because a corresponding one does not seem to be attested for Turkish or Azeri. Nevertheless, the base **olmuš* (probably ‘ripe’) is missing from GagTS, and so the word will be considered a special case here. See the comment at the beginning of 2.5.3 above.

sopsoluk (GagTS)

GagTS defines *sop* as ‘intensifying prefix’ and gives *sopsoluk* as an example. Nevertheless, the base **soluk* (probably ‘pale’) does not seem to be attested, and the word will be considered a special case here. See the comment at the beginning of 2.5.3 above.

sypsyky ‘very tight, very narrow’ (GagTS)

GagTS defines *syp* as ‘word-intensifying prefix’ and gives *sypsyky* as an example. Nevertheless, the base **syky* does not seem to be attested, and the word will be considered a special case here. See the comment at the beginning of 2.5.3 above. Perhaps a borrowing from Turkish (see 2.16.2)?

2.5.4 Structure

Five closing consonants of *C*-type are attested in a total of 43 examples derived from 41 unique bases, in a quite uneven distribution:

m: 6 examples: *bijaz*, *düz*, *jaš*, *ješil*, *kajyl* and *pak*,

p: 27 examples: *ačyk*, *ajdynnyk*, *ak*, *ansyz*, *ansyzdan*, *ansyzyn*, *diri*, *dolu*, *dōru*, *duruk*, *düz*, *jalabyk*, *jeni*, *kara*, *kirli*, *kuru*, *kyrmyzy*, *kyzgyn*, *sā*, *sary*, *sert*, *silme*, *sökük*, *sūk*, *tūlü*, *ūlen*, and *uzun*,

r: 1 example: *čyplak*, and

s|z: 9 examples: *belli*, **čevre*, *dolaj*, *dōru*, *katy*, *koža*, *kožamiti*, *māvi* and *mor*.

The status of *z* is not entirely clear. It appears to be but a phonetic variant of *s*, presumably before voiced non-liquid consonants. However, three words provide such environment (*belli*, *dolaj* and *dōru*), and it is only in the first two that *z* is attested. Also the expected **b* as a variant of *p* is missing. Despite these doubts, *z* will not be considered an independent closing consonant here.

Two words have more than one closer possible: *düz* and *dōru* (*m* and *p*, and *p* and *s*, respectively). The only significant regularity is that both have *p* and some other consonant as options.

In six words, the first vowel of the stem is long (*dōru*, *māvi*, *sā*, *sūk*, *tūlu* and *ūlen*). In all these cases, the reduplicated vowel is short. Note, however, that it is only in *māvi*, that the length is original (≪ Ar. مَائِي *māʾī* ‘water(y), aquatic’, see Nişanyan ÇTES for Tksh. *māvi* ‘blue’); in the remaining four words, the reduplication can, with equal probability, be assumed to have taken place before or after of the dropping of *g/j* and

the subsequent lengthening of the preceding vowel. See also 3.1.20 on shortening of the reduplicated vowel.

Two groups of words are derivatives: *ansyz*, *ansyzdan*, *ansyzyn*, and *koža*, *kožamiti*. In all cases the closing consonant is the same: *p* in the first group, and *s* in the second. It does not seem possible to determine whether derivation preceded or followed reduplication.

2.5.5 Semantics

In the great majority of cases, the reduplicated meaning is a simple intensification or apparently the same as the base meaning. Most cases are also clear adjectives, with the degree of conventionality necessary with the Turkic languages.

In two cases, *düz* and *ülen*, a modest evolution can be observed in the reduplicated meaning; see also below. In two other, *koža* and *kožamiti*, the reduplicated meaning suggests that perhaps not the full spectrum of the base meaning has been attested in the source. Such occasions must, unfortunately, be expected when only a limited amount of data is available. Finally, in two more, *čevre* and *jaš*, the reduplicated meaning clearly could not have arisen from what meaning is attested for the base. Again, the scarcity of the data is probably to blame; see 2.5.3 above.

The meanings of *düz* and its reduplications are perhaps of greater interest: ‘smooth, even, straight’ → *dümdüz* ‘1. intens.; 2. openly, frankly’, and *düpdüz* ‘openly, frankly’. It seems as if the reduplication with *m* intensified the meaning both literally and figuratively, while the one with *p* only figuratively. This would have been similar to Bshk. *ak* i.a. ‘1. white; 2. good, kind’ → *apak* ‘snow-white’, and **appak* in *appayym* ‘my dear, my darling’. However, until a considerably more exhaustive dictionary of Gagauz is compiled, this is a too far-fetched conclusion. See 3.1.2 on other examples of reduplications with different closing consonants and different meanings.

2.6 Karaim

Judging from their number, *C*-type reduplications seem to be hardly popular in the western Karaim languages, but much more so in Eastern Karaim. There is also an unusually large number of special cases in the latter, most of them loanword suspects. This harmonizes very well with the areas of influence of Ottoman (rich in reduplications) and of non-Turkic languages (which do not know the method at all) – suggesting that the actual native Karaim stock of reduplications is in fact much smaller. Unfortunately, separating native formations from old loanwords appears to be generally impossible at present, and so all reduplications for which the base is independently attested, are treated here as native.

In the western Karaim languages, only *m*, *p* and *pp* are attested as closing consonants; Eastern Karaim additionally has *r* and *s*. See 2.6.4.

Examples which are not clearly adjectival in character are very few. However, in Eastern Karaim, one reduplicated verb is attested, which is a very rare occurrence. See 2.6.5, and also 3.1.22 on reduplication of verbs.

Perhaps most noteworthy among the special cases are *appačyk* ‘snow-white’ because of its double *pp*, *bomboš* ‘completely empty’ because of its geography, and *čöpčövre* ‘all around’ because of its phonetics. See 2.6.3.

2.6.1 Sources

The Karaim material had been collected in Stachowski K. 2010. Here, it is mostly only summarized for convenience, but certain words have been reclassified and certain parts modified and extended; see especially 2.6.3 below (the special cases).

The main sources are KarRPS, RKarS-Haf and RKarS-Lev, containing in total ca. 17 400 + 7300 + 8120 = 32 820 entries. Some attestations have also been found in Berta 1998b, Grønbech 1942, Józefowicz 2008, Kakuk 1991, Kowalski 1929, Mardkowicz 1935, Sulimowicz 1973, and in the grammars listed below.

For grammatical descriptions, Zajączkowski 1931 (Kar.SW), Musaev 1964: 183 (Kar. NW and SW), Prik 1976: 85 (Kar.E) and Musaev 1977: 36 (Kar.NW and SW) have been used.

All are similar. They point to *p* as the most common closing consonant, some also mention the exceptionality of *m* in *jemješil* ‘very green’, and only Prik 1976: 85 enumerates all four possible closers, *m*, *p*, *r* and *s*. Neither attempts to formulate a rule of distribution.

Notation is quite inconsistent in some of the sources. I will retain here the unified phonological transcription used in Stachowski K. 2010 with only one modification, i.e. with the ⟨x⟩ for the guttural voiceless fricative being here replaced with ⟨h⟩.

2.6.2 Standard cases

The data from all the three Karaim languages are all treated together here. The affiliation is only marked after the \diamond symbol, as *E*, *NW* or *SW* for ‘eastern (= Crimean)’, ‘northwestern (= Trakai)’, and ‘southwestern (= Lutsk/Halych)’, respectively.

ačyk ‘open’ \diamond E **apačyk** (KarRPS, RKarS-Haf s.v. *открытый настез* and *распахнуть*, RKarS-Lev); \diamond E **appačyk** (KarRPS, RKarS-Lev)

See 2.6.4 below.

ačmak ‘to open’ \diamond E **apačmak** ‘to break open, to throw open’ (RKarS-Haf)

See 2.6.5 below.

ah see *ak*

ak &c. ‘white’ ◇ E **apak** (KarRPS, RKarS-Lev) ~ NW **apah** (Józefowicz 2008 s.v. *bialuteńki* and *bieluteńki*) ~ SW **apak** (Zajączkowski 1931, Mardkowicz 1935, Musaev 1964: 183, KarRPS) ◇ E **appak** (RKarS-Haf s.v. *белоснежный*) ~ SW **appak** (KarRPS)

al ‘red’ ◇ E **apal** (KarRPS, RKarS-Lev)

ansyz ‘sudden(ly), unawares’ ◇ E **apansyz** (KarRPS, RKarS-Haf, RKarS-Lev)

See 2.6.4 below.

ansyzyn ‘sudden(ly), unawares’ ◇ E **apansyzyn** (RKarS-Lev)

See 2.6.4 below.

aryh see *aryk*

aryk ‘thin, lean’ ◇ E **aparyk** (KarRPS, RKarS-Lev) ~ NW **aparyh** (Józefowicz 2008)

baška ‘other’ ◇ E **bambaška** (KarRPS, RKarS-Lev)

bedava ‘free of charge’ ◇ E **besbedava** (KarRPS, RKarS-Haf s.v. *совершенно даром*, RKarS-Lev)

belli ‘clear, obvious’ ◇ E **besbelli** ‘probably’ (KarRPS, RKarS-Lev)

beter ‘worse’ ◇ E **besbeter** (KarRPS, RKarS-Lev)

bijaz ‘white’ ◇ E **bimbijaz** (KarRPS, RKarS-Haf s.v. *белоснежный*, RKarS-Lev)

bos see *boš*

boš &c. ‘empty’ ◇ E **bomboš** (KarRPS, Prik 1976, Musaev 1977: 7, RKarS-Haf s.v. *совершенно пустой*) ~ NW **bomboš** (KarRPS, Józefowicz 2008) ◇ NW **bopboš** (Musaev 1964: 183; Musaev 1977: 7) ~ SW **bopbos** (Musaev 1977: 7)

See *bomboš* in 2.6.3 below.

bošyna ‘in vain’ ◇ E **bombošyna** (RKarS-Haf, RKarS-Lev)

See 2.6.4.

bütün ‘whole’ ◇ E **büsbütün** (KarRPS, RKarS-Lev)

čebik ‘quick(ly)’ ◇ E **čarčebik** (KarRPS, RKarS-Lev) ~ E **čerčebik** (KarRPS, RKarS-Haf, RKarS-Lev) ~ E **čyrčebik** (RKarS-Haf)

See *čarčebik* in 2.6.3 below.

čevre see *čövre*

ćivre see *čövre*

čövre &c. ‘around’ ◇ E **čepčevre** (KarRPS) ~ E **čöpčövre** (Grønbech 1942, KarRPS, RKarS-Lev) ~ E **čöpčüvre** (KarRPS, RKarS-Lev) ~ E **čüpčüvre** (KarRPS, RKarS-Haf) ~ SW **ćipćivore** (Mardkowicz 1935: 71, KarRPS)

See *čöpčövre* and *čöpčövetin* in 2.6.3 below.

čüvre see *čövre*

čyplak ‘naked’ ◇ E **čyrčyplak** (KarRPS, Prik 1976, RKarS-Haf s.v. *догола, нагишом* and *совершенно голый*, RKarS-Lev)

ińćke &c. ‘thin’ ◇ NW **iپیńćke** (Musaev 1964: 183) ~ NW **iپیńćka** (KarRPS, Musaev 1977: 36, Józefowicz 2008) ~ SW **iپیńćke** (Mardkowicz 1935, KarRPS)

jalynyz ‘lone’ ◊ E **japjalynyz** (RKarS-Lev)

jahši see *jahsy*

jahšy &c. ‘good’ ◊ NW **japjahši** (Musaev 1977: 36)¹¹ ~ NW **japjahšy** (Musaev 1964: 183, Berta 1998, Józefowicz 2008 s.v. *dobry-przedobry*) ~ SW **japjaksy** (Musaev 1977: 36)

jaksy see *jahšy*

ješil &c. ‘green’ ◊ E **jemješil** (Prik 1976, RKarS-Lev) ~ E **jymješly** (RKarS-Haf s.v. *совершенно зелёный*) ~ NW **jeńješil** (Józefowicz 2008 s.v. *zieloniutki*) ~ SW **jemješil** (Zajączkowski 1931)
See *čarčebik* in 2.6.3 below.

ješił see *ješil*

ješil see *ješil*

ješly see *ješil*

jumalak ‘round’ ◊ E **jumjumalak** (KRPS, RKarS-Lev)

See *jumjumarlak* in 2.6.3 below.

kara ‘black’ ◊ E **kapkara** (KarRPS, Prik 1976, RKarS-Haf s.v. *до черна, совершенно чёрный* and *чёрный-пречёрный*, RKarS-Lev) ~ NW **kapkara** (Kowalski 1929, Józefowicz 2008 s.v. *czarniusieńki*) ~ SW **kapkara** (Zajączkowski 1931)

karavlyk ‘darkness’ ◊ E **kapkaravlyk** ‘darkness, blackness’ (KarRPS, RKarS-Lev)

keńaťa see *kenete*

keńeťa see *kenete*

kenete &c. ‘sudden(ly)’ ◊ E **kepkenete** (Sulimowicz 1973, KarRPS, RKarS-Lev) ~ NW **kepkeńaťa** (Kowalski 1929, KarRPS, Józefowicz 2008) ~ **kepkeńeťa** (Kowalski 1929, KarRPS, KKS, Józefowicz 2008) ~ SW **kepkenete** (Sulimowicz 1973)

kenete see *kenete*

kök ‘blue’ ◊ E **kömök** (KarRPS, RKarS-Haf s.v. *темно-голубой*, RKarS-Lev)

kuru ‘dry’ ◊ E **kupkuru** (RKarS-Haf, RKarS-Lev)

kyzył ‘red’ ◊ NW **kypkzył** (Musaev 1964: 183, Józefowicz 2008 s.v. *czerwony-czerwonutki*) ~ SW **kypkzył** (Zajączkowski 1931, Musaev 1964: 183)

mavu see *mavy*

mavy &c. ‘blue’ ◊ E **masmavu** (KarRPS, RKarS-Lev) ~ E **masmavy** (Prik 1976, RKarS-Lev)

mor ‘violet’ ◊ E **mosmor** ‘dark violet’ (RKarS-Haf s.v. *темно-лиловый*, RKarS-Lev)

saglam ‘1. healthy; 2. whole’ ◊ E **sapsaglam** (KarRPS s.v. *сагълам*, RKarS-Haf s.v. *совершенно здоровый*, RKarS-Lev)

sary ‘yellow’ ◊ E **sapsary** (RKarS-Haf s.v. *совершенно желтый* and *желтый-прежелтый*) ~ NW **sapsary** (Musaev 1964: 183, KarRPS) ~ SW **sapsary** (Zajączkowski 1931, Musaev 1964: 183)

¹¹ The form with *-i* is most probably a spelling mistake under the influence of Russian orthography; on pp. 33, 35 &c., it is given as «яхшы».

slah see *sylak*

sylak &c. ‘wet’ ◊ E **symsylak** (KarRPS s.v. *кõт*, RKarS-Lev) ◊ E **sypslah** (RKarS-Haf s.v. *насквозь промокнуый*) ◊ E **syrsylak** (KarRPS s.v. *сылакъ*; RKarS-Lev)
See *sypslah* in 2.6.3 below.

tamam &c. ‘whole, wholly’ ◊ E **tastamam** (KarRPS s.v. *тамам*, RKarS-Lev) ~ E **testaman** ‘precisely, accurately’ (RKarS-Haf)¹²

taman see *tamam*

temiz ‘clean’ ◊ E **teptemiz** (RKarS-Haf) ◊ E **tertemiz** (KarRPS s.v. *тэр II* and *тэр-тэз миз*, RKarS-Haf s.v. *совершенно чистый*)

tez ‘quick(ly)’ ◊ E **teptez** (KarRPS s.v. *тэз*, RKarS-Lev)

tok ‘full, satiated’ ◊ E **tomtok** (RKarS-Lev) ◊ E **toptok** (KarRPS, RKarS-Lev)

tokal ‘blunt’ ◊ E **tostokal** (Kakuk 1991)

tolu &c. ‘full’ ◊ E **toptolu** (RKarS-Haf s.v. *битком набито* and *полным-полно*) ~ E **toptoly** (KarRPS, RKarS-Lev) ~ NW **toptolu** (Musaev 1964: 183, KarRPS, Musaev 1977: 36, Józefowicz 2008 s.v. *pełniutki* and *przepelniony*) ~ SW **toptolu** (Zajączkowski 1931, Mardkowicz 1935, Musaev 1964: 183, KarRPS, Musaev 1977: 36)

tołu see *tolu*

toly see *tolu*

tomalak ‘round’ ◊ E **tostomalak** (KarRPS, RKarS-Lev)

tögerek ‘round’ ◊ E **tömtögerek** (KarRPS, RKarS-Lev)

tüz ‘straight’ ◊ E **tümtüz** (KarRPS, Kakuk 1991 s.v. *tımqara*, RKarS-Haf s.v. *совершенно ровный*, RKarS-Lev) ◊ E **tüptüz** (RKarS-Lev)

tyk ‘vertical, upright’ ◊ E **tyntyk** (RKarS-Haf)

See *tyntyk* in 2.6.3 below.

uzun ‘long’ ◊ E **upuzun** (Prik 1976)

2.6.3 Special cases

Similarly to the subsection on standard cases above (2.6.2), data from all the three Karaim languages are all treated together here. The affiliation is marked inside parentheses, before the sources, as *E*, *NW* or *SW*.

afansyz ‘sudden(ly), unawares’ (*E*; KarRPS)

The base **ansyz* does not seem to be attested. The closer *f* is most probably just a spirantized *p* which, together with the archaic *ŋ*, suggests that the form must be a very old one. Borrowing seems less likely as there apparently is no obvious source for one. See Stachowski K. 2010: 141, and also *apansyzdan* below, and 3.1.21.

¹² The form is unclear. See Stachowski K. 2010: 146.

apansyzdan ‘sudden(ly), unawares’ (E; KarRPS, RKarS-Haf s.v. *от внезапности*, RKarS-Lev)

The base **ansyzdan* does not seem to be attested. Possibly then, the form is a derivative of *apansyz* ‘sudden(ly), unawares’, rather than a simple reduplication. It would seem that the relations between these words are as presented in fig. 2.2.

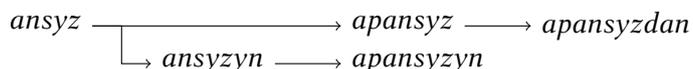


Figure 2.2: Reduplications of Kar.E *ansyz* ‘suddenly, unawares’ &c.

However, it needs to be borne in mind that this conclusion might well be a hasty one because in fact all it is based on is that **ansyzdan* is not attested in the most comprehensive dictionary of Karaim, KarRPS.

See also *ansyz* and *ansyzyn* in 2.6.2, and *afansyz* above.

bomboš (E; KarRPS, Prik 1976, Musaev 1977: 7, RKarS-Haf s.v. *совершенно пустой*) ~ **bomboš** (NW; KarRPS, Józefowicz 2008) ~ **bopboš** (NW; Musaev 1964: 183; Musaev 1977: 7) ~ **bopbos** (SW; Musaev 1977: 7) \diamond E, NW **boš** ~ SW **bos** ‘empty’

These words are interesting because of their geography.

The situation would have been clear if northwestern Karaim only had *p*, similarly to other Kipchak languages, eastern Karaim only had *m*, which would have been a rather likely influence of Ottoman, and southwestern Karaim had both, being in an intermediary position between the two.

However, the distribution in the two western languages is inverted: in the southern dialect only *p* seems to be attested, while in the northern one, there is both *p* and *m*.

It seems, then, that *m* in the reduplication of *boš* should be regarded as a common Karaim trait, which happened to be the weakest in southwest, stronger in the north, and the strongest in the east, where it was surely also reinforced by Ott. *bomboš* id. (see 2.13.2).

See 3.2.5 for further discussion of *boš*, and also 3.1.4 for other reduplications where the closing consonant is the same as the initial consonant of the base.

čarčebik (E; KarRPS, RKarS-Lev) ~ **čyrčebik** (E; RKarS-Haf) \diamond **čebik** ‘quick(ly)’
The form *čarčebik* might be a contamination with CTat. *čarčabik* id. (AiM) or Ott. *čarčabuk* id. (see 2.13).

The form *Čyrčebik* probably results from the eastern Karaim manner of pronouncing *e* higher and more backed in the first syllable (Prik 1976: 25f, Janowski 1997: 7f).

See Stachowski K. 2010: 142 and also *ješil* ‘green’ \rightarrow *jymješly* in 2.6.2 above, and 3.1.19 for other reduplications with anlaut not matching the base.

čöpčevirtin (E; KarRPS, RKarS-Lev) ~ **čöpčövretin** (E; KarRPS, RKarS-Lev) ~ **čüpčüvretin** ‘from all around’ (E; KarRPS, RKarS-Lev)

The bases *čevirtin, *čövretin and *čüvretin do not seem to be attested. The latter two appear to be ablatives of the attested čövre and čüvre; so does the first one, although is a less transparent way.

See čöpčövre below.

čöpčövre (E; Grønbech 1942, KarRPS, RKarS-Lev) ~ **čöpčüvre** (E; KarRPS, RKarS-Lev) ~ **čüpčüvre** (E; KarRPS, RKarS-Lev) ◊ **čevre** &c. ‘around’

These forms are not entirely clear. The oldest attested shapes are čövre, which exists no more, and čöpčövre (both in *Codex Cumanicus*). The present ones are: čevre, čivre, čüóra, and čüvre. The first and the last of them appear in reduplications and are not phonetically transparent. It seems that the most likely course of events was as follows:

1. †čövre → čöpčövre.

This must have happened before 1330, when the oldest surviving parts of *Codex Cumanicus* have been copied.

2. †čövre ≥ čevre, čüvre.

This can be explained by the general Turkic tendency to avoid ö in the first syllable.

3. čevre → čepčevre, čüvre → čüpčüvre.

After this stage, it would appear that the tendency to avoid ö must have significantly weakened, and the morphological transparency of reduplications disappeared, in order for čöpčüvre to arise in the next step.

4. čöpčövre ◊ čüpčüvre → čöpčüvre.

Especially the last conclusion seems somewhat grandiose for just a few vowel alternations in what is essentially one word. A different, and at least equally acceptable scheme, however, is not known to me.

See *čöpčevirtin above, and 3.1.19 for other reduplications with anlaut not matching the base.

čöpčö|üvretin see čöpčevirtin

čö|üpčüvre see čöpčövre

čymčyrt ‘complete silence’ (E; KarRPS, RKarS-Haf s.v. *полная тишина*, RKarS-Lev)

The base *čyrt does not seem to be attested, but the word is not an isolate. See 3.1.15 for what are probably its cognates.

jumjumarlak (E; KRPS, RKarS-Lev) ◊ **jumalak** ‘round’

The base *jumarlak does not seem to be attested. Most probably, it is the original shape of the word which had been later simplified to jumalak (→ jumjumalak, see 2.6.2 above), and eventually went entirely out of use.

A more interesting peculiarity of these forms is that they are closed by *m*, which is also the second consonant of the base. See 3.1.6 for other such examples.

kaskat ‘stupor, stupefaction’ (E; RKarS-Lev)

The origin of this word is unclear. It might be a native formation whose base has gone out of use, a borrowing from Ottoman (see Ott. *kaskaty* in 2.13.2), or other. It is present in two idioms: ~ *kalmak* ‘to be taken aback’ and ~ *külmek* ‘to laugh loudly’. See Stachowski K. 2010: 144.

komkos ‘very stupid’ (E; RKarS-Haf s.v. *совершенно глупый*)

This word resembles a reduplication both phonetically and semantically but the hypothetical base **kos* does not seem to exist in Karaim or any of the neighbouring languages.

koskoža ‘incredibly’ (E; RKarS-Haf), ‘1. incredible; 2. huge’ (RKarS-Lev)

This word is probably a borrowing from Ott. *koskoža* ‘huge’; see 2.13.2 and also Stachowski K. 2010: 144.

öpjuúš (NW; Józefowicz 2008) ◊ **juúš** ‘wet’

This word is not clear. The base is also present in other Turkic languages, where it has the form *jibi* and similar; see Zajączkowski 1932: 154. The labial vowels in *juúš* can be explained, so can be the lack of the expected *j-* in the reduplicated *öp-*, and the frontness of the initial vowel in it (see Stachowski K. 2010: 143), but the fact that it is an *ö* rather than an **ü*, remains mysterious. Explanation by a composition is not likely because **öp* does not seem to exist independently.

simsijah ‘completely black’ (E; KarRPS, RKarS-Lev) ~ **simsijak** ‘completely black’ (E; KarRPS, RKarS-Lev), ‘completely (dark) blue’ (E; RKarS-Haf s.v. *совершенно синий*)’

The base **sijah|k* does not seem to be attested in unanimously Karaim texts. It does appear multiple times in at least one anthology, its language however is mixed and under heavy influence of Turkish (Aqtay 2009: 33 and 706).

From the point of view of eastern Karaim as a whole, it is probably a borrowing from Ott. *simsijah* ‘completely black’ (see 2.13.2). The form with *-k* is likely the result of hypercorrectness. See Stachowski K. 2010: 145, and *sypslah* below.

simsijak see *simsijah*

sypslah ‘completely wet’ (E; RKarS-Haf s.v. *насквозь промокший*)

The base **slah* does not seem to be attested. It is most probably a slovenly pronunciation of *sylak*; on *-k* ~ *-h* see also *simsijah* above.

Regardless of the specific pronunciation, the base *sylak* is unusual in that its reduplications can be closed with as many as three different consonants. There is only one more such word in the present collection, Yak. *kyžy* ‘contrariness, ...’, see 2.21.3 and also 3.1.1 on alternative closing consonants.

The three forms are not entirely clear. The *m*-variant could possibly be seen as a Crimean Tatar influence (CTat. *symsylak* (AiM)), but that would be just moving the problem beyond the scope of the present work (Crimean Tatar is not discussed here), as all of the Kipchak languages typically close their reduplications with *p(p)* and no other consonant. The *p*-variant, then, may be a

native, Kipchak reduplication (see 3.2.1). Finally, the *r*-variant could be tied to a possible influence of Ott. and Tksh. *syrsyklam* ‘completely wet’ (see 2.13.2 and 2.16.2, respectively).

Although Urum is essentially not analysed in the present work, it should be noted that it, too, has three possible reduplications of *sylah* &c. ‘wet’: *symslah*, *sypsylah*, and *syrsylah* (Garkavecъ 2000), and the similarity to Eastern Karaim is unlikely to be accidental.

See also 3.1.19 for other reduplications with anlaut not matching the base.

syrsyklam ‘completely wet’ (E; KarRPS, RKarS-Lev)

The base **syklam* does not seem to be attested. Probably a borrowing from Ott. *syrsyklam* id. (see 2.13.2). See Stachowski K. 2010: 145.

tentek ‘very stupid, very sloven’ (E; KarRPS, RKarS-Haf s.v. *разгильдяй*, RKarS-Lev s.v. *нерадивый, разгильдяй* and *рассеянный*) ~ **teńtak** (NW; KarRPS, Józefowicz 2008) ~ **tentek** (SW; Mardkowicz 1935, KarRPS)

The word resembles a reduplication both phonetically and semantically but the hypothetical base **tek* does not seem to exist in Karaim or any of the neighbouring languages. See also Kirg. *tentek* ‘1. naught, imp usually of children; 2. fool, stupid’, e.g. in *ak tentek* ‘oafish’, *kypkyzyl tentek* ‘blooming fool’.

tüztümüz ‘straight’ (E; RKarS-Haf)

The base **tümüz* does not seem to be attested. However, the word might be a double intensification, a reduplication with the base prepended to it again: *tüz* → *tümtüz* → **tüzütümtüz* → *tüztümüz*, similarly to Russ. *белый-пребелый*. See 3.1.16 on reduplications with the base prepended to them.

tyntyk (E; RKarS-Haf) ◊ **tyk** ‘vertical, upright’

The spelling with ⟨ы⟩ in RKarS-Haf is unusual and probably only meant to denote the lack of palatalization in the preceding *t*, accompanied perhaps by a slightly retracted pronunciation of *i*. In KarRPS, the word is given as *tik*; the Armeno-Kipchak, Crimean Tatar, Ottoman and Urum forms are all front.

2.6.4 Structure

Altogether, attested are 73 examples of *C*-type, derived from 45 unique bases. The distribution is strongly skewed in the two western languages, and much less so in the eastern one.

Eastern (= Crimean) Karaim is the richest in reduplications. Five closing consonants of *C*-type are attested in it in a total of 49 examples derived from 43 bases, in a quite uneven distribution:

m: 12 examples: *baška*, *bijaz*, *boš*, *bošyna*, *ješil*, *jumalak*, *kök*, *sylak*, *tögerek*, *tok*, *tüz*, and *tyk*,

p: 22 examples: *ačmak, ačyk, ak, al, ansyz, ansyzyn, aryk, čövre, jalyjyz, kara, karaž vlyk, kenete, kuru, saglam, sary, sylak, temiz, tez, tok, tolu, tüz, and uzun,*

pp: 2 examples: *ačyk*, and *ak*,

r: 4 examples: *čebik, čyplak, sylak, temiz*, and

s: 9 examples: *bedava, belli, beter, бүтүн, mavy, mor, tamam, tokal, and tomalak.*

Five words have more than one closing consonant possible: *ačyk* (*p* and *pp*), *ak* (*p* and *pp*), *sylak* (*m*, *p* and *r*), *temiz* (*p* and *r*), *tok* (*m* and *p*), and *tüz* (*m* and *p*). The only regularity is that all have *p*. Unusual among them is *sylak*, as it is one of the two examples in this work, which have more than two closing consonants possible; see 3.1.1 for other examples of alternative closers.

Eastern Karaim is also one of two Turkic languages where double *pp* can be used as a closing consonant with any other word than *ak* ‘white’. The word is *ačyk* ‘open’), but as it begins with a vowel, it is at least theoretically possible that the form *appačyk* is in fact a contraction of **appa.ačyk*, i.e. a CV-type reduplication where double *pp* appears to be more common, as in Az. *sappasay* ‘absolutely healthy’ (Zeynalov 1993: 149) &c. See also 3.1.8 on double *pp* as a closer.

Also noteworthy are the forms *jumjumalak* and *jumjumarlak* ‘completely round’, where the closing consonant is the same as the second consonant of the base. See 3.1.6 for parallel examples.

There are two small groups of derivatives: one based on *ansyz* ‘suddenly’, and one on *boš* ‘empty’. For the first, see *apansyzdan* in 2.6.3. The latter only contains *boš* itself, and *bošyna* ‘in vain’. However, since both bases are attested, there are no grounds to suspect anything else than simple reduplications of two independent words.

North- (= Trakai) and **southwestern** (= Luck/Halych) Karaim have a very similar and much more limited stock of reduplications. Three closing consonants are attested in them in a total of 24 examples (12 in NW and 12 in SW), and again, in a rather uneven distribution.

m: 2 + 1 examples: NW *boš*, and NW + SW *ješil*,

p: 10 + 10 examples: NW + SW *ak, boš, incke, jahšy, kara, kenete, kyzyl, sary* and *tolu*, NW *aryk*, SW *čivore*, and

pp: 1 example: SW *ak*.

Two words have more than one closing consonant possible: NW *boš* (*m* and *p*), and SW *ak* (*p* and *pp*).

Seven words have reduplications in both eastern Karaim and the two western languages: *ak* ‘white’, *boš* ‘empty’, *ješil* ‘green’, *kara* ‘black’, *kenete* ‘sudden(ly)’, *sary* ‘yellow’ and *tolu* ‘full’. The closing consonants are almost always the same: *m* in *ješil* and *p* in all the other ones, but there are two exceptions: one is *ak* for which the variant with a

double *pp* is not attested in northwestern Karaim, and the other is *boš* which has *m* in eastern Karaim, both *m* and *p* in the northwestern and only *p* in the southwestern. See *bomboš* in 2.6.3 above.

2.6.5 Semantics

In the great majority of cases, the reduplicated meaning is a simple intensification of the base meaning, or apparently the same. The latter should probably be attributed to dictionary definitions not being sufficiently exhaustive.

Almost all of the base words are closer to being adjectives, if we regard Turkic nomina as a continuous scale. There are two exceptions: *karavlyk*, which is primarily a noun ('darkness'), and *ačmak*, which is quite explicitly a verb ('to open'). The latter is a surprise as reduplicated verbs are most rare in the Turkic languages. See 3.1.22 for the other few examples.

2.7 Karakalpak

Karakalpak *C*-type reduplications are moderately numerous and very uniform.

Only *p* and *pp* are attested as the possible closing consonants, if *besbeter* '?', *kum* ≠ *kuwyt* 'commotion, ...' and *žymžyrt* 'quiet(ly), ...' are put aside as being foreign to the Karakalpak system (borrowed or inherited from a period long predating independent Karakalpak).

One notable feature is to be found in the structure of Karakalpak reduplications. Namely, words beginning with *ža-* have their first vowel palatalized in the reduplication. Out of eight such examples, a half also have alternate, standard, reduplications, e.g. *žaman* 'bad' → *žapžaman* ~ *žäpžaman*. See 2.7.4.

Reduplications of non-adjectives almost do not occur, *tomalak* '1. ball, ...; 2. bit, ...' being the single exception. The meaning of its reduplication, however, is adjectival. See 2.7.5.

2.7.1 Sources

The main source of the material is KklpRS which contains ca. 30 000 entries. Some attestations have also been found in Baskakov 1951–52: 2/1: 210f and Pomorska 2004: 143f.

Transcription:

Baskakov 1951–52: 2/1: 210f: *дж* → *ž* | *кб* → *k* | *нб* → *ŋ* | *ю* → *ü*¹³,

KklpRS: *ə* → *ä* | *ε* → *γ* | *κ* → *k* | *η* → *ŋ* | *θ* → *ö* | *ŷ* → *w* | *γ* → *ü* | *χ* → *h* | *ж* → *ž*.

¹³ Only appears once, in *юлкен*.

2.7.2 Standard cases

- ak** ‘white’ ◇ *appak* (Baskakov 1951–52: 2/1: 210, KklpRS)
ansat ‘easy, toilless’ ◇ *apan̄sat* (KklpRS)
anyk ‘clear, understandable’ ◇ *apanyk* (KklpRS)
ašyk ‘1. open; 2. light, bright’ ◇ *apašyk* (KklpRS)
awyr ‘heavy’ ◇ *apawyr* (KklpRS)
bijik ‘high, tall’ ◇ *bipbijik* (Baskakov 1951–52: 2/1: 211)
boz ‘grey, pale’ ◇ *bopboz* (Baskakov 1951–52: 2/1: 211)
danyyl ‘smooth, even’ ◇ *dapdanyyl* (KklpRS: only attested in ~ *žol* ‘a very smooth road’)
domalak ‘round’ ◇ *dopdomalak* ‘1. completely round; 2. completely full’ (KklpRS)
 See *toptomalak* in 2.7.3 below.
durys ‘correct, faithful, accurate’ ◇ *dupdurys* (KklpRS)
kara ‘black’ ◇ *kapkara* (Baskakov 1951–52: 2/1: 211, KklpRS)
karanyy ‘dark(ness), gloom’ ◇ *kapkaranyy* (KklpRS)
katty ‘hard, solid’ ◇ *kapkatty* (KklpRS)
kiškene ‘small, little’ ◇ *kipkiškene* (KklpRS)
kiškentaj ‘small, little’ ◇ *kipkiškentaj* (KklpRS)
kojyw ‘thick, dense’ ◇ *kopkojyw* ‘1. *intens.*; 2. Okay!, Good!’ (KklpRS)
kuw ‘hoary, pale’ ◇ *kupkuw* (KklpRS)
kyska ‘short’ ◇ *kypkyska* (KklpRS)
kyzyl ‘red’ ◇ *kypkyzyl* ‘very red; bright red’ (Baskakov 1951–52: 2/1: 210, KklpRS, Pomorska 2004: 144)
melle ‘yellow’ ◇ *mepmelle* (Baskakov 1951–52: 2/1: 211)
möldir ‘transparent, clear, pure’ ◇ *möp möldir* (KklpRS)
šak i.a. ‘fit, just’ ◇ *šapšak* (KklpRS)
sary ‘yellow’ ◇ *sapsary* (Baskakov 1951–52: 2/1: 211, KklpRS)
saw ‘healthy’ ◇ *sapsaw* (KklpRS)
semiz ‘fatty, greasy’ ◇ *sepsemiz* (KklpRS)
šukyr ‘deep’ ◇ *šupšukyr* (KklpRS)
suluw ‘beautiful’ ◇ *supsuluw* (KklpRS)
sur ‘1. grey, pale; 2. formidable, fearsome’ ◇ *supsur* (KklpRS)
šymyr ‘strong, thick, dense’ ◇ *šypšymyr* (KklpRS)
synyk ‘meek, gentle, soft’ ◇ *sypsynyk* ‘very suave, very courteous’ (KklpRS)
tajar ‘ready’ ◇ *taptajar* (KklpRS)
takyr ‘naked, bare’ ◇ *taptakyr* (KklpRS)
taza ‘clean, neat’ ◇ *taptaza* (KklpRS)
tegis ‘smooth, even’ ◇ *teptegis* (KklpRS)
teñ ‘smooth, even’ ◇ *tepten* (KklpRS)
tez ‘fast, quick’ ◇ *teptez* (Baskakov 1951–52: 2/1: 211)

- tik** ‘straight, erect, vertical’ ◇ *tiptik* (KklpRS)
- tikke** ‘straight, erect, vertical’ ◇ *tiptikke* (KklpRS)
- tokalak** ‘lopped’ ◇ *toptokalak* ‘hummel, hornless’ (KklpRS)
- tomalak** ‘1. ball; hank; 2. bit, morsel, crumb’ (Baskakov 1951–52: 1: 194, 386)
 ◇ *toptomalak* ‘1. completely round; 2. full of a person’ (KklpRS)
 See *toptomalak* in 2.7.3 below.
- tompak** ‘1. full, thick; 2. convex’ ◇ *toptompak* (KklpRS)
- tuwra** ‘straight, direct’ ◇ *tuptuwa* (KklpRS)
- tynyk** ‘1. transparent, clear; 2. quiet, calm’ ◇ *typtynyk* (KklpRS: only attested in *hawa* ~ ‘a perfectly clear sky, a completely cloudless sky’)
 See also 3.1.11 for the entire family.
- ülken** ‘big, huge’ ◇ *üpülken* (Baskakov 1951–52: 2/1: 211, KklpRS)
- uzak** i.a. ‘distance, distant, far’ ◇ *upuzak* (KklpRS)
- uzun** (Baskakov 1951–52: 2/1: 211) ~ *uzyn* (KklpRS) ‘long’ ◇ *upuzun* (Baskakov 1951–52: 2/1: 211) ~ *upuzyn* (KklpRS)
- yras** ‘true, faithful, accurate’ ◇ *ypyras* (KklpRS)
- žaksy** ‘good’ ◇ *žäpžaksy* (KklpRS)
- žakty** ‘light, bright’ ◇ *žäpžakty* (KklpRS)
- žakyn** ‘close, near’ ◇ *žapžakyn* (Baskakov 1951–52: 2/1: 211, KklpRS) ◇ *žäpžakyn* (KklpRS)
- žalyyz** ‘only, single, lonely,’ ◇ *žäpžalyyz* (KklpRS)
- žaman** ‘bad’ ◇ *žapžaman* (Baskakov 1951–52: 2/1: 211) ◇ *žäpžaman* (KklpRS)
- žana** (KklpRS) ~ *žany* (Baskakov 1951–52: 2/1: 211) ‘new’ ◇ *žapžana* (KklpRS) ~ *žapžany* (Baskakov 1951–52: 2/1: 211) ◇ *žäpžana* (KklpRS)
- žas** ‘young’ ◇ *žäpžas* (KklpRS)
- žasyl** ‘green’ ◇ *žapžasyl* (Baskakov 1951–52: 2/1: 211, KklpRS) ◇ *žäpžasyl* (KklpRS)
- ženil** ‘light’ ◇ *žepženil* (KklpRS)
- žuka** ‘thin’ ◇ *župžuka* (KklpRS)
- žumsak** ‘soft’ ◇ *župžumsak* (KklpRS)
- žumyry** ‘round’ ◇ *župžumyry* (KklpRS)
- žuwan** ‘thick, fat’ ◇ *župžuwan* (KklpRS)

2.7.3 Special cases

The general tendency in the Kipchak languages is that reduplications are relatively numerous, and *p* and *pp* are the only possible closing consonants. Karakalpak seems to follow this pattern, but three words go against it: two unclear forms that might be reduplications closed by *m* (*kumkuwyt* and *žymžyrt*), and one reduplication closed by *s* (*besbeter*).

In all three cases, alternative and plausible explanations can be found, which assume that these words have been formed outside of Karakalpak. In lack of any conclusive evidence, they will be treated as special here, and ignored in further considerations.

besbeter (Baskakov 1951–52: 2/1: 211, KklpRS, Pomorska 2004: 144) \diamond **beter** ‘intensifier’ (Baskakov 1951–52: 1: 327); *bad (Baskakov 1951–52: 2/1: 211, Pomorska 2004: 144); more (Baskakov 1951–52: 1: 327, KklpRS); worse (e-sozlik)’

There is a certain confusion regarding the meaning of *beter* in Karakalpak. A closer inspection of the examples adduced in the sources suggests that perhaps ‘worse’ might be the most accurate translation. All the other propositions (‘intensifier’, ‘bad’ and ‘more’) seem correct in some sentences but much less so in others, and only ‘worse’ appears to fit all. This conclusion is consistent with the history of the word which ultimately stems from Pers. *بَدْتار* *badtar*, the superlative of *بَد* *bad* ‘bad’. But see below.

Closed by *s*, *besbeter* has no parallel in Karakalpak. This invites the idea of borrowing, and a perfect candidate can be found in Turkmen. There are as many as 12 or 13 reduplications closed by *s* in it (more than 20% of the total number), it is the natural geographic intermediary between Persian and Karakalpak, and what is more, *besbeter* has two meanings in it: ‘1. more so, very; 2. worst of all’ which very well fits the inconsistency seen in the translations of Kklp. *besbeter*. See 2.17.2 on Trkm. *beter*.

Note that the general rule adopted in this work is to only exclude reduplications from further considerations, if their base is not attested independently in the given language. Kklp. *besbeter* is an exception from this practice that was only made due to the unusual accumulation of premises pointing to a borrowing.

A parallel exception, also concerning *beter*, has been made in Kumyk, see 2.11.3.

kumkuwyt ‘excitation, commotion, agitation’ (KklpRS)

The base **kuwyt* does not seem to be attested, and the whole form is not clear. The *m* as the supposed closing consonant is quite unusual – but see *besbeter* above and *žymžyrt* below. Perhaps a borrowing or merely an apparent reduplication (see 3.1.3 for other examples)? A parallel situation with dominating *p* and two unclear cases in supposed *m* can be found in Kazakh (see 2.8.3).

toptomalak ‘1. completely round; 2. full of a person’ (KklpRS) \diamond **tomalak** ‘1. ball; hank; 2. bit, morsel, crumb’ (Baskakov 1951–52: 1: 194, 386)

This word and *domalak* ‘round’ (see 2.7.2 above) are no doubt related. The exact nature of this relationship, however, is not clear. Importantly, the closing consonant is the same in both cases (*p*).

žymžyrt ‘quiet(ly), silence’ (Baskakov 1951–52: 2/1: 211, KklpRS)

The base **žyrt* does not seem to be attested, but the word is not an isolate. See 3.1.15 for possible cognates.

2.7.4 Structure

Technically, two closing consonants of *C*-type are attested in a total of 60 examples derived from 60 unique bases, in a very uneven distribution:

p: 59 examples: *anyk, aḡsat, ašyk, awyr, bijik, boz, daḡḡyl, domalak, durys, kara, karayy, katty, kiškene, kiškentaj, kojyw, kuw, kyska, kyzyl, melle, möldir, šak, sary, saw, semiz, šukyr, suluw, sur, šymyr, synyk, tajar, takyr, taza, tegis, teḡ, tez, tik, tikke, tokalak, tomalak, tompak, tuwra, tynyk, ülken, uzak, uzun, yras, žaksy, žakty, žakyn, žalyyz, žaman, žaḡa, žas, žasyl, žeḡil, žuka, žumsak, žumyry, žuwan,* and

pp: 1 example: *ak*.

Not even *ak* ‘white’ has more than one closer possible.

Unusual are the eight words in *ža-*, because all can have their reduplicated vowel palatalized to *ä*. In four cases (*žaksy* ‘good’, *žakty* ‘light, bright’, *žalyyz* ‘only, single, lonely’, and *žas* ‘young’), only the *-ä*-form is attested; in the remaining four (*žakyn* ‘close, near’, *žaman* ‘bad’, *žaḡa* ‘new’ and *žasyl* ‘green’) – both, the *-a-* and *-ä-*forms are. All attestations are from KklpRS, but since they are listed s.v. *жап* and *жап* as examples illustrating the use of these ‘intensifiers’, it must be suspected that alternate forms might possibly exist for *all* reduplicated words in *ža-*. A plausible, albeit not necessarily *sine qua non* condition that would allow such a phonetic assimilation to occur, is that reduplications lose their morphological lucidity. From the available data, no definite conclusions can be drawn.

2.7.5 Semantics

In almost all cases, the reduplicated meaning is either a simple intensification of the base meaning, or very close to it. Also, almost all words are, for all practical purposes, adjectives.

The only two exceptions are *kojyw*, the reduplication of which has evolved an additional meaning not very directly related to the meaning of the base (‘thick, dense’ → ‘Okay!, Good!’), and *tomalak*, the meaning of which is of a quite clear substantival character (‘1. ball; hank; 2. bit, morsel, crumb’). Note that the meaning of the reduplicated *toptomalak* is no longer a noun, but already a plain adjective (‘1. completely round; 2. full of a person’).

2.8 Kazakh

C-type reduplications are very numerous in Kazakh, but also very uniform and standard.

Only *p* and *pp* are attested as closing consonants. Reduplications of non-adjectives or non-trivial semantic evolutions effectively do not occur. See 2.8.4 and 2.8.5.

Unusually, however, special cases seem to fall into distinct groups rather than being a loose collection of unrelated words. See in particular *tymtyrakaj* ‘in disarray, ...’ and *žypžylmajaj* ‘without a trace, ...’ in 2.8.3.

2.8.1 Sources

The main sources of the material are KzkRS and SKzkP, which contain ca. 21 000 and 15 000 entries, respectively. Some attestations have also been found in Balakaev 1959, Pomorska 2004: 144, Ščerbak 1977: 120, and in the grammars listed below.

For grammatical descriptions, Somfai Kara 2002: 33 and Kulikovskaja/Musaeva 2006: 69 have been used. Both are rather brief and effectively limited to the statement that the reduplicated syllable is closed by *p*.

Transcription:

KzkRS, Kulikovskaja/Musaeva 2006, SKzkP: $\text{ə} \rightarrow \text{ä} \mid \text{ɛ} \rightarrow \text{ɣ} \mid \text{u} \rightarrow \text{ij} \mid \text{i} \rightarrow \text{i} \mid \text{κ} \rightarrow \text{k} \mid \text{ŋ} \rightarrow \text{ŋ} \mid$
 $\text{ø} \rightarrow \text{ö} \mid \text{y} \rightarrow \text{uw} \mid \text{V} \text{y} \rightarrow \text{w} \mid \text{ɣ} \rightarrow \text{ü} \mid \text{ɣ} \rightarrow \text{u},$
 Ščerbak 1977: $\text{ĩ} \rightarrow \text{y}.$

2.8.2 Standard cases

ädemi ‘beautiful’ \diamond *äpädemi* (KzkRS, Kulikovskaja/Musaeva 2006, SKzkP)
ajkyn ‘clear, precise, explicit’ \diamond *apajkyn* (KzkRS)
ak ‘white’ \diamond *appak* (KzkRS)
ala ‘motley, colourful; palomino’ \diamond *apala* ‘all in patches’ (SKzkP)
alasa ‘low’ \diamond *apalasa* (KzkRS)
alys ‘far’ \diamond *apalys* (SKzkP)
anyk ‘clear, obvious’ \diamond *apanyk* (KzkRS, SKzkP)
aryk ‘thin, lean’ \diamond *aparyk* (KzkRS)
arzan ‘cheap’ \diamond *aparzan* (KzkRS)
äsem ‘beautiful’ \diamond *äpäsem* (SKzkP)
ašyk ‘open’ \diamond *apašyk* (KzkRS, SKzkP)
awyr ‘heavy; hard, difficult’ \diamond *apawyr* (KzkRS)
äzir ‘ready’ \diamond *äpäzir* (SKzkP)
bajayy ‘1. old, past, earlier; 2. just this, precisely this; 3. ordinary’ \diamond *bapbajayy* ‘1. very old; 2. as before, as used to be’ (SKzkP)
bajsaldy ‘calm, staid’ \diamond *bapbajsaldy* (SKzkP)
begili ‘known’ \diamond *bepbegili* ‘commonly known, very obvious’ (SKzkP)
berik ‘tough, durable’ \diamond *bepberik* (SKzkP)
bijik ‘tall’ \diamond *bipbijik* (SKzkP)
bos i.a. ‘empty’ \diamond *bopbos* (SKzkP)

- bujra** ‘curly’ ◇ *bupbujra* (SKzkP)
bütün ‘whole’ ◇ *büpbütün* (SKzkP)
dajar ‘ready’ ◇ *dapdajar* (KzkRS)
dajyn ‘ready’ ◇ *dapdajyn* (KzkRS)
däl ‘just, precisely’ ◇ *däpdäl* ‘precisely so’ (SKzkP)
dämdi ‘tasty’ ◇ *däpdämdi* (Kulikovskaja/Musaeva 2006: 69)
dardaj ‘stalwart, portly, strapper’ ◇ *dapdardaj* ‘quite stalwart, quite portly’ (KzkRS, SKzkP)
domalak ‘round’ ◇ *dopdomalak* (KzkRS, SKzkP)
durys ‘correct, accurate, proper’ ◇ *dupdurys* (KzkRS, SKzkP)
erte ‘(early) in the morning’ ◇ *eperte* (KzkRS)
kara ‘black’ ◇ *kapkara* (KzkRS, Somfai Kara 2002: 33, SKzkP)
karaŋgy ‘dark’ ◇ *kapkaraŋgy* (KzkRS, SKzkP)
katty ‘hard, solid’ ◇ *kapkatty* (KzkRS, SKzkP)
kiškentaj ‘small, little’ ◇ *kipkiškentaj* (KzkRS, SKzkP)
kurçak ‘dry’ ◇ *kupkurçak* (KzkRS, SKzkP)
kuw ‘1. dry, dried up; 2. yellowed, faded’ ◇ *kupkuw* (SKzkP)
kuwnak ‘brisk, sprightly’ ◇ *kupkuwnak* (SKzkP)
kyzyl ‘red’ ◇ *kypkyzyl* (KzkRS, SKzkP)
majda ‘small, fine’ ◇ *mapmajda* (SKzkP)
möldir ‘transparent, clear’ ◇ *möpmöldir* (KzkRS, SKzkP)
momakan ‘obedient, well-behaved’ ◇ *mopmomakan* (SKzkP)
momyn ‘humble, quiet, conciliatory’ ◇ *mopmomyn* (SKzkP)
muzdaj ‘cold’ ◇ *mupmuzdaj* (KzkRS, SKzkP)
nyk ‘tough, durable’ ◇ *nypnyk* (SKzkP)
oŋaj ‘easy’ ◇ *opoŋaj* (KzkRS, SKzkP)
öŋdi ‘1. with (nice) skin, with (nice) complexion, with (nice) colour; 2. obverse’
 ◇ *öpöŋdi* ‘very beautiful’ (SKzkP)
ötirik ‘a lie’ ◇ *öpötirik* ‘a blatant, outright lie’ (SKzkP)
šayyn ‘small, little’ ◇ *šapšayyn* (KzkRS, SKzkP)
šak ‘just right’ ◇ *šapšak* (SKzkP)
salmakty ‘serious, weighty, ponderous’ ◇ *sapsalmakty* (KzkRS)
sary ‘yellow, flaxen, fair(-haired)’ ◇ *sapsary* (Ščerbak 1977: 120, KzkRS, Pomorska 2004: 144, SKzkP)
saw ‘healthy’ ◇ *sapsaw* (KzkRS, SKzkP)
šeber ‘master’ ◇ *šepšeber* (Balakaev 1959: 35)
seldir ‘loose, sparse, thin’ ◇ *sepseldir* (KzkRS, SKzkP)
semiz ‘fat’ ◇ *sepsemiz* (SKzkP)
sergek ‘alert, snappy’ ◇ *sepsergek* (KzkRS, SKzkP)
šijki ‘raw, green’ ◇ *šipšijki* (SKzkP)

- šolak** ‘short, scanty’ ◇ **šopšolak** (SKzkP)
šošak ‘pointed, conical’ ◇ **šopšošak** (SKzkP)
šubar ‘dappled, mottled’ ◇ **šupšubar** (SKzkP)
süjir ‘sharp, pointed’ ◇ **süpsüjir** (SKzkP)
šunɣyl ‘1. deep; 2. sunken, hollow’ ◇ **šupšunɣyl** (SKzkP)
šunɣyr ‘hole, pit, hollow’ ◇ **šupšunɣyr** ‘all in: holes, pits’ (SKzkP)
sur ‘grey, pale’ ◇ **supsur** (KzkRS, SKzkP)
synyk ‘crack(ed), fracture(d)’ ◇ **sypsynyk** ‘completely broken, shattered’ (SKzkP)
tajaw ‘close, near’ ◇ **taptajaw** (SKzkP)
tapal ‘low’ ◇ **taptapal** (SKzkP)
 See 3.1.6 on the closer being identical to C_2 .
tar ‘tight, narrow’ ◇ **taptar** (SKzkP)
tastaj ‘like a stone’ ◇ **taptastaj** ‘1. rock-solid; 2. stone-cold’ (SKzkP)
tätti ‘sweet’ ◇ **täptätti** (SKzkP)
täwir ‘good, not bad’ ◇ **täptäwir** (KzkRS, SKzkP)
tegin ‘free of charge’ ◇ **teptegin** (SKzkP)
tegis ‘smooth, even’ ◇ **teptegis** (KzkRS, SKzkP)
tez ‘fast, quick’ ◇ **teptez** (SKzkP)
tike ‘straight, vertical, steep’ ◇ **tiptike** (SKzkP)
tolyk ‘full, complete’ ◇ **toptolyk** ‘1. *intens.*; 2. puffy, fat’ (SKzkP)
tügel ‘everything, everyone’ ◇ **tüptügel** ‘full, complete’ (KzkRS, SKzkP)
tunyk ‘transparent’ ◇ **tuptunyk** (SKzkP)
tutas i.a. ‘full, complete’ ◇ **tuptutas** (SKzkP)
tuwra ‘straight’ ◇ **tuptuwra** (SKzkP)
tyyyz ‘dense, close, tight’ ◇ **typtyyyz** (SKzkP)
tynyk ‘quiet, peaceful’ ◇ **typtynyk** (SKzkP)
 See 3.1.11 on possible cognates.
tynyš ‘quiet, peaceful’ ◇ **typtynyš** (KzkRS, SKzkP)
 See 3.1.11 on possible cognates.
ülken ‘big, huge’ ◇ **üpülken** (KzkRS, SKzkP)
usak ‘small, fine’ ◇ **upusak** (KzkRS, SKzkP)
uzyn ‘long’ ◇ **upuzyn** (SKzkP)
žaksy ‘good’ ◇ **žapžaksy** (KzkRS, Kulikovskaja/Musaeva 2006: 69)
žakyn ‘1. close, near; 2. recent’ ◇ **žapžakyn** (SKzkP)
žalaŋaš ‘naked, bare’ ◇ **žapžalaŋaš** (SKzkP)
žalpak ‘flat’ ◇ **žapžalpak** (SKzkP)
žaman ‘bad’ ◇ **žapžaman** (SKzkP)
žana ‘new(ly)’ ◇ **žapžana** (KzkRS, SKzkP)
žaryk ‘clear, bright’ ◇ **žapžaryk** (KzkRS, SKzkP)
žas i.a. ‘young’ ◇ **žapžas** (SKzkP)

žasyl ‘green’ ◊ **žapžasyl** (SKzkP)

žatyk ‘1. lying, horizontal; 2. sloping, inclined’ ◊ **žapžatyk** ‘even, smooth, sleek’ (KzkRS)

ženil ‘light’ ◊ **žepženil** (SKzkP)

žiji ‘often’ ◊ **žipžiji** (SKzkP)

žijren ‘sorrel, red (hair)’ ◊ **žipžijren** (SKzkP)

žiniške ‘thin, narrow’ ◊ **žipžiniške** (SKzkP)

žuka ‘thin’ ◊ **župžuka** (KzkRS, Kulikovskaja/Musaeva 2006: 69, SKzkP)

žumsak ‘soft, tender, gentle, subtle’ ◊ **župžumsak** (KzkRS, SKzkP)

žumyr ‘round’ ◊ **župžumyr** (KzkRS, SKzkP)

žuwan ‘fat’ ◊ **župžuwan** (KzkRS, SKzkP)

žuwas ‘1. quiet; 2. calm, obedient’ ◊ **župžuwas** ‘1. very quiet; 2. very calm’ (SKzkP)

žyjnaky ‘arranged, orderly’ ◊ **žypžyjnaky** ‘1. *intens.*; 2. very tidy, sparkling clean’ (KzkRS, SKzkP)

žyldam ‘quick, fast, hasty’ ◊ **žypžyldam** (SKzkP)

žyltyr ‘shiny, glossy’ ◊ **žypžyltyr** ‘very smooth, polished’ (SKzkP)

žyly ‘warm’ ◊ **žypžyly** (KzkRS, SKzkP)

See also *žypžylmayaj* ‘without a trace, clean; very smooth’ in 2.8.3 below.

2.8.3 Special cases

Kazakh special cases, with the exception of *taptapal* ‘very low’ and *župžubymen* ‘in pairs, in twos’, seem to fall into five partially overlapping groups:

1. closed by *m* (save apparent reduplications and loanwords): *tymtyrakaj*, *tymtyrys*, *žymžylas* and *žymžyrt*,
2. with the root *tyr*: *tymtyrakaj* and *tymtyrys*,
3. with the root *žyl*: *žypžylmayaj*, *žymžylas* and, although not a special case, *žypžyly*,
4. with the root *žyl* ~ *žyr*: *žypžylmayaj*, *žymžylas* and *žymžyrt*,
5. *žymžyrt*, which clearly is related to words in Karakalpak, Turkish and Uighur, and *tymtyrys* which might belong to the same family.

Group 1 collects four words which appear to be reduplications closed by *m* but whose bases do not seem to be attested.

Against 107 Kazakh reduplications closed by *p*, this group strongly resembles the situation in Karakalpak where along 59 standard examples in *p*, two special cases can be found in *m*. Incidentally, one of them is common to both languages: Kklp. *žymžyrt* :: Kzk. *žymžyrt*. See also 3.1.15 for other possible cognates of the word.

It is not clear whether these words are apparent reduplications, loanwords, or simply their bases had gone out of use (see also 3.1.3 and 3.1.15 for examples of these categories). In theory, they could also be the last relics of hypothetical bygone reduplications

in *m* in Kazakh; see in particular *žypžylmayaj* below, which is probably related to the standard *žypžylu* in 2.8.2 above.

Group 2 might be just a coincidental phonetic similarity. There is no obvious semantic tertium comparationis to connect its two members.

Group 3 is fragile from the semantic point of view, but not altogether impossible. See *žypžylmayaj* below.

Group 4 is semantically more plausible than group 3 ('without a trace' : 'silence') but somewhat uncertain phonetically. Maybe variants of an onomatopoeic root?

Group 5, finally, includes the Kazakh representative in a small family of words which obviously are related, but the exact relationships remains unclear. There is no clue by which to judge with certainty whether it is a case of multiple borrowing, or an archaism whose structure had long become obscure to the speakers, and hence the surviving *m*. See 3.1.15 for more commentary.

To sum up, as far as grouping is concerned, groups three and five are possible but definitely not trivial. Groups one, two and four might well be coincidental.

tyntyrakaj 'in disarray, disorderly, in all directions' (KzkRS, SKzkP)

The base **tyrakaj* does not seem to be attested. The closer *m* is highly unusual as all the 108 reduplications in Kazakh are closed by *p* or *pp* – but see the comment on group 1 in 2.8.3 above.

The similarity of the hypothetical root **tyrakaj* to that of *tyntyrys* is probably illusory, in view of the semantics. A connection of the latter with *žymžyrt* (see below) seems more plausible. See also Kirg. *tyntyrakaj* 'in disarray, every which way' in 2.10.3.

tyntyrys '(in) complete silence, completely silent, completely mute' (Balakaev 1959: 173, KzkED, KzkRS, SKzkP)

The base **tyrys* does not seem to be attested. The word is surely related to Kirg. *tyntyrys* 'complete silence' (see 2.10.3), and possibly several other forms in different languages; see 3.1.15. See also *tyntyrakaj* above and *žymžyrt* below.

župžubymen 'in pairs, in twos' (SKzkP)

The base **žubymen* does not seem to be attested, and the whole form is not clear.

žypžylmayaj 'without a trace, clean (KzkRS), very smooth (SKzkP)'

The base **žylmayaj* does not seem to be attested. It is probably related to *žymžylas* below.

The origin of this word is not clear. Maybe it is a derivative from *žylu-* 'to (get) warm'? See *žylymšy* 'rotten, musty' and *žylmakaj* 'sleek, smarmy, slick, smooth-tongued, slippery' (KzkRS). The semantic evolution would have to go along these lines: 'warm' → 'rotten' → 'smarmy' → 'slippery' → 'clean' → 'without a trace'.

See the standard case *žyly* ‘warm’ in 2.8.2 above. The exact relations between these words are not clear but if they were found to eventually be related, a hypothesis could be put forward based on them that *m* had been at some remote point in time a valid closing consonant in Kazakh.

žymžylas ‘without a trace’ (KzkRS)

The base **žylas* does not seem to be attested. For *-m-*, see *tyntyrakaj* above; for *žyl-*, see *žypžylmajaj* above.

žymžyrt ‘complete silence’ (KzkRS)

The base **žyrt* seems to only be attested as a verb root with the meaning ‘to tear, to rip’. This root is probably the parent of a small family of related forms in six different languages, and ultimately also of Kzk. *tyntyrys* ‘(in) complete silence, completely silent, completely mute’. See 3.1.15.

2.8.4 Structure

Two closing consonants of *C*-type is attested in a total of 108 examples derived from 108 unique bases, in a rather one-sided distribution:

p: 107 examples: *ädemi, ajkyn, ala, alasa, alys, anyk, aryk, arzan, äsem, ašyk, awyr, äzir, bajajy, bajsaldy, begili, berik, bijik, bos, bujra, bütin, dajar, dajyn, дәl, дәmdi, dardaj, domalak, durys, erte, kara, karajy, katty, kiškentaj, kuryak, kuw, kuwnak, kyzyl, majda, möldir, momakan, momyn, muzdaj, nyk, oňaj, öňdi, ötirik, šajyn, šak, salmakty, sary, saw, šeber, seldir, semiz, sergek, šijki, šolak, šošak, šubar, süjir, šujyyl, šuňkyr, sur, synyk, tajaw, tapal, tar, tastaj, tätti, täwir, tegin, tegis, tez, tike, tolyk, tügel, tynyk, tutas, tuwra, tyyyz, tynyk, tynyš, ülken, usak, uzyn, žaksy, žakyn, žalaňaš, žalpak, žaman, žaňa, žaryk, žas, žasyl, žatyk, žeňil, žiji, žijren, žiňiške, žuka, žumsak, žumyr, žuwan, žuwas, žyjnaky, žyldam, žyltyr, žyly, and*

pp: 1 example: *ak*.

One quite specific form is *taptapal* ‘very low’; see 3.1.6 for other reduplications with the closing consonant identical to *C*₂.

2.8.5 Semantics

In almost all the cases, the reduplicated meaning is a straightforward intensification or (apparently) the same as the base, the latter resulting probably from imprecise dictionary definitions. The not entirely trivial evolutions can be found in *öňdi*, where i.a. ‘with nice complexion’ → ‘very beautiful’, and *tügel*, where ‘everything, everyone’ → ‘full, complete’.

The great majority are quite clear adjectives. A stronger substantival component can only be found in *ötirik* ‘a lie’, *šeber* ‘master’, *šuňkyr* ‘hole, pit, hollow’ and *synyk* ‘crack(ed), fracture(d)’.

2.9 Khakas

C-type reduplications are pronouncedly more numerous in Khakas than in most South Siberian languages, and on par only with Tuvinian. Otherwise, they are quite compatible with the others in that *p* is the only closing consonant, and that semantics is confined within the borders of relatively basic adjectival and adverbial meanings – albeit not as strongly limited to colour names as grammars tend to picture it.

Only two cases can be considered special: *appagas* ‘snow-white’ and *köppeges* ‘very blue’. Both belong to a large family of related forms in Khakas, Oirot and Shor. See 2.9.3.

2.9.1 Sources

The main source of the material is HakRS which contains ca. 14 000 entries. Some attestations have also been found in Anderson 1998: 23, Butanaev 1999, HakOS, Li et al. 2007, Ščerbak 1977: 120, Serebrennikov/Gadžieva 1986: 112, Subrakova 2006, and the grammatical descriptions listed below.

For grammatical descriptions, Dyrenkova 1948: 42, HakRS: 409, Patačakova 1962: 151, Babuškin 1975: 94, and Anderson 1998: 23 have been used.

HakRS: 409 does not mention closing consonants at all. Interestingly, the process is described in it as happening “обычным удвоением или удвоением прилагательных с сокращением первого компонента до одного слога”.

Dyrenkova 1948: 42 mentions the prevalence of colour names, gives an inaccurate description of the process and some examples, and states that the closing consonant is always *p*.

Patačakova 1962: 151 effectively limits herself to a very general information and some examples, without specifying what the closing consonant is.

Babuškin 1975: 94 only gives a somewhat unclear description of the process and several examples, but does not specify what the closing consonant is.

Anderson 1998: 23 states that reduplication is limited in Khakas to “a small number of basic color terms” and two more words (*čön* ‘fat’ and *čaryh* ‘light, bright’), and that the closing consonant is always *p*. Also, he interprets the process as prefixation of *CVp-*, but theoretical issues of this kind will not interest us here (see fn. 3).

Transcription:

Anderson 1998: $i \rightarrow y \mid x \rightarrow h \mid VV \rightarrow \bar{V}$,

Babuškin 1975: 94, Dyrenkova 1948: 42, HakOS, HakRS, Patačakova 1962, Serebrennikov/Gadžieva 1986: $u \rightarrow \check{z} \mid z \rightarrow g \mid i \rightarrow \check{i} \mid n\bar{b} \rightarrow \eta \mid \bar{o} \rightarrow \bar{o} \mid VV \rightarrow \bar{V}$,

Li et al. 2007: $\gamma \rightarrow g \mid \check{i} \rightarrow y$.

Ščerbak 1977: $\bar{b} \rightarrow g \mid \check{i} \rightarrow y$,

2.9.2 Standard cases

agrin &c. ‘slowly, quietly, gradually’ ◇ **apagrin** ‘very quietly’ (HakRS) ~ **apagyryn** (HakOS: no meaning given)

agyryn see *agrin*

ajas ‘light, bright’ ◇ **apajas** (HakRS, HakOS)

amyr ‘peace(ful), calm, quiet’ ◇ **apamyr** (HakOS)

aryg ‘clean; bright, clear, serene; transparent, limpid; cleanly’ ◇ **aparyg** (HakRS: ‘quite clean; very cleanly’; HakOS, Li et al. 2007: no meaning given)

čagyn ‘close, near’ ◇ **čapčagyn** (HakRS)

čaryh ‘light, bright’ ◇ **čapčaryh** (HakRS, Babuškin 1975: 94, HakOS, Anderson 1998: 23¹⁴)

čjke ‘straight, direct’ ◇ **čjpčjke** (HakRS, HakOS)

čōn ‘fat’ ◇ **čopčōn** (Patačakova 1962: 151, Babuškin 1975: 94, Anderson 1998: 23, Pomorska 2004: 144)

čylbyraŋ ‘smooth, even’ ◇ **čypčylbyraŋ** (Patačakova 1962: 151, HakOS)

hara ‘black’ ◇ **haphara** (Dyrenkova 1948: 42, HakRS, HakOS, Anderson 1998: 23)

harashy ‘dark’ ◇ **hapharashy** (Babuškin 1975: 94)

hū ‘pale, white’ ◇ **huphū** ‘very pale’ (HakRS)

hyjyr ‘oblique, slanted, crooked’ ◇ **hyphyjyr** (Patačakova 1962: 151)

hyr ‘grey, hoary, roan’ ◇ **hyphyr** (HakRS)

hyzyl ‘red’ ◇ **hyphyzyl** ‘extremely red, the reddest’ (Dyrenkova 1948: 42, Babuškin 1975: 94, Ščerbak 1977: 120, Anderson 1998: 23); bright red (HakRS)

kinetjŋ ‘suddenly’ ◇ **kipkinetjŋ** (HakRS, HakOS)

nā ‘new’ ◇ **napnā** (HakRS, HakOS, Li et al. 2007)

orta ‘correct’ ◇ **oporta** (HakRS)

sah ‘sober’ ◇ **sapsah** (HakRS, HakOS)

saryg ‘yellow’ ◇ **sapsaryg** (Dyrenkova 1948: 42, HakRS, Patačakova 1962: 151, Babuškin 1975: 94, Ščerbak 1977: 120, Serebrennikov/Gadžieva 1986: 112, HakOS)

sirgek ‘sensitive, sharp-eyed’ ◇ **sipsirgek** (Patačakova 1962: 151)

symsyryh ‘silence, quiet(ly)’ ◇ **sypsymsyryh** ‘very quiet(ly)’ (HakRS)

See *symsyryh* in 2.9.3 below.

syn ‘truth(ful), true’ ◇ **sypsyn** (HakRS)

tadylyg ‘sweet’ ◇ **taptadylyg** (HakRS: 409, Patačakova 1962: 151)

tegilek see *tigilek*

tērpek ‘circle, round’ ◇ **teptērpek** (HakOS)

tigilek &c. ‘1. wheel; 2. round’ ◇ **tiptiglek**¹⁵ (Dyrenkova 1948: 42) ~ **teptegilek** (HakOS)

¹⁴ Not explicitly. Only mentions that “[...] and čarix ‘bright, light’–have an expressive reduplicative prefix [RD] of the shape CVp-”.

¹⁵ With a not unusual syncope of the middle high vowel in a three-syllable word.

tin ‘1. same, equal; 2. even, smooth’ ◊ **tiptin** (HakRS: only attested in \sim *kis-* ‘to cut: evenly, smoothly’, HakOS)
toglah ‘round’ ◊ **toptoglah** (HakRS)
togyr ‘across, opposite, against’ ◊ **toptogyr** (HakRS)
tōj ‘similar, like’ ◊ **tōptōj** (HakRS, HakOS)
tüs ‘smooth, even’ ◊ **tüptüs** (HakOS)
tyñ ‘strong(ly), powerful(ly)’ ◊ **typtyñ** ‘very strongly, very powerfully’ (HakRS)
uzun &c. ‘long’ ◊ **upuzun** (HakRS) \sim **upuzyn** (Patačakova 1962: 151)
uzyn see *uzun*

2.9.3 Special cases

In Khakas, Shor and Oiro, *ak* and *kök* have grown into entire families of 17 or 18 forms in total. The patterns are very similar and often the commentary for Oiro, *mutatis mutandis*, also applies for Khakas. See 2.12.3 for more commentary, and also 3.1.11 for other such families.

appagas (HakRS: 409, Patačakova 1962: 151) ◊ **ah** ‘white’

According to HakRS, *appagas* is composed of *ap* + *ah* + *as*. The last morpheme must be the diminutive suffix *-as*, as in *ajah* ‘cup’ → *ajagas* ‘small cup’ (HakRS: 404). Since the shape *agas* seems to only exist with the meaning ‘tree; wood’, which is a different word altogether, the diminutive was most probably added to the already reduplicated form. The whole is then a multiple intensification; see 3.1.13 for more examples.

As for the double *pp*, different possible explanations exist (see *mutatis mutandis* Oir. *appāš* in 2.12.3), but the most probable one is perhaps that of emphatic lengthening (see 3.1.12).

Interestingly, the simple reduplication **apah* does not seem to exist at all.

The word, together with *köppeges* ‘light blue, very blue’ below, belongs to a large family of related forms in Khakas, Oiro and Shor; see 2.12.3, and also 3.1.11 for other such families.

köppeges ‘light blue; very blue’ (HakRS, HakRS: 409, Patačakova 1962: 151) ◊ **kök** ‘1. green; 2. blue’

Patačakova 1962: 151 confirms that this word stems from *kök* ‘1. green; 2. blue’. It might have arisen in at least three different ways (see *mutatis mutandis* Oir. *köppōš* in 2.12.3), but neither seems to be more likely than the others. Here, additionally a backward vowel harmony will most probably need to be assumed to explain the *e* in the middle syllable. Similarly in Shor *köpegeš* (see 2.14.3).

The diminutive *-es* was probably added to the already reduplicated shape as **ke|ög|kes* does not seem to be attested on its own.

Similarly to *appagas* ‘snow-white’ above, a simple **köpkök* does not seem to exist. Both words are part of a large family of related forms in Khakas, Oirod and Shor; see 2.12.3, and also 3.1.11 for other such families.

symsyryh ‘silence, quiet(ly)’ (HakRS)

This word might be an ancient reduplication and belong to one family with such words as Kar.E *čymčyrt* ‘complete silence’ and similar.

The base *sryh* is attested in Khakas with several meanings: ‘1. fine snow; 2. drizzle; 3. arrow with a tetrahedral head; 4. bone plate used for whistling arrows¹⁶ (Subrakova 2006); 5. whistling arrow (Butanaev 1999)’. None of these can be directly linked to ‘silence’ and, it seems, indeed to one another. A certain connection, however, would be conceivable if *sryh* really were cognate with *čyrt*.

See 3.1.15 for more commentary.

2.9.4 Structure

Only one closing consonant of *C*-type is attested in a total of 33 examples derived from 33 unique bases:

p: 33 examples: *agrin, ajas, amyr, aryg, čagyn, čaryh, čike, čōn, čylbyraŋ, hara, harashy, hū, hyjyr, hyr, hyzyl, kinetŋn, nā, orta, sah, saryg, sirgek, symsyryh, syn, tadylyg, tērpek, tigilek, tiŋ, toglah, togyr, tōj, tūs, tyŋ* and *uzun*.

Note that this does not include *appagas* and *köppeges* in 2.9.3 above.

In five words, the first vowel of the stem is long: *čōn* ‘fat’, *hū* ‘pale, white’, *nā* ‘new’, *tērpek* ‘circle, round’ and *tōj* ‘similar, like’. In all it has been shortened in the reduplication. See 3.1.20 on shortening of the reduplicated vowel.

2.9.5 Semantics

The reduplicated meaning is always simply the intensification of the base meaning. Only in the case of *hyzyl*, the definition is not strictly compatible between the sources – though not mutually exclusive, either – but the difference is ignorable from the point of view of the present work (‘extremely red, the reddest’ : ‘bright red’).

All words can be considered adjectives or, rarely, adverbs, with only occasional additional substantival meanings.

2.10 Kirghiz

C-type reduplications in Kirghiz are very numerous and very uniform. In just one single case, can the closing consonant be other than *p*. It is with the word *ak* ‘white’, which, beside *apak*, may also be reduplicated to *appak*.

¹⁶ An arrow with a hollow bone bulb attached below the head, which emits a whistle in flight.

ak is generally rather productive, and the eventual base for three out of five multiple intensifications in Kirghiz. This is a relatively high number, and it includes such formations as *ap.ap.ak.aj* where double reduplication has been combined with diminutive to create even more vivid semantics. See 2.10.4.

Reduplications of non-adjectives are relatively rare and, in most cases, quite straightforward. However, Kirghiz has the only reduplicated pronoun attested in the present work. See 2.10.5.

2.10.1 Sources

The main source of the material is KirgRS which contains ca. 40 000 entries. Some attestations have also been found in Radloff 1883: 280, Ščerbak 1977: 120, and in the grammars listed below.

Grammatical descriptions have been extracted from Sabdan uulu / Barmanov 1933: 50f, Wurm 1949: 107, Hebert/Poppe 1963: 27, Imart 1981: §1109f, Abduldaev et al. 1987: 169, and Kasapoğlu Çengel 2005: 188.

Sabdan uulu / Batmanov 1933: 50f give a brief description of the process, complete with a few examples, only mention *p* as the closing consonant, and note that adjectives in *-lū* and *-syz* do not form reduplications.

Wurm 1949: 107 is even more brief as he only gives the basic description of the process and one example. The closing consonant *p* is mentioned as the only one possible.

In Hebert/Poppe 1963: 27, the description is in fact one sentence: “Adjectives differ from true nouns in that the former can be reduplicated”, followed by two examples. So worded, it suggests that *all* Kirghiz adjectives can be reduplicated. I believe that it is merely an unhappy wording.

Imart 1981: §1109f states clearly that only certain adjectives can be reduplicated, and notes that the reduplicated syllable is closed by *p*. Next, he offers the following remark:

Noter parallèlement à la réduction des voyelles longues un cas de gémination consonantique expressive :

ак < апак ~ аппак ~ апаппак extrêmement blanc

After it, an apparent support is expressed for M.L. Bazin’s proposition (the exact source is not specified) to equal reduplication with the Turkish construction of the *hava güzel mi güzel* type. It is accompanied by the observation that the initial *m-* of *mi* alternates with *b* and *p*, which would have yielded **ak + py + ak > appak*.

Altogether, Imart’s stance is somewhat indefinite. How his diagnosis of *gémination expressive* for *p* in *(ap)appak* can be made compatible with the scheme inspired by Bazin’s idea, is not conspicuous in itself, and not explained, either.

A view similar to Bazin's is expressed by Abduldaev et al. 1987: 169, where reduplications are interpreted as a juxtaposition of a truncated form of an adjective with its full form, and glued together by a semantically empty structural element *p*. Also, it is noted that stress is initial in them.

Finally, Kasapoğlu Çengel 2005: 188 limits herself to stating that reduplications are formed by appending *p* to the initial syllable.

Transcription:

Imart 1981: *i* → *y*

Kasapoğlu Çengel 2005: *c* → *ç* | *ι* → *y* | *ş* → *ş*,

Radloff 1883: *q* → *a* | *ğ* → *ç* | *o* → *u* | *q* → *k*,

Sabdān uulu / Batmanov 1933: *o* → *ö* | *q* → *k* | *b* → *y*,

Wurm 1949: *i* → *y* | *l* → *l* | *q* → *k*,

Other: *VV* → *V̄* | *ɥ* → *ɥ* | *o* → *ö* | *ɣ* → *ü* | *ɣ* → *ç*.

2.10.2 Standard cases

ačū '1. bitter; 2. pungent' ◇ **apačū** (KirgRS)

ačyk '1. open; 2. clear, obvious' ◇ **apačyk** (KirgRS)

ak 'white' ◇ **apak** (KirgRS, Imart 1981: §1110) ◇ **appak** (KirgRS, Imart 1981: §1110)

See also *apapakaj* and *apappak* in 2.10.3.

alys 'far' ◇ **apalys** (KirgRS)

apakaj 'very white' ◇ **apapakaj** (KirgRS) 'very nice, very good'

See 2.10.4 below.

appak 'very white' ◇ **apappak** (KirgRS, Imart 1981: §1110)

See 2.10.4 below.

balpak 'fat and squat of a person' ◇ **bapbalpak** (KirgRS)

beker 'vain' ◇ **bepbeker** (KirgRS: only attested in *maktanganyŋ* ~ 'you boast completely in vain')

bijik 'high, tall' ◇ **bipbijik** (KirgRS)

birdej 'same, identical' ◇ **bipbirdej** (KirgRS)

boz 'light grey, ash grey, earthy grey' ◇ **bopboz** (KirgRS)

bütün 'whole' ◇ **büpbütün** (KirgRS)

čak 'fit, just' ◇ **čapčak** 'exactly, just right'

čoŋ 'big, huge' ◇ **čopčoŋ** (KirgRS, Imart 1981: §1110)

čunkurčak 'depression, basin' ◇ **čupčuŋkurčak** (KirgRS: only attested in ~ *kyl-* 'to shatter to the ground')

dajar 'ready, prepared, preparation' ◇ **dapdajar** 'fully ready' (KirgRS)

dajyn '(well-)known, certain' ◇ **dapdajyn** 'accurately known' (KirgRS)

dāna '1. (well-)known; 2. knowing, wise' ◇ **dapdāna** 'clearly, distinctly' (KirgRS)

dardaj 'strapper, whopper, hefty' ◇ **dapdardaj** (KirgRS)

döngölök ‘cart-wheel’ ◇ **döpdöngölök** ‘completely round’ (KirgRS)

See 2.10.5 below.

ēn ‘desert, secluded, lorn’ ◇ **epēn** (KirgRS)

ičke ‘thin, slim’ ◇ **ipičke** (KirgRS)

kačan ‘when?’ ◇ **kapkačan** ‘(very) long ago’ (KirgRS)

See *kapkačan* in 2.10.3 below.

kačanky ‘1. related to what time? 2. old, past, earlier’ ◇ **kapkačanky** ‘(very) old, past’ (KirgRS)

See *kapkačan* in 2.10.3 below.

kajdagy ‘located at any place’ ◇ **kapkajdagy** ‘located god knows where; god knows what sort of’ (KirgRS)

kara ‘black’ ◇ **kapkara** ‘1. *intens*; 2. species of golden eagle (KirgRS s.v. *канкара*)’ (Sabdan uulu / Batmanov 1933: 50, KirgRS, Imart 1981: §1110, Abduldaev et al. 1987: 169)

karangy ‘dark(ness)’ ◇ **kapkarangy** ‘complete darkness’ (KirgRS)

kenj ‘wide, spacious’ ◇ **kepkenj** (KirgRS)

kenen ‘sufficient, ample, generous, spacious’ ◇ **kepkenen** ‘very spacious’ (KirgRS)

kičine ‘very little, very small’ ◇ **kipkičine** ‘very small’ (KirgRS)

See *kipkičinekej* in 2.10.3 below.

kojū ‘thick, dense’ ◇ **kopkojū** (KirgRS)

kök ‘blue’ ◇ **köpkök** (Sabdan uulu / Batmanov 1933: 50f, Hebert/Poppe 1963: 27, Abduldaev et al. 1987: 169, Johanson 1998: 39f, Kasapoğlu Çengel 2005: 188)

kū i.a. ‘1. pale, white; 2. dry’ ◇ **kupkū** (KirgRS: only attested in *öñü* ~ ‘he is very pale’, ~ *karagaj* ‘a completely dried up spruce’)

kurgak ‘dry’ ◇ **kupkurgak** (KirgRS)

kyzyl ‘red’ ◇ **kypkyzyl** (Sabdan uulu / Batmanov 1933: 50, Wurm 1949: 107: <q̄ipq̄iz̄>, Hebert/Poppe 1963: 27, KirgRS)

möldür i.a. ‘clean, clear, transparent’ ◇ **möpmöldür** (KirgRS)

muzdak ‘cold, chilly’ ◇ **mupmuzdak** (Imart 1981: §1110)

okšoš ‘same, similar’ ◇ **opokšoš** (KirgRS)

oñoj ‘easy’ ◇ **opoñoj** (KirgRS)

ōr ‘heavy’ ◇ **opōr** (KirgRS, Ščerbak 1977: 120, Imart 1981: §1110)

sakaldū ‘a bearded man, a man of age’ ◇ **sapsakaldū** (KirgRS: only attested in ~ *bašyñ menen* ‘you and your huge beard’)

sary ‘yellow’ ◇ **sapsary** (Radloff 1883: 280, KirgRS)

semiz ‘fatty, greasy’ ◇ **sepsemiz** (KirgRS)

širin ‘sweet’ ◇ **šipširin** (KirgRS)

sō ‘healthy’ ◇ **sopsō** (KirgRS)

sonun ‘(very) good’ ◇ **sopsonun** (KirgRS)

- sūk** i.a. ‘1. cool, cold; 2. unpleasant, surly; 3. opposed to’ ◇ **supsūk** (Sabdan uulu / Batmanov 1933: 50, KirgRS)
- sulū** ‘beautiful’ ◇ **supsulū** (KirgRS, Imart 1981: §1110)
- sur** ‘grey’ ◇ **supsur** (KirgRS)
- syjda** ‘smooth’ ◇ **sypsyjda** (KirgRS)
- sylyk** ‘delicate, polite, suave’ ◇ **sypsylyk** (KirgRS)
- synyk** i.a. ‘delicate, polite, suave’ ◇ **sypsinyk** (KirgRS)
- tak** i.a. ‘exactly, precisely’ ◇ **taptak** ‘1. *intens.*; 2. a sharp jump, sharp movement’ (KirgRS)
- takyr** ‘naked, bare’ ◇ **taptakyr** ‘smooth, even’ (KirgRS)
- takyrčak** ‘naked’ ◇ **taptakyrčak** ‘entirely, clean’ (KirgRS)
- tatynakaj** ‘nice, kind’ ◇ **taptatynakaj** (KirgRS: only attested in ~ *kyz* ‘a very nice, very kind girl’)
- taza** ‘clean’ ◇ **taptaza** (Sabdan uulu / Batmanov 1933: 50, KirgRS)
- tegerek** ‘round’ ◇ **teptegerek** (KirgRS)
- tegiz** ‘smooth, even’ ◇ **teptegiz** (KirgRS)
- teŋ** ‘equal’ ◇ **tepten** (KirgRS)
- tik** ‘vertical, steep’ ◇ **tiptik** (KirgRS)
- tirū** ‘live’ ◇ **tiptirū** (KirgRS)
- tok** ‘full; satisfying’ ◇ **toptok** (Sabdan uulu / Batmanov 1933: 50)
- toltura** ‘completely full’ ◇ **toptoltura** (KirgRS)
- toluk** ‘full’ ◇ **toptoluk** (KirgRS)
- tügöl** ‘all, completely, entirely’ ◇ **tüptügöl** (KirgRS)
- tūra** ‘rightly, correctly’ ◇ **tuptūra** (KirgRS)
- tüz** i.a. ‘flat, smooth, straight; plain, flatland’ ◇ **tüptüz** (KirgRS, Kasapoğlu Çengel 2005: 188)
- tyjpyl** ‘entirely, to a man’ ◇ **typtyjpyl** ‘1. entirely, clean; 2. completely empty’ (KirgRS)
- tyrmaktaj** ‘with nails, with claws’ ◇ **typtyrmaktaj** (KirgRS: only attested in ~ *bala* ‘a tiny child’)
- uzun** ‘long’ ◇ **upuzun** (Imart 1981: §1110)
- ynak** i.a. ‘a close friend’ ◇ **ypynak** (KirgRS)
See 2.10.5 below.
- yras** ‘truly, really, as should be, good, skillfully’ ◇ **ypyras** ‘pure truth, very accurately’ (KirgRS)
- žaj** i.a. ‘calmly, slowly, quietly’ ◇ **žapžaj** (KirgRS)
- žakšy** ‘good’ ◇ **žapžakšy** (KirgRS)
- žalgyz** ‘only, single, lonely’ ◇ **žapžalgyz** (KirgRS)
- žaman** ‘bad’ ◇ **žapžaman** (KirgRS)
- žanaša** ‘in file, cheek by jowl’ ◇ **žapžanaša** ‘very close, by the side’ (KirgRS)
- žany** ‘new’ ◇ **žapžany** (KirgRS, Kasapoğlu Çengel 2005: 188)

žaš ‘young’ ◊ **žapžaš** (KirgRS, Imart 1981: §1110)

žašyl ‘green’ ◊ **žapžašyl** ‘light green (KirgRS), very green (Kasapoğlu Çengel 2005: 188)’

žeke ‘separate, singular, only’ ◊ **žepžeke** ‘completely alone’ (KirgRS)

ženil ‘light, mild’ ◊ **žepženil** (KirgRS, Kasapoğlu Çengel 2005: 188)

žōn ‘fatty, greasy’ ◊ **žopžōn** (Imart 1981: §1109)

žuka ‘thin’ ◊ **župžuka** (KirgRS)

žumšak ‘soft’ ◊ **župžumšak** (Radloff 1883: 280, KirgRS)

žumuru ‘round’ ◊ **župžumuru** (KirgRS)

žyłaṇač ‘naked’ ◊ **žypžyłaṇač** (KirgRS)

See also 3.1.10 on other intensifications of the word.

žylas i.a. ‘gone, demolished’ ◊ **žypžylas** ‘gone without a trace’ (KirgRS)

žylma ‘1. smooth, even; 2. well-mannered, polite, diplomatic’ ◊ **žypžylma** ‘1. very smooth, slippery; 2. well-mannered, diplomatic, sly, crafty fellow’ (KirgRS)

žylmakaj ‘1. smooth, slippery; 2. entirely, clean e.g. *rob*’ ◊ **žypžylmakaj** (KirgRS)

2.10.3 Special cases

apapakaj ‘1. very white; 2. *child.* very nice, very good’ (KirgRS) ◊ **apakaj** ‘very white’

This word is a triple intensification: *ap-* redup. + *ap-* redup. + *ak* + *aj* dimin. Most probably, the order of formation was *ak* → *apak* → *apakaj* → *apapakaj* as all these forms are attested, and **akaj* does not seem to be so.

The second meaning is marked as *демское* in KirgRS which explains both the exaggerated emphasis and nothing about the productivity of reduplication in Kirghiz or the actual emotional load of the formation.

See *apappak* below, and 3.1.13 for other examples of multiple intensification, and 3.1.2 for related examples of alternative reduplications with different meanings.

apappak (KirgRS, Imart 1981: §1110) ◊ **ak** ‘white’

This word is a double reduplication. Apparently, *appak* was no longer felt as sufficiently intense by some speakers. See 3.1.13 for other examples of multiple intensification, and also *apapakaj* above and *kipkičinekej* below.

KirgRS confirms the interpretation as double reduplication but decomposes the word to *ap-ap-ak*, rather than the more manifest **ap-app-ak*. This is surprising as both *apak* and *appak* exist in Kirghiz. Moreover, the doubling of *p* cannot be easily and surely explained on a purely Kirghiz ground. Perhaps the two *ap-*s are to be read symbolically for ‘reduplication’, and the notation is meant to explain the morphological build rather than phonetic intricacies – as a definition in a dictionary might be expected to.

***čypčyrǵa** ‘all, everything’ (KirgRS) ◊ ? **čyrǵa** ‘lure; bait for Golden Eagle’

In KirgRS, *čyp* is defined as ‘intensifier for words beginning with *čy-*’ (s.v. *чын* I), and exemplified by the untranslated phrase *čypčyrǵasyn korotpoj* which refers to the entry for *čyrǵa*. There, *čyrǵa* is translated as ‘lure; bait for Golden Eagle’, and among the examples, two phrases with the form in *čyp-* are listed, *čyp čyrǵasy* (or *čypčyrǵasy*) *koroboj, bāry esimde* ‘I have every last detail in my memory’, and *čyp čyrǵasyn korotpoj saktap žürdüm* ‘I kept everything intact and untouched’.

The semantic evolution suggested by this attestation is not impossible, but also not trivial. The attestation itself is also not entirely clear. Here, the word will be considered a special case, and excluded from further considerations.

kapkačan ‘(very) long ago’ (KirgRS) ◊ **kačan** ‘when?’

The semantic shift ‘when?’ → ‘(very) long ago’ might seem unusual at first, but it is not necessarily so. The primary meaning of reduplication is intensification. Intensified ‘when?’ might be rendered as ‘when, oh when?’ in English, and that can quite naturally be understood as ‘(very) long ago’ – especially if used as an introduction to a tale or in a similar context.

Very similar semantics can be found in an adjectival derivative from *kačan*, *kačanky*, where the two meanings are already present in the base: ‘1. related to what time? 2. old, past, earlier’, and it is only the etymologically later one that is reduplicated.

See also Kirg. *alda kačan* ‘1. long ago; 2. god knows when’ (KirgRS).

To be noted about this example is that *kačan* is primarily a pronoun, which makes it the only case of a reduplicated pronoun in the present work.

kapkačanky see *kapkačan* above.

kepkenedej ‘very small, tiny’ (KirgRS)

The base **kenedej* does not seem to be attested, and the word is not clear. Maybe related to *kipkičinekej* ‘very small’ below?

kipkičinekej ‘very small’ (KirgRS)

The base **kičinekej* does not seem to be attested. However, *kičine* is, therefore the process was here apparently the same as in *apakaj* (see *apapakaj* above) and Oir. *apač|š* (see 2.12.3), i.e. it was the diminutive that was derived from a reduplicated base, rather than the other way round. However, see also Kklp. *kipkiškene, kipkiškentaj*, and Kzk. *kipkiškentaj*, all ‘very: small, little’ in 2.7.2 and 2.8.2, respectively.

kyp žyłaṅač (KirgRS) ◊ **žyłaṅač** ‘naked’

Technically, this is not a reduplication at all. The word is, nevertheless, interesting because it shows how the morphological awareness of reduplication vanishes, and how the reduplicated anlauts are being promoted to independent words, if with an as yet somewhat unspecified meaning. This particular form, in all probability, is derived from *kypkyzyl žyłaṅač* ‘red-naked = stark-naked’ (KirgRS s.v. *кыпкызыл*).

See 3.1.10 for other cases of emancipation of reduplicated anlauts.

taptatynakaj *‘very nice, very kind’ (KirgRS: only attested in \sim *kyz* ‘a very nice, very kind girl’) \diamond **tatynakaj** ‘nice, kind’

This word is only unusual in that it is a double intensification, a combination of reduplication with a diminutive. Apparently, the amplification of the semantics ensured by the latter must have waned away over time as KirgRS gives the meaning of *tatyna* as ‘*то же, что татынакай*’.

See also 3.1.13 for other examples of multiple intensification.

tymtyrakaj ‘in disarray, every which way’ (KirgRS)

In KirgRS, the base **tyrakaj* only refers to *tymtyrakaj*. The word is certainly related to the equally unclear Kzk. *tymtyrakaj* ‘in disarray, disorderly, in all directions’ in 2.8.3.

tymtyrs ‘complete silence’ (KirgRS)

The base *tyrs* seems to only be attested with the meaning ‘click, snap, crack(le), crunch’. The evolution from it to ‘complete silence’ is not trivial, but not impossible.

2.10.4 Structure

Technically, two closing consonants of *C*-type are attested in a total of 92 examples derived from 91 unique bases, in a rather one-sided distribution:

p: 91 examples: *ačū, ačyk, ak, alys, apakaj, appak, balpak, beker, bijik, birdej, boz, bütün, čak, čoң, čuңkurčak, dajar, dajyn, dāna, dardaj, döngölök, ēn, ičke, kačan, kačanky, kajdagy, kara, karaңgy, keң, kenen, kičine, kojū, kök, kū, kurgak, kyzyl, möldür, muzdak, okšoš, oңoj, ōr, sakaldū, sary, semiz, širin, sō, sonun, sūk, sulū, sur, syjda, sylyk, synyk, tak, takyr, takyrčak, tatynakaj, taza, tegerek, tegiz, tej, tik, tirū, tok, toltura, toluk, tügöl, tūra, tüz, tyjpyl, tyrmaktaj, uzun, ynak, yras, žaj, žakšy, žalgyz, žaman, žanaša, žaңy, žaš, žašyl, žeke, žeңil, žōn, žuka, žumšak, žumuru, žyлаңаč, žylas, žylma, žylmakaj, and*

pp: 1 example: *ak*.

Only one word has more than one closer possible. It is *ak* ‘white’, and its two closers are *p* and *pp*; see also *apapakaj* and *apappak* in 2.10.3 above.

In eight words the first vowel is long: *dāna* ‘(well-)known, ...’, *ēn* ‘desert, ...’, *kū* ‘1. pale, white; 2. daffodil’, *ōr* ‘heavy’, *sō* ‘healthy’, *sūk* ‘cold, ...’, *tūra* ‘rightly, ...’, *žōn* ‘fatty, ...’. In all cases, the reduplicated vowel has been shortened. See 3.1.20 on shortening of the reduplicated vowel.

The word *kačan* ‘when?’ is the only case of a reduplicated pronoun in the present work. The thought of intensifying an interrogative pronoun might seem unusual at first to a speaker of English – incidentally, much less so to a speaker of Polish or other Slavonic language, see e.g. *któ.ż* ‘who, oh who?’ &c. – but the case is actually quite straightforward. See 2.10.3 above.

The base *ak* ‘white’ is unusually productive in Kirghiz, with as many as five different intensifications: *apak*, *apakaj*, *aparakaj*, *appak*, and *apappak*. The middle one is particularly noteworthy, being the only example in the present work of a combination of double reduplication and a diminutive. See *aparakaj* and *apappak* in 2.10.3 above.

There are two more words created by combining reduplication and diminutive: *kipkičinekej* and *taptatynakaj*, see 2.10.3 for both. The total number of multiple intensifications in Kirghiz amounts thus to five, which is relatively many. Notably, one of them is a triple intensification: *ap.ap.ak.aj*.

2.10.5 Semantics

In most examples, the reduplicated meaning is apparently the same as the base meaning or its simple intensification.

The only stronger deviation can be observed in *apakaj* and *kačan*. For the latter, see 2.10.3 above. As for the former, apart from the expected ‘very white’, its reduplication also has the meaning ‘very nice, very good’, which can be neither found in the base word nor in *ak*. Among the meanings of *ak*, there are also such components as ‘clean’ and ‘innocent’, and they are perhaps the closest match to be found here. See 3.1.24 on reduplications of ‘white’.

With the notable exception of *kačan* ‘when?’ (see 2.10.3 above), most examples are also quite clear adjectives. Only the following five words have a less determined character:

Perhaps the most interesting is the case of *ynak* ‘1. чистый, без примесы; 2. близкий друг’, where apparently only the second, substantival, meaning is present in the reduplication while the first, adjectival one, is lost: *урynak* ‘очень близкий друг’. This goes somewhat against the essence of reduplication which generally is adjective intensification.

In the case of *karągy*, an insufficiently exhaustive definition might be suspected. A word translated as ‘тьма-тьмушая; тьма непроглядная’ (KirgRS) is likely to also be used as an adjective in a Turkic language, even if the Russian definition does not quite reflect this.

The remaining three cases appear to be simpler: *čurkurčak* ‘углубление, впадина, котловина’, *döngölök* ‘1. колесо (телеги); 2. шутол[ивое слово, выражение] автомобиль’, and *sakaldū* ‘бородач, пожилой мужчина, мужчина почтенного возраста’. In each of them, there is a quite straightforward link between the meaning of the base and that of the reduplication. Note that the first and the last of these three are only attested in what seem to be fixed phrases.

There are five multiple intensifications in Kirghiz, three of them derived from *ak* ‘white’. See 2.10.4 above.

2.11 Kumyk

C-type reduplications are fewer in Kumyk than in the majority of Kipchak languages, but equally uniform.

Three closing consonants are attested. Two of them only have one example each, and one might likely be a loanword (*büsbütün*).

Two points in the build of Kumyk reduplications need to be noted. One is the word *parahat* ‘calm’, one of the generally very few beginning with *p*, and reduplicating interestingly to *papparahat*. The other point is the existence of three reduplications which are apparently pronounced differently than their origin and spelling would suggest. See 2.11.4.

As far as semantics is concerned, Kumyk reduplications are quite standard. See 2.11.5.

2.11.1 Sources

The main source of the material is KmkRS which contains ca. 13 000 entries. Some attestations have also been found in Ščerbak 1977: 120, Doniyorova 2004: 19, and in the grammars listed below.

For grammatical descriptions, Dmitriev 1940: 71f and, to a certain degree, KmkRS, have been used.

Dmitriev 1940: 71f states that the closing consonant is *p* and, more surprisingly, that the majority of Kumyk adjectives can be reduplicated and, therefore, that intensive needs to be recognized “как живую категорию кумыкских (и вообще тюркских) прилагательных”. The general Turkic part of this remark is incorrect, and this casts doubt on the Kumyk bit. The scarcity of examples attested in KmkRS, too, makes this opinion seem exaggerated, but it must be remembered that this dictionary only contains ca. 13 000 entries in total.

KmkRS does not contain a grammatical description. The reduplicated anlauts, however, are usually described as “препозитивная усил[ительная] частица, присоединяемая к словам, начинающимся на ...”, but this description is not consistent: sometimes it limits the scope of the ‘particle’ to ‘adjectives’ or ‘certain adjectives’ only, sometimes to merely one or two concrete examples. In any case, it quite clearly disagrees with Dmitriev’s diagnosis.

The matter cannot be resolved without extensive research directed specifically at Kumyk, which goes beyond the scope of the present work. The material discussed here will be limited to the explicitly confirmed examples.

Transcription:

Dmitriev 1940: 71f: *ə* → ä | *къ* → k | *лъ* → l,

Doniyorova 2008: *къ* → k,

KmkRS, RKmkS: *e-* → je- | *гъ* → g | *къ* → k | *ю* → ü.

2.11.2 Standard cases

ačyk ‘1. open; 2. clear of weather’ ◇ **apačyk** ‘1. completely open; 2. very clear, very bright, very obvious’ (Dmitriev 1940: 72, KmkRS)

ak ‘white’ ◇ **apak** (KmkRS, Ščerbak 1977: 120, Doniyorova 2008: 19¹⁷) ◇ **appak** (Dmitriev 1940: 72, KmkRS)

See 2.11.4 below.

belgili ‘clear, (well-known)’ ◇ **bepbelgili** (Dmitriev 1940: 72)

boš ‘empty’ ◇ **bopboš** (KmkRS)

bütün ‘whole’ (RKmkS s.v. *целый* and others¹⁸) ◇ **büsbütün** ‘1. intens.; 2. very, quite, altogether’ (KmkRS)

See 2.11.4 below.

gavajyn ‘free of charge’ ◇ **gapgavajyn** (KmkRS)

gazir ‘ready’ ◇ **gapgazir** (KmkRS)

gerti ‘true, faithful, accurate’ ◇ **gepgerti** (KmkRS: only attested in *~sin ajtmak* ‘to speak the absolute truth’)

jangy ‘new’ ◇ **japjangy** (KmkRS)

jangyz ‘only, single, lonely’ ◇ **japjangyz** (Dmitriev 1940: 72, KmkRS, Doniyorova 2008: 19)

jašil ‘green’ ◇ **japjašil** (Dmitriev 1940: 72)

jengil ‘light, easy’ ◇ **jepjengil** (KmkRS)

kara ‘black’ ◇ **kapkara** (Dmitriev 1940: 71, KmkRS, Doniyorova 2008: 19)

karangy ‘dark(ness)’ ◇ **kapkarangy** (KmkRS)

kuru ‘dry’ ◇ **kupkuru** (KmkRS)

kyzyl ‘red’ ◇ **kypkyzyl** ‘bright red’ (Dmitriev 1940: 72, KmkRS)

parahat ‘calm’ ◇ **papparahat** (KmkRS)

sangyrav ‘deaf’ ◇ **sapsangyrav** (KmkRS)

sari ‘yellow’ ◇ **sapsari** (Dmitriev 1940: 72, KmkRS)

See 2.11.4 below.

takyr ‘naked, bare (RKmkS s.v. *голый*)’ ◇ **taptakyr** ‘completely naked, completely bare; with no vegetation at all’ (KmkRS)

taza ‘clean’ ◇ **taptaza** (Dmitriev 1940: 72, KmkRS)

tegiš ‘smooth, even’ ◇ **teptegiš** ‘1. intens.; 2. dog-poor’ (KmkRS)

tuvra ‘straight, direct’ ◇ **tuptuora** (Dmitriev 1940: 72)

tüz ‘1. accurate; 2. straight, direct’ ◇ **tüptüz** (KmkRS: only attested in *~ ok eki* ‘exactly two’, *~ üstüne ürümek* ‘to go directly’)

¹⁷ In Doniyorova 2008: 19, *apak* is spelt ⟨ап-ак⟩ rather than *⟨ап-акъ⟩. It is most probably a misprint.

¹⁸ In KmkRS, *бютюн* only references to *бютюнлей* without any commentary.

2.11.3 Special cases

bešbeter ‘even worse’ (KmkRS) ◊ **beter** ‘1. more; 2. worse’ (RKmkS s.v. *наче* (in *ondan da* ~ ‘even more so’), and *сколько*, KmkRS: only attested in ~*inden allag saklasyn* ‘god forbid it gets worse’, KmkRFS s.v. *айтгъандан...*, *бетерин...*, *бесчестье...*, and *сверх всякого...*)

This word has a definite air of a loanword in Kumyk. The attestations of the base *beter* in the dictionaries are somewhat circumstantial. The closing consonant *p* is overwhelmingly the most frequent in Kumyk, to the point, in fact, where any other must immediately seem suspicious. On the other hand, *š*, among all the Turkic languages, only occurs as a closing consonant with *beter* in Azeri, Kumyk and Turkish. The word is eventually of Persian origin (بَدْتَار *badtar* ‘worse’). From the point of view of geography, it would have been expected to pass through Azeri – where not only the possible closing consonants are much more diversified (see 2.2.4 above), but also the meanings of Az. *betār* match rather closely those attested for Kumyk: ‘1. worse; 2. more’, and dialectally, ‘even worse’ (AzRS). See also Trkm. *besbeter* ‘1. more so, very; 2. worst of all’ in 2.17.2.

It is because of this unusual combination of premises, that this reduplication will be treated here as a loanword and excluded from further considerations, despite its base being attested independently – which is otherwise considered sufficient to treat a word as a native formation.

A parallel exception, also concerning *beter*, has been made in Karakalpak, see 2.7.3.

zepzemre ‘very wet’ (Dmitriev 1940: 72)

The base **zemre* is missing from both KmkRS and RKmkS. It seems that only the former contains a similar shape, namely *zemire* ‘arch. ritual song to invoke rain in a dry summer’. Also MifyNM attests *земурé* as the name of a (*the?*) goddess of fertility in pre-Islamic Kumyk mythology, whose cult has been preserved in the rainmaking ritual. It is obvious that there is a semantic relation between the two words, but less clear what the nature of this relation exactly is.

2.11.4 Structure

Three closing consonants of *C*-type are attested in a total of 25 examples derived from 24 unique stems, in a distribution so even, that it is rather suggestive of borrowing in the two outstanding cases:

p: 23 examples: *ačyk*, *ak*, *belgili*, *boš*, *gavajyn*, *gazir*, *gerti*, *jangy*, *jangyz*, *jašil*, *jengil*, *kara*, *karangy*, *kuru*, *kyzyl*, *parahat*, *sangyрав*, *sari*, *takyr*, *taza*, *tegiš*, *tuвra*, and *tüz*,

pp: 1 example: *ak*, and

s: 1 example: *bütün*.

Only *ak* ‘white’ has more than one closer possible: *p* and *pp*.

The use of *s* as the closing consonant in *büsbütün* is surprising against the remaining examples. It is quite probable that it is, like *bešbeter* (see 2.11.3 above), a loanword from perhaps Azeri, but in the case of *bütün* the base is attested independently in Kumyk, and there are no more arguments than the uniqueness of the closing consonant, to demonstrate its possible foreign origin. See also 3.2.5 on the distribution of closing consonants across the Turkic languages.

Kumyk is one of just three languages which have a reduplication derived from a word beginning with *p*-, and the only one which also employs *p* as a closing consonant in this case. See 3.1.6 on the closing consonant being identical to *C*₂.

Three of the examples given by Dmitriev 1940: 72 are annotated as being pronounced differently than spelt. They are: *apak*, pronounced [appak] («аппакъ»), *kappa* pronounced [kappagart], and *sapsari* [säpsäri] («сәп-сәри»).

2.11.5 Semantics

In the great majority of cases, the reduplicated meaning is a simple intensification of the base meaning, or apparently the same. The latter should probably be attributed to dictionary definitions not being sufficiently exhaustive.

The only slight diversions are provided by *ačyk* and *bütün*, where the meaning of the reduplication has evolved one little step further away from the base (respectively, ‘1. open; 2. clear of weather’ → ‘1. completely open; 2. very clear, very bright, very obvious’, and ‘whole’ → ‘1. *intens*; 2. very, quite, altogether’), and by *tegiš*, where it has developed an additional, figurative meaning (‘smooth, even’ → ‘1. *intens*.; 2. dog-poor’).

2.12 Oirot

Relatively few *C*-type reduplications are attested in Oirot, pronouncedly less than in Khakas and Tuvian. The majority are standard and uniform, but *ak* ‘white’ and *kök* ‘blue’ have evolved into unusually large families.

Technically, two closing consonants are attested: *p* in 16 examples, and *pp* in one (*appak*). See 2.12.4.

Reduplications of non-adjectives almost do not exist. Semantic evolutions going beyond simple intensification are just slightly more numerous. See 2.12.5.

Ak and *kök* have proven unusually productive as they have sprouted a total of 15 or 16 derivatives in Oirot and Shor combined, plus two in Khakas. Also interesting is what appears to be emphatic lengthening of *k* in *čipčikke* ‘exactly, precisely’. See 2.12.3.

2.12.1 Sources

The main sources of material are AltTS, which contains more than 13 000 entries, and Dyrenkova 1940: 77. Some attestation have also been found in Baskakov 1972 and 1985, Li et al. 2007, Ramstedt 1952: 249f, Ščerbak 1977: 120, Schönig 1998b: 408, and Serebrennikov/Gadžieva 1986: 112.

For grammatical descriptions, Dyrenkova 1940, Baskakov 1985: 30, and AltTS have been used.

Dyrenkova 1940: 77 gives a slightly more elaborate description of the process which, however, only informs that the closing consonant is “один из губных”. This is confusing, given that all Oïrot examples have their reduplicated anlauts closed by either *p* or *pp*.

Baskakov 1985: 30 limits himself to the information that the superlative is formed by reduplication of the first syllable.

AltTS does not contain a chapter on grammar but some information can be deduced from how *ap* and *žap* are defined, namely “Pekiştirme sıfatlarında p sesi ile yapılan ön ek” and “Pekiştirme sıfatı yapmada kullanılan p sesi ile ön ek”, respectively.

Transcription:

| | |
|---|--|
| AltTS: $c \rightarrow \check{z} \mid \iota \rightarrow y \mid \tilde{n} \rightarrow \eta \mid \varsigma \rightarrow \check{s} \mid$ | Li et al. 2007: $d' \rightarrow \check{z} \mid i \rightarrow y,$ |
| $y \rightarrow j,$ | Schönig 1998b: $q \rightarrow k,$ |
| Dyrenkova 1940: $\partial b \rightarrow \check{z} \mid VV \rightarrow \bar{V},$ | Serebrennikov/Gadžieva 1987: $\kappa \rightarrow k.$ |

2.12.2 Standard cases

ak ‘white’ \diamond *apak* (Dyrenkova 1940: 77) \diamond *appak* (Ramstedt 1952, Baskakov 1972)

See also *apag* in 2.12.3 below

boro ‘grey’ \diamond *bopboro* (Ščerbak 1977: 120)

čike ‘straight, right, accurate’ \diamond *čipčik(k)e* ‘exactly, precisely’ (Dyrenkova 1940: 77)

See *čipčikke* in 2.12.3 below.

čokur ‘motley’ \diamond *čopčokur* (Dyrenkova 1940: 77, Ščerbak 1977: 120)

kara ‘(pitch) black’ \diamond *kapkara* (Baskakov 1985: 30, Schönig 1998b: 408, AltTS)

kök ‘blue’ \diamond *köpkök* (Dyrenkova 1940: 77)

kozyr ‘large’ \diamond *kopkozyr* (Dyrenkova 1940: 77)

kyzyl ‘red, crimson’ \diamond *kypkyzyl* ‘1. carmine, crimson; 2. bright red’¹⁹ (Dyrenkova 1940: 77, Serebrennikov/Gadžieva 1986: 112, AltTS)

sary ‘yellow’ \diamond *sapsary* (Dyrenkova 1940: 77)

temej ‘to no avail, in vain’ \diamond *teptemey* (AltTS)

¹⁹ The original meanings in AltTS are: *kızıl* ‘1. Kırmızı, kızıl; [...]’ : *kıp-kızıl* ‘kıp kızıl, tamamiyle kırmızı [...]’.

ten ‘same, identical’ ◊ **tepten** ‘very straight, very smooth’ (AltTS)

tünej ‘same, similar’ ◊ **tüptünej** ‘1. completely the same; 2. all right, okay’ (AltTS)

uzun ‘high’ ◊ **upuzun** (Dyrenkova 1940: 77)

žany ‘new’ ◊ **žapžany** ‘1. brand-new; 2. news, novelty’ (AltTS, Li et al. 2007)

žžyl ‘green’ ◊ **žapžžyl** (Dyrenkova 1940: 77)

2.12.3 Special cases

In Oiro, Khakas and Shor (see 2.14.3), the reduplications of *ak* ‘white’ and *kök* ‘blue’ have evolved into entire families of 17 or 18 forms in total (Oir. *apagaš* is not clear). All derive eventually from a composition of the reduplicated form with a diminutive suffix, and further phonetic simplifications or semantic amplification.

Discussion and the appropriate schemes are given s.v. Khak. *appagas* and *köppeges*, Oir. *apagaš* and *köpögöš*, and Shor *apagaš* and *köpegeš* (respectively, 2.9.3, here, and 2.14.3). The recommended entry to start with is Oir. *apagaš*, and later *köpögöš*.

See also 3.1.11 for other big families of related reduplications.

akpāš (Dyrenkova 1940: 77) ◊ **ak** ‘white’

Dyrenkova 1940: 77 seems to interpret this form as a contraction of *ap* + *ak* + *-aš* (see *apagaš* below). Within this position, it appears that the only way to understand the *k* in *akpāš* is through dissimilation of **pp* in **app.ak.aš*, or else there is no reason for the *ā* in the second syllable to be long.

An alternative solution would be a composition of **ak* + *apagaš* (< *ap.ak.aš*). The loss of the second *a* in **akapagaš* appears to be quite natural within Turkic phonotactics (see e.g. Ölmez 2011: 402f for a similar example, KB *adakšu* : Uigh. *adkašu* ‘together’). See 3.1.16 for parallel constructions with the base prepended again to the reduplication.

In theory, *akpāš* can also be connected with *apaš* below: *apagaš* > **apāš* > *apaš*, **akapāš* > *akpāš*. However, **akapāš* gives less ground for a contraction such as in *adakšu*, and so this possibility will be dismissed as being less plausible.

See *apagaš* below.

apač|š (Baskakov 1985: 30) ◊ **ak** ‘white’

According to Baskakov 1985: 30, *apaš* is a contraction of **ap.ak.aš*. If so, *apač* should be a contraction of **ap.ak.ač*. Both, *-ač* and *-aš* are diminutive suffixes (see Baskakov 1985: 27). Neither **ag|kač* nor **ag|kaš* seem to be attested, which implies that the suffixes have been added to the already reduplicated form.

See *apagaš* below.

apag (Baskakov 1972) ◊ **ak** ‘white’

This form is not clear. **ag* ‘white’ appears to be missing from Oiro. If not a misprint, perhaps a secondary shortening of *apagaš* (see below)?

apagaš ‘1. *intens.* (Dyrenkova 1940: 77, AltTS); 2. rabbit (AltTS)’ \diamond **ak** ‘white’

Dyrenkova 1940: 77 derives this form from *ap-* + *ak* + *-aš*, which seems very plausible. The suffix *-aš* is diminutive (see Baskakov 1985: 27), and it must have been added as the last component because **ag|kaš* does not seem to be attested; see 3.1.13 for parallel examples.

There are five similar forms based on *ak*, five more based on *kök* ‘blue’ (see *köpögöš* below), and in addition, two in Khakas (see 2.9.3), and five in Shor (2.14.3). Those with a double *pp* present the greatest difficulty (in Oïrot, these are *appāš* and *köppōš*). The *ak* side of the Oïrot family can be presented as in fig. 2.3.

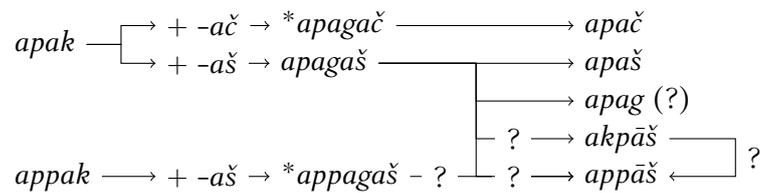


Figure 2.3: Reduplications of Oïr. *ak* ‘white’ &c.

See *köpögöš* and also Khak. *appagas* and *köppeges* in 2.9.3, and Shor *apagaš* and *köpegeš* in 2.14.3.

appāš (Dyrenkova 1940: 77) \diamond **ak** ‘white’

Dyrenkova 1940: 77 appears to interpret this form as a composition of *ap-* + *ak* + *-aš*, which does not explain the double *pp*.

The form *appak* exists in Oïrot (see 2.12.2), most probably continuing an ancient emphatic lengthening in *apak* (see 3.1.12 on the phenomenon). The present word, then, should perhaps be understood as a contraction of **appak.aš*.

However, at least two other interpretations are also possible, yielding in total three versions:

- contraction of **appak.aš*,
- contraction of **ap.apagaš*, and
- simplification of **-kp-* in **akpagaš* < *ak* + *apagaš*.

As for the last two options, both double reduplication and prepending the base to the reduplication, can be observed in various Turkic languages as methods of strengthening the intensification; see 3.1.9 and 3.1.16, respectively.

Double reduplications do not seem to be attested for Oïrot, but one is present in Shor (*apapagaš*, see 2.14.3). Prepending of the base does seem to be attested in Oïrot, in *akpāš* above. Additionally, support for the second option could be sought in *köppōš* below but that word itself requires a parallel example, so this would create a vicious circle.

In lack of arguments that could lend support to the two alternative options, the simplest solution must be tentatively accepted, and that is the first option, **appak.aš* > *appāš*.

See *apagaš* above.

berbek ‘fat’ ◊ **bek** ‘healthy’ (AltTS)

The base word appears to be native: cf. Tuv. *be^δk* ‘strong, sturdy’ &c. = general Tkc. *bek* id. &c. (Tatarincev 2000–08) ~ **berk* id. = Tksh. *berk* ~ *pek* ~ dial. *perk* id. = Tat. dial. *bek* id. = Kklp., Kmk., Nog. *bek* ‘very’ &c. (Stachowski [in print] s.v. *berk*). Also the Oirots meanings ‘strong, sturdy’, ‘healthy’, and ‘fat’ seem to be conceivably close to one another.

The use of *r* as a closing consonant, however, is absolutely extraordinary. This would have been the only South Siberian reduplication with any other closer than *p* or *pp*, which invites the idea that *berbek* is either not a native formation, or actually not a *C*-type reduplication at all.

In Turkish, the variants with and without *r* coexist (*berk* ~ *pek*). Composition of synonyms, such as in Tksh. *pek çok* lit. ‘much much’ or ‘very very’, or *güçlü kuvvetli*, lit. ‘strong powerful’, is no less a popular method of intensification in Turkic than elsewhere. If a hypothetical Oir. **berk* were combined with *bek*, a simplification of the resulting consonant cluster would have to be expected, most probably leading to **berbek*.

In the present work, *berbek* will be assumed to not be a reduplication, and as a consequence, excluded from further considerations.

čipčikke ‘exactly, precisely’ (Dyrenkova 1940: 77) ◊ *čike* ‘straight, right, accurate’

It seems that the base word is only attested in one form, with a single *k*. The reduplication has two variants: *čipčike* and *čipčikke*. The latter results most probably from an expressive lengthening (a cross-linguistic phenomenon, see e.g. Blevins 2004: 174, and *appāš* below), which makes it one of the very few multiple intensifications of this kind (see 3.1.12 for more examples).

kökpögöš (AltTS) ◊ **kök** ‘blue’

Despite the superficial similarity to *akpāš* above, here a composition of *kök* + **köp.kög.öš* does not at all appear likely. The simplest solution is perhaps to assume a metathesis in **köpkögöš*.

See *köpögöš* below.

kökpöš (Dyrenkova 1940: 77) ◊ **kök** ‘blue’

This form is most likely a contraction of *kökpögöš* above, parallel to *köpögöš* > *köpöš* below. See *köpögöš* below.

köpögöš (Dyrenkova 1940: 77) ◊ **kök** ‘blue’

By the same token as *apagaš* above, this form is most probably a composition of *köp-* + *kök* + *-aš* (diminutive, see Baskakov 1985: 27), only here with an additional simplification of **-pk-* to *-p-*. The diminutive must have been added to the already reduplicated form because **kögöš* does not seem to be attested.

There are four similar forms based on *kök*, and also five based on *ak* ‘white’ (see *apagaš* above), two more in Khakas (see 2.9.3), and five in Shor (see 2.14.3). The *kök* side of the Oirat family can be presented as in fig. 2.4.

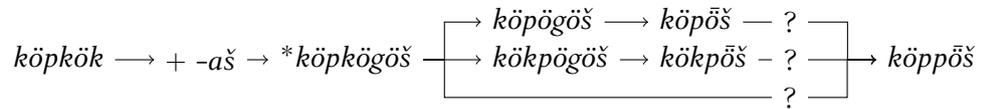


Figure 2.4: Reduplications of Oir. *kök* ‘blue’ &c.

Note that other schemes are also conceivable. This one assumes one simplification of an intervocalic consonant cluster, one metathesis and twice a loss of intervocalic *g*, i.e. more of a frequent phenomenon and less of rarer ones. The proportion would be turned in alternative orderings, which makes them less plausible.

See *apagaš* above and also Khak. *appagas* and *köppeges* in 2.9.3, and Shor *apagaš* and *köpegeš* in 2.14.3.

köpöš (Dyrenkova 1940: 77) \diamond *kök* ‘blue’

This form is most likely a contraction of *köpögöš* above, parallel to *kökpögöš* > *kökpöš* above, although other schemes are also conceivable, see *köppöš* below, and also *köpögöš* above.

köppöš (Dyrenkova 1940: 77) \diamond *kök* ‘blue’

This form is phonetically very similar to *appāš* above, but its history must have been different.

At least three explanations are available:

- simplification of **-pk-* in **köpköš* < **köpkögöš* < *köpkök* + *-aš*,
- simplification of **-kp-* in *kökpöš* < **köpköš* < **köpkögöš*, and
- emphatic lengthening of *p* in *köpöš*.

The first option is simpler and thus appears more plausible than the second, but a metathesis is not at all impossible, as attested by the existence of *kökpögöš* and *kökpöš* (see above). Note that the order of transformations presented here is for illustration only. It is not possible to determine whether the simplification of the consonant cluster would occur before or after the merger of **-ögV-* > *ō*, or whether the second vowel was already **ö* at that point, or still **e* (the primary shape of the suffix is *-eš*), &c.

The last option requires one of the first two to have had occurred first, and the resulting **-pp-* to have been simplified to **-p-* to create a phonetic shape appropriate for it to operate. This is much less improbable than it might seem as in fact *köpöš* already exists, see above.

Theoretical scenarios can be multiplied with relative ease, but at present the actual data seem to be missing with which to select the most probable one of them.

See *köpögöš* above.

2.12.4 Structure

Technically, two closing consonants of *C*-type are attested in a total of 16 examples derived from 15 unique bases, in a very uneven distribution:

p: 15 examples: *ak, boro, čike, čokur, kara, kök, kozyr, kyzyl, sary, temej, teŋ, tünej, uzun, žany, žažyl*, and
pp: 1 example: *ak*.

The only word with more than one closer possible is *ak* with *p* and *pp*.

2.12.5 Semantics

In most cases, the reduplicated meaning is a simple intensification, but exceptions are relatively numerous. They include: *čike* where it is not the first meaning that is intensified in the reduplication ('straight, right, accurate' → 'exactly, precisely'), *kyzyl* where the intensification is not entirely straightforward ('red, crimson' → '1. carmine, crimson; 2. bright red', but see also fn. 19), and *teŋ, tünej* and *žany* where a further semantic evolution can be observed (respectively, 'same, identical' → 'very: straight, smooth', 'same, similar' → '1. completely the same; 2. all right, okay', and 'new' → '1. brand-new; 2. news, novelty').

A non-adjectival or non-adverbial meaning only appears once, in *žapžany* 'news, novelty', a not particularly complex evolution from 'brand-new' ← 'new'.

2.13 Ottoman

With 58 examples, the Ottoman collection is not small, but it is nevertheless quite incomplete. The real number of reduplications might have been even about twice as high. See 2.13.4.

Five closing consonants are attested, which as many as in modern Turkish. There is one difference: Ottoman has one example with a double *pp* (*appak*) and none with *š*; in Turkish it is reversed: *appak* is missing but there is *bešbeter*. Unusually, it seems unclear on which syllable the Ottoman reduplications were stressed. See 2.13.4.

Semantically, Ottoman reduplications are quite standard. Non-trivial shifts and non-adjectives almost do not occur. See 2.13.5.

In *aphāzyr* 'absolutely ready', ⟨biz butoun⟩ 'absolutely all', *öpuzun* 'very long' and *topdolu* 'absolutely full', the reduplicated anlaut does not match the anlaut of the base. Especially the last of them is interesting, even if in no way central to the question of reduplication. Also, in the cases of *māvi* 'blue', *sāfi* 'pure', and *sāry* 'yellow', there is a suspicion that the reduplicated vowels might have retained their length through

reduplication, and in the case of *tamām* ‘proper, ...’ that the reduplicated vowel might have been actually lengthened. See 2.13.3.

2.13.1 Sources

The main sources are: Comidas de Carbognano 1794, von der Berswordt 1839, Németh 1916: 41, and TaS. Some attestations have also been found in Abdülbâkî 1934, Hızır, Mes’ūd, Qorqut, Tuḥfe, Argenti 1533, Ferraguto 1611, Harsány 1672, Mascis 1677, Köroğlu, Vefik Paşa 1890, Jehlitschka 1895: 56f, Sami 1901, Deny 1921: 236, Räsänen 1949: 239, Ramstedt 1952: 249f, Räsänen 1957: 74, Clauson 1972, Šupa/Aleksanovič-Miškinene 1995, and in the grammars listed below.

Mentions in grammars are surprisingly rare: Molino 1641, Seaman 1670, Vaughan 1709, Holderman 1730, Viguier 1790, Romero [18th c.], Jaubert 1833, Schroeder 1835, Davids 1836, Fu’ad-Efendi/Ġävdät-Efendi 1855, Dubeux 1856, Mallouf 1862, and Redhouse 1884 all appear to skip the phenomenon entirely. De Preindl 1790: 19, Comidas de Carbognano 1794, Bereswordt 1839: 11f, Jehlitschka 1895: 56f and Guzev 1979: 46 give some examples but limit the grammatical description to actually less than the minimum, and it is only Meninski 1680: V 39, Németh 1916: 41, Deny 1921: 236 and Özer 2008: 29f who provide slightly more elaborate descriptions, albeit still far from comprehensive.

De Preindl 1790: 19 speaks about the adjectives’ *propres particles* and Bereswordt 1839: 11 about *eigenthümliche Wörter*, so it is understandable that neither touches the question of the closing consonant.

Meninski 1680: V 39 and Jehlitschka 1895: 56f are closer to the truth when they describe the reduplicated anlauts as *particulis certis ad sonum quafi effictis* and *Vorsatzsilben [...], welche mit demselben Konsonanten und Vokale beginnen, sonst aber ziemlich willkürlich sind*, respectively, but they do not take their analyses any further.

Németh 1916: 41, Deny 1921: 236, Guzev 1979: 46 and Özer 2008: 29f recognize the phenomenon as reduplication. Németh 1916: 41 gives a description of the process and lists *m*, *p* (~ *b*), *r* and *s* (~ *z*) as the possible closing consonants but without attempting to formulate rules of distribution. Similarly, if much less clearly, in Deny 1921: 236. Guzev 1979: 46 does not even go this far. Özer 2008: 29f offers a somewhat wordy description according to which, it appears, *b* and *p* are the only possible closing consonants, and the reduplicated anlauts are prefixes.

The diversity of spellings employed in the sources is high, but their consistency and accuracy is considerably less so. The forms used for the main entries are (possibly symbolic) *tertia comparationis* between the different spellings, which in most cases can be found to be the same as the modern Turkish continuants. The original notations are given after the historical sources and skipped after the modern ones, if not necessary.

2.13.2 Standard cases

ačyk ‘open’ ◇ **apačyk** (Ferraguto 1611: 228, 234: ⟨áp accíhc⟩, Meninski 1680: V 39: ⟨ap aciuk⟩, Comidas de Carbognano 1794: 27: ⟨ap acýk⟩, von der Berswordt 1839: 11: ⟨ap ačyk⟩)

ak ‘white’ ◇ **apak** (von der Berswordt 1839: 11: ⟨ap ak⟩, Jehlitschka 1895: 57: ⟨ap aq⟩, Räsänen 1949: 239) ◇ **appak** (Clauson 1972: 3 (14th c.), TaS: ⟨ābīq abbak⟩ (15th c.), ⟨ābīq⟩ (15–16th c.), ⟨ābīq apbak⟩ (16th c.), ⟨ābīq appak⟩ (16th c.))
See *ap(p)ak* in 2.13.3 below.

alaža ‘multicoloured, motley’ ◇ **apalaža** (Qorqut: ⟨abalaca⟩, ⟨apalaca⟩)

ansyz &c. ‘suddenly, unawares’ ◇ **apansyz** (Mes‘ūd 586b: ⟨apanşuz(da)⟩, TaS: ⟨apanşuzda⟩ (15–16th c.); Argenti 1533: 137b: ⟨apanfís⟩, Harsány 1672: 268b: ⟨apanβiz⟩, Mascic 1677: 4, 235b: ⟨abanfisde(n)⟩ ‘accidentally’, *abanfifden* ‘immediately’, Meninski 1680: I 5f, 11, V 39: ⟨ap anfyz⟩ and similar, ⟨ap an-fyzin⟩, de Preindl 1790: 372, 544: ⟨apanfiz⟩, Stachowski S. 2002: ⟨apansez⟩, Schroeder 1835: ⟨apansziz⟩, von der Berswordt 1839: 12: ⟨ap anşiz⟩, Jehlitschka 1895: 57: ⟨apanβýz⟩, Clauson 1972: 3: ⟨apanşizda⟩, *apanşizina*)

ansyzda(n) see *ansyz*

ansyzyn(a) see *ansyz*

ary ‘clean’ ◇ **apary** (Abdülbâkî 1934: 266)

barāber see *berāber*

bejaz ‘white’ ◇ **bembejaz** (Meninski 1680 s.v. *albus* and *candidus*, V 39: ⟨ben-bejāz⟩, Comidas de Carbognano 1794: 26: ⟨ben-bejāz⟩, Šupa/Alekšandrovíč-Miškinene 1995: ⟨benbeyas⟩, von der Berswordt 1839: 11: ⟨bem bejaz⟩, Jehlitschka 1895: 57: ⟨bem bejáz⟩)

belli ‘clear; well-known’ ◇ **besbelli** (Argenti 1533: 149a: ⟨bef belí⟩, ⟨bes belí⟩, Ferraguto 1611: 234: ⟨bés bellí⟩, Meninski 1680: V 172: ⟨bes-belü⟩, Jehlitschka 1895: 56: ⟨bez bellí⟩, Németh 1916: 41: ⟨bez-belli⟩ ‘absolutely certain, absolutely sure’)

berāber ‘together’ ◇ **besberāber** (TaS: ⟨basberaber⟩ (17th c.), ⟨besberaber⟩ (18–19th c.))

See ⟨basberāber⟩ in 2.13.3 below.

boš ‘empty’ ◇ **bomboš** (de Preindl 1790: 19: ⟨bom bošch⟩, Comidas de Carbognano 1794: 27: ⟨bom boš⟩, von der Berswordt 1839: 12: ⟨bon boš⟩, Jehlitschka 1895: 57: ⟨bom boš⟩, Németh 1916: 41: ⟨bom-boš⟩, Deny 1921: 236: ⟨bom boš⟩, Räsänen 1957: 74)

büjük ‘great’ ◇ **büsbüjük** (Meninski 1680: V 39: ⟨büz büjük⟩, Comidas de Carbognano 1794: 27: ⟨böş böjük⟩)

bütün ‘all, whole’ ◇ **büsbütün** (Meninski 1680: V 39: «بوز بتون büz bütün», de Preindl 1790: 19: «biz butoun», Comidas de Carbognano 1794: 27: «بوز بتون büs bütün», von der Berswordt 1839: 11: «بز بتون büz bütün», Jehlitschka 1895: 57: «بستون büsbütün», Németh 1916: 41: «بستون büz-bütün», Deny 1921: 236: «بوس بوتون büsbütün», Ramstedt 1952: 250, Räsänen 1957: 74)
See «biz butoun» in 2.13.3 below.

čabuk ‘quick(ly)’ ◇ **čarčabuk** (Vefik Paşa 1890)

čevre ‘around’ ◇ **čepčevre** (Mes‘ūd: 597a: «çepçevre», Meninski 1680: I 1677: «چپ چوره çep čevre», Özer 2008: 29f)

čürük ‘rotten’ ◇ **čüpčürük** (Comidas de Carbognano 1794: 27: «چوپ چورك čiuپ čüürük»)

čyplak ‘naked’ ◇ **čymčyplak** (Meninski 1680: I 1593: «چم چپلاق čimčiplak») ◇ **čyrčyplak** (Mascic 1677: 224b: «cir ciplach», Meninski 1680: I 1593: «چر چپلاق čir čiplak», Jehlitschka 1895: 57: «چر چپلاق čyr cyplaq», Deny 1921: 236: «چیر چپلاق čír číplaq»)

degirmi ‘round’ ◇ **desdegirmi** (TaS: «دس دکرمی desdeğirmi» (16th c.))

diri ‘alive’ ◇ **dipdiri** (TaS: «dipdirilice» (16th c.), Meninski 1680: II 2020: «دب دری dibdirī», II 2022: «دبدری dibdirī», Comidas de Carbognano 1794: 27: «دب دیری dip diri», von der Berswordt 1839: 12: «دپ دری dip diri»)

doγru ‘straight’ ◇ **dosdoγru** (Meninski 1680: II 2165: «دوز دوغری doz doghry», II 3144, V 39: «دوز طوغرو doz doghru», Comidas de Carbognano 1794: 27: «دوز طوغرو doz doghrū», von der Berswordt 1839: 11: «دوز طوغرو doz doghru»)

dolajinže ‘around’ ◇ **dosdolajinže** (Meninski 1680: II 2165: «دوز دولاینجه doz dolainge»)

dolu ‘full’ ◇ **dopdolu** (Mes‘ūd: 640b: «toptolu», Harsány 1672: 262a: «topdoli», Mascic 1677: 241b: «top dolu», Meninski 1680: II 3085: «طبطلو toptolü», II 3135: «طوپطلو toptolü», de Preindl 1790: 19: «dop dolou», Comidas de Carbognano 1794: 27: «طوپ طولو top tolü», von der Berswordt 1839: 12: «طوپ طولو top dolu», Jehlitschka 1895: 57: «طوپ دولو top dolú», Németh 1916: 41: «طوب ضولو dob-dolu», Özer 2008: 29)

See *topdolu* in 2.13.3 below.

duru ‘clear, limpid’ ◇ **dupduru** (Mes‘ūd: 641a: «tupturu»)

düz ‘smooth, even, straight’ ◇ **dümdüz** (Meninski 1680: II 2020: «دم دوز düm düz», Comidas de Carbognano 1794: 27: «دوم دوز düm düz», von der Berswordt 1839: 12: «دوم دوز düm düz» ‘ganz vereint’, Jehlitschka 1895: 56: «دوم دوز düm düz» ‘completely flat’, Räsänen 1949: 239: «düm-düz» ‘intens.’) ◇ **düpdüz** (TaS: «دپدوز» (13th c.), «دپدن» (14th–16th c.), «دوپدن» (16th c.), «دوپدوز» (16th c.), Hızır: 52, Tuḥfe (1514–15): 52, Köroğlu: 595, Mascic 1677: 150b, 223a, Meninski 1680: II 2020: «دوب دوز dübdüz», II 2022: «دب دن dübdüz», Guzev 1979: 46, Özer 2008: 29)

- eji** ‘good’ ◇ **epeji** (Comidas de Carbognano 1794: 27: «اپ ايو ep ejì», Jehlitschka 1895: 56: «اپ ايی ep ejí») ~ Tksh. *epej(i)(že)* (Hatiboğlu 1973, Müller 2004: 108, BTS) See Tksh. *epej* in 2.16.3.
- gök** ‘blue’ ◇ **gömgök** (TaS: «كومكك» ‘extremely’²⁰ (15th c.), Németh 1916: 41: «كومكك göm-gök»)
- götürü** ‘all, whole’ ◇ **gösgötürü** (TaS: «كوسكوتوری» and similar (16–17/18th c.))
- hāzyr** ‘ready, prepared’ ◇ **aphāzyr** (Meninski 1680: I 6: «اپ حاضر ap hāzyr», de Preindl 1790: 487: «ap hazir») See *aphāzyr* in 2.13.3 below.
- jalyрыз** ‘alone’ ◇ **japjalyрыз** (Comidas de Carbognano 1794: 27: «ياپ يالكر yap ja=lyn-ÿz», von der Berswordt 1839: 11: «ياپ يالكر yap jaleniz», Jehlitschka 1895: 57: «ياپ يالكر yap jalyñýz»)
- jaš** ‘wet’ ◇ **jamjaš** (Meninski 1680: V 39: «يام ياش jam jaš», Comidas de Carbognano 1794: 27: «يام ياش jam jaš», von der Berswordt 1839: 11: «يام ياش jam jaš»)
- jassy** ‘flat’ ◇ **jamjassy** (Comidas de Carbognano 1794: 27: «يام ياصی jam jasy») ◇ **japjassy** (TaS: «ياپ ياصی (14–15th c.)»)
- ješil** ‘green’ ◇ **jemješil** (Meninski 1680: V 39: «يم يشيل jem ješil», Comidas de Carbognano 1794: 27: «يم يشيل jem ješil», von der Berswordt 1839: 11: «يم يشيل jem ješil», Jehlitschka 1895: 57: «يم يشيل jem jezil»)
- kara** ‘black’ ◇ **kapkara** (Meninski 1680: II 3607, V 39: «قپ قره kap kara, kapkara», Comidas de Carbognano 1794: 26: «قپ قره kap karà», von der Berswordt 1839: 11: «قپ قره kap kara», Jehlitschka 1895: 57: «قاپ قره qap qarà», Räsänen 1949: 239, Özer 2008: 29)
- karšy** ‘opposite’ ◇ **kapkaršy** (TaS: «قپقرشو kap karšu» (14th c.), Mes‘ūd 618a: «kapkaršu», *kap karšu*)
- katy** ‘hard, solid’ ◇ **kaskaty** (Argenti 1533: 213a: «chas chattí, kaskatí», Šupa/Aleksandrovič–Miškinene 1995)
- koža** ‘big, huge’ ◇ **koskoža** (Comidas de Carbognano 1794: 27: «كوس قوجه kos kogià», Jehlitschka 1895: 57: «قوس قوجا qoß qočá»)
- kuru** ‘dry’ ◇ **kupkuru** (Meninski 1680: V 39: «قوپ قوری kup kuru|y», Comidas de Carbognano 1794: 27: «قوپ قوری kup kurù», von der Berswordt 1839: 11: «قپ قورو kup kuru», Németh 1916: 41: «قوپقورو kup-kuru»)
- kyrmyzy** ‘red’ ◇ **kypkyrmyzy** (Meninski 1680: II 3607, V 39: «قپ قرمزی kyp kyrmyzy», Comidas de Carbognano 1794: 27: «قپ قرمزی kyp kyrmyzè», von der Berswordt 1839: 11: «قپ قرمزی kyp kyrmyzy», Jehlitschka 1895: 57: «قپ قرمزی qyp qyrmyzý»)
- kyryk** ‘broken’ ◇ **kypkyryk** (Comidas de Carbognano 1794: 27: «قپ قرق kyp kyrÿk»)

²⁰ Erroneously extracted from *gömgök deli* in the example. See also *top* below.

- kyvrak** ‘firm’ ◇ *kyskyorak* (Comidas de Carbognano 1794: 27: «قوراك قس kys kyvrak»)
kyzyl ‘red’ ◇ *kypkyzyl* (Meninski 1680: II 3607, V 39: «قپ قزل kypkyzyl», Guzev 1979: 46: «kybkyzyl», Özer 2008: 29f)
- māvi** ‘blue’ ◇ *māsmāvi* (Comidas de Carbognano 1794: 27: «ماس ماوی mas mavi», von der Berswordt 1839: 11: «ماس ماوی mas mawi», Jehlitschka 1895: 57: «ماس ماوی maß mawí», Sami 1901: «ماس ماوی», Räsänen 1957: 74: «mas-mavi») See *māsmāvi* in 2.13.3 below.
- mor** ‘violet’ ◇ *mosmor* (Comidas de Carbognano 1794: 27: «موس مور mos mor», Jehlitschka 1895: 57: «موس مور moß mor»)
- sāfi** ‘pure’ ◇ *sāmsāfi* (Comidas de Carbognano 1794: 27: «صافی sam safi») See *sāmsāfi* in 2.13.3 below.
- sāry** ‘yellow’ ◇ *sapsāry* (Comidas de Carbognano 1794: 27: «صاپ صاری sap sarý», von der Berswordt 1839: 12: «صاپ صاری sap sary», Jehlitschka 1895: 57: «صاپ صاری ßap ßarý») See *sāpsāry* in 2.13.3 below.
- sijāh** ‘black’ ◇ *simsijāh* (Comidas de Carbognano 1794: 26: «سم سیاه sim sijāh», von der Berswordt 1839: 11: «سم سیاح sim sijah», Jehlitschka 1895: 57: «سم سیاه ßim ßijāh»)
- sivri** ‘pointed’ ◇ *sipsivri* (Comidas de Carbognano 1794: 27: «سپ سوری sip sivri»)
- syklam** ‘wet’ ◇ *syrsyklam* (Jehlitschka 1895: 56: «صر صقلام ßyr ßyqlám», Németh 1916: 41: «صر صقلام syr-syklam»)
- syky** ‘tight, firm’, ◇ *symšky* (Jehlitschka 1895: 57: «صیم صیقی ßym ßyqý»)
- syrlak** ‘smooth, sleek, glazed, shiny’ (DS) ◇ *sypsyrlak* (TaS: «صیپیرلق sypsyrlak» (15th c.))
- tamām** ‘proper, right, just’ ◇ *tāstamām* (Comidas de Carbognano 1794: 27: «تاز تمام tas tamām») See *tāstamām* in 2.13.3 below.
- tatlu** ‘sweet’ ◇ *tamtatlu* (TaS: «طمطلو طم تلو» (15th c.))
- tāze** ‘fresh’ ◇ *tamtāze* (TaS: «تمتازه tamtaze» (16th c.))
- temiz** ‘clean’ ◇ *tertemiz* (Räsänen 1949: 239)
- tolu** see *dolu*
- top** ‘a round body’ ◇ *tortop* (TaS: «طرطوب» ‘derlenip toparlanmış olarak’²¹ (17th c.), Németh 1916: 41: «طور طوپ tor-top» ‘completely round’)
- uzun** ‘long’ ◇ *öpuzun* (Meninski 1680: I 481: «اوپ اوزون öp uzun») ~ *upuzun* (Deny 1921: 236: «اوپ اوزون up uzun», Räsänen 1957: 74) See *öpuzun* in 2.13.3 below.

²¹ Perhaps erroneously extracted from a sentence; see also *gök* above.

2.13.3 Special cases

ap(p)ak (Clouston 1972: 3 (14th c.), TaS: ⟨آبق abbak⟩ (15th c.), ⟨آپق⟩ (15–16th c.), ⟨آپق apbak⟩ (16th c.), ⟨آپق appak⟩ (16th c.), von der Berswordt 1839: 11: ⟨آپ اق ap ak⟩, Jehlitschka 1895: 57: ⟨آپ اق ap aq⟩, Räsänen 1949: 239: ⟨ap-ak⟩) ◊ **ak** ‘white’
This word is attested with two closing consonants: *p* and *pp*. Strangely, double *pp* seems to only occur till the 16th century, followed first by a gap in the 17th–18th century, and then by single *p* which does not appear before the 19th century. Both forms seem to be even missing from Meninski’s large dictionary of 1680.

One is tempted to conclude that reduplication must have been productive in Ottoman at least as late as the 19th century. This is not necessarily wrong; see *topdolu* below, and also 3.1.8 on double *pp*.

aphāzyr (Meninski 1680: I 6: ⟨آپ حاضر ap hāzyr⟩, de Preindl 1790: 487: ⟨ap hazir⟩) ◊ **hazyr** ‘ready, prepared’

The two attestations are clearly missing the expected initial *h-*. It seems that at least Meninski would have been too familiar with Ottoman for this shape to be ascribed to his mishearing.

More probably, it was the general lack of understanding of how exactly reduplication operates, that combined with the natural poor audibility of *h* to cause the modification of the anlaut, faithfully recorded by Meninski and later de Preindl.

It might be that e.g. *autá*, *angisi* or *epsindén* written down by Ferraguto 1611: 216, 223 and 232 for **hafta*, **hangisi* and **hepsinden*, illustrate the same phenomenon rather than the author’s limited knowledge of Ottoman, that could be otherwise suspected.

See also 3.1.19 for other reduplications with anlaut not matching the base.

⟨**basberāber**⟩ (TaS: ⟨بص برابر basberaber⟩ (17th c.)) ◊ **barābar** ~ **berāber** ‘together’

The word stems from Pers. برابر /barābar/ [bārābār], and existed in Ottoman with two vocalizations: ⟨a-a-a⟩ and ⟨e-a-e⟩ (VEWT, where vowel lengths have been omitted). Hence only the forms **basbarābar* and *besberāber* are to be expected in Ottoman.

Here, TaS most probably regarded the spelling with ⟨ص⟩ to be a sign of back vocalization, but it is not clear why the authors should suggest reading the base with *e-ā-e* rather than with *a-ā-a*. Perhaps an influence of modern Turkish phonetics? (See *berāber* in 2.16.2.)

⟨**biz butoun**⟩ (de Preindl 1790: 19) ◊ **bütün** ‘all, whole’

This spelling is most probably a mistake. Note that the base *bütün* itself is spelt ⟨butoun⟩ in this work until p. 475, but ⟨butun⟩ starting with p. 561, and Arabic spelling, unfortunately, is not given at all. Also, one earlier and four later sources, including von der Berswordt which is just 49 years later, all only give shapes with *ü* in the reduplicated anlaut.

čymčyy ‘pure’ (TaS: چىمچىغ *çımçığ*) (16th c.)

The base *čyy does not seem to be attested. Possibly a variant of *čak* ~ *čay* ‘1. only, ...; 2. exact(ly)’ (TaS (13–19th c.))?

dardayan ‘scattered, cluttered, all over the place’ (TaS: طار طغان *tartağan, dardağan*) (15–19th c.), Mascis 1677: 87a: ⟨dardaghan⟩

The base **d|tayan* does not seem to be attested. The word *dardayan* is surely linked to Ott. طاغتمق *dayytmak* ‘to disperse’ (Meninski 1680: II 3070), دغدغه ~ دغدغه *daydaye(t)* i.a. ‘confusion, commotion’ (Meninski 1680: II 2093f), and داغداغان ~ طاغداغان *daydayan* ‘1. dispersed; confused; 2. confusion’ (Meninski 1680: II 3070). Its exact etymology, and the nature of this relation, however, are not clear.

The shape *dardayan* most probably evolved from *daydayan*. It might have been because the latter ceased to be morphologically transparent, or generally, because *C*-type reduplication was a more common phenomenon than whatever process created *daydayan*, so that analogy, facilitated by phonetic similarity, caused *daydayan* to shift to what was seen as a more regular shape.

As for *daydayan*, the first, and less probable possibility, is that it was originally a germ of a separate type of reduplication where the first three sounds are reduplicated and prepended to the base without any additional element in between. There are more examples which might be interpreted as supporting this idea (see 3.1.6) but it is more likely that, with perhaps one exception, they are just a collection of special cases within the standard *C*-type of reduplication.

The second, considerably more plausible possibility is that *daydayan* is a phonetically simplified participle in *-gan* from an intensified causative of *dayytmak*: **dayyt.a* > **dayyt.a.γan* > **dayt.a.γan* (the syncope of a high vowel in the middle syllable is natural in the Turkic languages) > *dayd.a.γan*, which was finally reinterpreted as *dar.dayan*. The weak point of this proposition is that it is based on the use of the *-gan* participle, which is rather atypical of the Oghuz languages.

See also Trkm. *duv dagyn* ‘scattered, diffused’ in 3.1.10.

dipdiñsüz in ~ *olmak* ‘to completely lose one’s peace’ (TaS: دىپدينسوز *dipdiñsüz*) (15th c.)

The base **diñ* does not seem to be attested. The word is probably linked in some way to Tat. *tyn* ‘quiet, peaceful’ and related forms (see Bshk. *tymyk, tyn, tynys*, Kklp. *tynyk*, Kzk. *tynyk, tynyš*, Tat. *tymyzyk, tyn, tynyč* and Uigh. *teč, tin* and *tinč* in 3.1.11).

Semantically, this would seem to be a more plausible connection than دىنسىز *dĩnsyz* ‘without faith; infidel’ (Meninski 1680: II 2218) or دىنسىز *densyz* ‘unrestrained, lax, shameless’ (Meninski 1680: II 2107).

māsmāvi (Comidas de Carbognano 1794: 27: «ماس ماوی mas mavī», von der Berswordt 1839: 11: «ماس ماوی mas mawi», Jehlitschka 1895: 57: «ماس ماوی maß mawí», Sami 1901: «ماس ماوی», Räsänen 1957: 74: «mas-mavi»)

Of the sources used here, this form is only attested in four, and neither of them marks vowel length in its Latin transcription, if the word has one. The Arabic spelling is the same in all cases, and suggests quite clearly a long *ā* in the reduplicated anlaut. This goes against the general Turkic practice (see 3.1.20).

A guess could be ventured that the spelling with «ماس» merely attempts to preserve the original orthography of the Arabic etymon (ماءى/ماوى *māwī/māī*, see Nişanyan ÇTES), and to mark the relation between the two parts of the reduplication, rather than to faithfully reflect the Ottoman pronunciation.

This conjecture can be strengthened by the example of *sāry* attested as both «صپ صارى» and «صاب صارى», in this chronological order (see below). Likewise, the reduplicated vowel is short in Ott. *aphāzyr* and *tamtāze*, but it is at the same time spelt long in *sāmsāfi*, and surprisingly, in *tāstamām* (see below on both).

öpuzun (Meninski 1680: I 481: «اوپ اوزون öp uzun») ♦ **uzun** ‘long’

The unusual reduplication is perhaps the result of fluctuations in labial vowels in 17th century Ottoman; see also «ايون ojun» ~ «öjün» ‘game’ and «بورى bor̄y» ~ «bur̄y» ‘tuba; acumen’ in Meninski 1680: I 560 and 917, respectively. The choice of this specific shape might have been dictated by Meninski’s general preference for front variants, see e.g. his remarks on declensional suffixes: *Dativus* [...] *pronuntiat* propriè e; *sæpe autem, præfertim apud plebem, ut a* (Meninski 1680: V 26), *Ablativus* [...] *den autem & ten regulariter pronuntiat, sed plerumque, in vulgari præfertim fermone, profertur ut tan & dan* (V 27).

See also 3.1.19 for other reduplications with anlaut not matching the base.

sāmsāfi (Comidas de Carbognano 1794: 27: «سام صافى sam safi») ♦ **sāfi** ‘pure’

Apart from the stress, which Comidas de Carbognano for some reason marks consistently on the ultima, this form is unusual in that it appears to have its vowel not shortened during reduplication. This is against the Turkic practice in general, and against the Turkish one in particular. Two more Ottoman words appear to reduplicate likewise (see *māsmāvi* and *sāpsāry* above and below), and in both cases it might be suspected that the long spelling is but an orthographic device to highlight the link between the base and the reduplicated anlaut. It seems plausible, that the case of *sāmsāfi* is no different. See also *aphāzyr* and *tamtāze*, but also the unusual *tāstamām* below.

sāpsāry (Comidas de Carbognano 1794: 27: «صپ صارى sap sar̄y», von der Berswordt 1839: 12: «صپ صارى sap sary», Jehlitschka 1895: 57: «صاب صارى Bap Bary») ♦ **sary** ‘yellow’

The spelling of this form is inconsistent: the earlier versions suggest a short vowel in the reduplicated anlaut, and the later one a long one.

This chronology is rather fortunate for us here. Since shortening of the reduplicated vowel is considered here to be the oldest method (see 3.2.6 and 3.4.4), it would be groundless to assume a lengthening in Ottoman and subsequent reshortening of the vowel by the end of the 19th century. Rather, it seems that the ⟨ا⟩ in ⟨صاپ⟩ serves no other purpose than to orthographically emphasize the relation between the two parts of the reduplication.

Possibly, *māvi* and *sāfi* are parallel cases (see above), but see also the rather atypical *tāstamām* below.

tastamām (Comidas de Carbognano 1794: 27: «تاز تمام tas tamàm») ◇ *tamām* ‘proper, right, just’

This form is quite extraordinary. In general, the long vowel in the reduplicated anlaut can be explained as a reduplication of more than just the initial mora (see 3.2.6 on morae), as a secondary lengthening, or as an inaccurate attestation.

Since the base comes from Ar. تمام ‘complete, proper, right’ (Nişanyan ÇTES), and does not seem to have ever had its first vowel lengthened on the Ottoman ground, the first possibility must be excluded.

The second possibility (secondary lengthening) would have been unusual, but it is not entirely impossible. What appears to be parallel cases can be found in Dolgan and Yakut, e.g. Yak. *būsütün*, *ūnutary* (see 2.21.4).

Finally, the third option (inaccurate attestation) seems to be quite likely. This explanation has been assumed in the cases of Ott. *māsmāvi*, *sāmsāfi*, and *sāpsāry* (see above), the last of which happens to provide a convenient chronological argument in favour of this interpretation.

At present, the final answer cannot be given. It might seem that there did exist in Ottoman something like an orthographic tradition of spelling the reduplicated vowel long, but actual material support for it is weak. Parallel cases of secondary lengthening, likewise, can only be found in geographically very remote and barely related languages.

topdolu (Mes‘ūd: 640b: «toptolu», Harsány 1672: 262a: «topdoli», Mascis 1677: 241b: «top dolu», Meninski 1680: II 3085: «طبطولو, طبطولو toptolü», II 3135: «طوطولو toptolü», de Preindl 1790: 19: «dop dolou», Comidas de Carbognano 1794: 27: «طوپ طول top tolü», von der Berswordt 1839: 12: «طوپ طول top dolu», Jehlitschka 1895: 57: «طوپ دولو top dolú», Németh 1916: 41: «طوب طول dob-dolu») ◇ *dolu* ‘full’

This reduplication is quite unusual in that the reduplicated anlaut often does not match the anlaut of the base. As many as four combinations are attested: *t-t* (14th c.), *t-t* (17th c.), *t-d* (17th and 19th c.), and *d-d* (18th–19th c.).

It seems as if the change *t- > d-* occurred between the 14th and 17th century in the base, but the reduplication has not caught up with it until the 19th century. The Arabic spelling generally does not reflect this differentiation, and employs ⟨ط⟩ for both places.

This suggests that the voicing in *dopdolu* occurred later than would have been expected by Doerfer 1975–76: 121; see fig. 2.5.

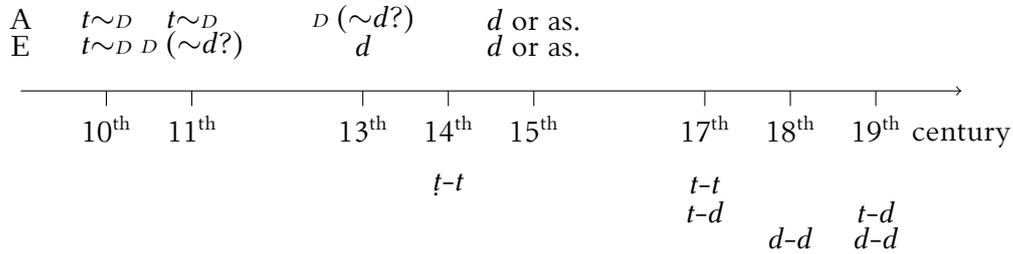


Figure 2.5: Voicing of Oghuz initial *t-*: 1. (above the line) according to Doerfer 1975–76: 121, in a back ('A') and front ('E') environment, and 2. (below the line) as observed in *dopdolu* 'absolutely full'. The abbreviation 'as.' denotes assimilation caused by a voiceless consonant occurring later in the word. Small caps *D* is for 'half-voiced *d*'.

However, Doerfer's impression (p. 124) is that "*t-* [...] kommt bei all solchen Wörtern vor, auf denen ein starker Nachd[r]uck ruht, [...] oder die Dinge mit starkem Gefühlsakzent [...] bezeichnen". This is accurately confirmed by what can be found in Meninski 1680, where *dolu* is always given with *d-* (‹دولي doḷy› (II 2190), ‹طلو doḷu› (II 3122), ‹طولو doḷu› and ‹طولی doḷy› (II 3153), and ‹طولي doḷy› (II 3154)), while *dopdolu* with two *t*'s (‹طبطلو, طبطلو toptolü› (II 3085), and ‹توپطولو toptolü› (II 3135)).

Apparently, four tendencies were at play here: **1.** voicing of initial *t-* > *d-*; **2.** reluctance of 1. to occur in emphatic words; **3.** assimilation to the following voiceless consonant (here *-p-* or *-pt-*), and **4.** desire for morphological transparency of reduplications.

With this conclusion, it might seem, the case can be closed. However, there are more reduplications in Ottoman which begin with a *d-*, and do not display similar variation: *degirmi*, *diri*, *doyru*, *dolajinže*, *duru*, and *düz*.

One possible explanation of the uniqueness of the behaviour of *dopdolu* is that the reduplicated *to.p-* happens to have a form identical to *top* 'ball, sphere' which, accidentally, is a semantically conceivable attributive for 'full'. The word might have been at some point in time reinterpreted as **top tolu* lit. 'ball-full', i.a. 'full like a ball = completely full', and thus have its phonetic evolution stalled.

Naturally, it must be kept in mind that the above considerations depend solely on the accuracy of romanized attestations. To ascertain this, however, research extending far beyond the scope of the present work would be necessary.

See also 3.1.19 for other reduplications with anlaut not matching the base.

2.13.4 Structure

Four closing consonants of *C*-type are attested in a total of 58 examples derived from 54 unique bases, in a relatively even distribution:

m: 13 examples: *bejaz*, *boš*, *čyplak*, *düz*, *gök*, *jaš*, *jassy*, *ješil*, *sāfī*, *sijāh*, *syky*, *tatly*, and *tāze*,

p: 25 examples: *ačyk*, *ak*, *alaža*, *ansyz*, *ary*, *čevre*, *čürük*, *diri*, *dolu*, *duru*, *düz*, *ejī*, *hāzyr*, *jalynyz*, *jassy*, *kara*, *karšy*, *kuru*, *kyrmyzy*, *kyryk*, *kyzyl*, *sary*, *siori*, *syrlak*, and *uzun*,

pp: 1 example: *ak*,

r: 5 examples: *čabuk*, *čyplak*, *syklam*, *temiz*, *top*, and

s: 14 examples: *belli*, *berāber*, *büjük*, *bütün*, *degirmi*, *doyru*, *dolajınže*, *götürü*, *katy*, *koža*, *kyvrak*, *māvi*, *mor*, and *tamām*.

Apart from *ak* with a single and double *p*, only three words have more than one closing consonant possible: *čyplak* (*m* and *r*), and *düz* and *jassy* (both *m* and *p*).

The first vowel of the base is long in five cases: *hāzyr*, *māvi*, *sāfī*, *sāry* and *tāze*. In the first and the last one, it clearly has been shortened in the reduplication. The cases of *māvi*, *sāfī*, and *sāry* are more ambiguous, and possibly parallel; see *māsmāvi*, *sāmsāfī*, and *sāpsāry* in 2.13.3 above. Rather unusual is the case of *tāstamām*, where the reduplicated vowel appears to have been actually lengthened; see *tāstamām* in 2.13.3 above. See also 3.1.20 on shortening of the reduplicated vowel in general

Beside the shortening of long vowels, the reduplicated anlaut does not exactly match the head of the base in four cases: *aphāzyr*, *«biz butoun»*, *öpuzun*, and *topdolu*. See 2.13.3 above.

One unexpected finding that needs to be made note of is accent. In all the sources used in the present work, not just the Ottoman, whenever accent is discussed or marked in the examples, it is said to fall on the reduplicated anlaut. Here, however, Comidas de Carbognano 1794 consistently marks it on the final syllable, and Jehlitschka 1895 – assuming that acute denotes accent in his notation, which he does not clarify – transcribes two of his examples as *«apañßýz»* and *«büßbütún»*, and the remaining twenty as two words, the second of which (the base) has an acute on its final syllable, if it is not monosyllabic. At the same time, no other source seems to make a statement about accentuation of reduplications at all.

The only possible conclusion, it seems, is that Ottoman reduplications were stressed on the final syllable, which is inconsistent with what we find in modern Turkish and, in fact, any other Turkic language. The situation is not sufficiently clear for me to attempt final judgements.

In all likelihood, the data collected here are incomplete, and must be considered an illustration rather than a solid base for far-reaching conclusions which, accordingly, will not be drawn in the present work.

The main reason for this supposition is that reduplications are primarily emotional formations, and a greater part of them was most probably considered just as, or even more colloquial in the Ottoman period as they are now. The lexicographical practice of the time tended to be highly selective and only focus on what was considered to be the beautiful variety of the language. Many reduplications might have been discarded, or even unknown to the authors of dictionaries.

But it is also intriguing that the Ottoman collection is just about a quarter of the size of that of modern Turkish. (Ottoman has 58 examples, seven of which are no longer attested in the 177-strong Turkish set; see 3.4.6 for a comparison.) It does not seem plausible, that the 133 reduplications which mark the difference between the two collections should all have been coined during the 20th century, more than tripling the stock accumulated over the previous ten centuries. Especially so, as in the 21st century reduplication is considered to be essentially not productive any longer.²²

In reality, however, apparently only about 84 examples are in actual widespread use across the territory of modern Turkish (see 2.16.4). The other half of the collection is, it seems, territorially or otherwise limited, and so probably not inherited directly from (literary) Ottoman.

Thus, the gap in attestations between Ottoman and Turkish can be estimated at about forty reduplications. A part of this group has probably entered literary Turkish from dialects after 1928, and possibly some might have also been coined after that date, but their exact number seems impossible to establish based on the data collected here, just as it is impossible to estimate the number of reduplications that might have arisen and gone out of use unrecorded, both entirely during the Ottoman period.

See 3.4.6 for another argument supporting these conclusions.

2.13.5 Semantics

The reduplicated meaning is almost always a simple intensification or the same as the base meaning. The only exceptions are: the meaning of *dümdüz* in von der Berswordt 1839: 12 ('entirely conjunct') which, in the light of the other attestations, appears to be imprecisely extracted from some sentence the author used as the source, and what seem to be two mistakes of the same kind in TaS in *gömgök* ('extremely') and *tortop* ('derlenip toparlanmış olarak').

The only word that is not clearly an adjective or an adverb, is *top* 'a round body' (→ 'completely round'). Taking into account, however, how porous the border between adjectives and nouns is in Turkic, it hardly is an improbable example.

²² Although, there are examples like *hyphyzly* 'very fast' or *koskomik* 'very funny', which can be found e.g. in the TS Corpus (Sezer [draft]) or the Turkish National Corpus (Akşan et al. 2012), but are apparently missing from traditional dictionaries. Their number and status are not clear.

2.14 Shor

Very few *C*-type reduplication are attested in Shor, pronouncedly less than in Khakas and Tuvinian, or even Oïrot. Only one closing consonant can be observed. All examples are adjectives or adverbs, and no non-trivial semantic evolutions are to be seen. It is only the derivatives of *ak* ‘white’ and *kök* ‘blue’ in 2.14.3 that are more unusual.

2.14.1 Sources

Sources for Shor are scarce. Here, Čispijakov 1992: 106, Dyrenkova 1941: 78, Kurpeško Tannagaševa / Akalın 1995 (ca. 4000 entries), and Ščerbak 1977: 120 have been used, and some further attestations have also been found in Schönig 1998b: 408 and Stachowski M. 1998b: 109.

Dyrenkova 1941: 78 gives a rather imprecise description of the formation of reduplications in Shor. Namely, it states that adjectives have a syllable prepended to them, whose “первые два звука [...] тождественны с первым слогом этого качественно-го имени, а последний согласный — губной *n*”, in spite of *apak* being given as an example just four lines below. She also suggests that reduplications might stem from doubled adjectives, as in *čōn čōn* lit. ‘fat fat’ = ‘very big, very fat’.

Transcription:

| | |
|---|--|
| Čispijakov 1992: $\kappa \rightarrow k$, | Schönig 1998b: $q \rightarrow k$, |
| Dyrenkova 1941: $\varepsilon \rightarrow e \mid \text{нв} \rightarrow \eta \mid \ddot{o} \rightarrow \ddot{o} \mid$ | ShorTS: $\iota \rightarrow y \mid \text{ş} \rightarrow \text{š}$, |
| $\ddot{y} \rightarrow \ddot{u} \mid VV \rightarrow \bar{V}$, | Stachowski M. 1998b: $\ddot{a} \rightarrow e$. |

2.14.2 Standard cases

ak ‘white’ \diamond *apak* (Dyrenkova 1941: 78, Ščerbak 1977: 120, Čispijakov 1992: 106)
See *apagaš* in 2.14.3 below.

apagaš ‘snow-white’ \diamond *apapagaš* (Dyrenkova 1941: 78)
See *apagaš* in 2.14.3 below.

kara ‘black’ \diamond *kapkara* (Dyrenkova 1941: 78, Čispijakov 1992: 106, Schönig 1998b: 408)

kök ‘blue’ \diamond *köpkök* (Dyrenkova 1941: 78)

kyzyl ‘red’ \diamond *kypkyzyl* (Dyrenkova 1941: 78, Čispijakov 1992: 106, ShorTS)

saryg ‘yellow’ \diamond *sapsaryg* (Dyrenkova 1941: 78, Čispijakov 1992: 106)

tegen ‘to no avail, in vain’ \diamond *teptegen* (Dyrenkova 1941: 78, Čispijakov 1992: 106, ShorTS)

2.14.3 Special cases

In Khakas, Shor and Oirot, *ak* and *kök* have grown into entire families of 17 or 18 forms in total. The patterns are very similar and often the commentary for Oirot, *mutatis mutandis*, also applies for Shor. See 2.12.3 for more commentary, and also 3.1.11 for other such families.

apagaš (Dyrenkova 1941: 78, Stachowski M. 1998b: 109) ◊ *ak* ‘white’

Similarly to Oir. *apagaš* in 2.12.3, this form is derived by Dyrenkova 1941: 78 and Stachowski M. 1998b: 109 from *ap-* + *ak* + *-aš*, which is very plausible. Here, too, *-aš* is a diminutive suffix (see Dyrenkova 1941: 32 (§6) and ShorTS), and, in absence of **ag|kaš*, must be concluded to have been added to the already reduplicated form.

Two more forms are linked to this one: *appagaš* above and *apapagaš* below, and can be presented as in fig. 2.6.

See also *köpegeš* for similar examples with *kök* ‘blue’.

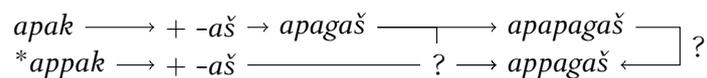


Figure 2.6: Reduplications of Shor *ak* ‘white’ &c.

apapagaš (Dyrenkova 1941: 78) ◊ *ak* ‘white’

This word is a re-reduplication of *apagaš* above, resulting in a triple intensification (two reduplications and one diminutive), perhaps due to the emphatic erosion of the original *apak*. See *apagaš* above, and also 3.1.13 for parallel examples.

appagaš (Dyrenkova 1941: 78, ShorTS, Stachowski M. 1998b: 109) ◊ *ak* ‘white’

Mutatis mutandis, the commentary to Oir. *appāš* in 2.12.3 applies. See *apagaš* above.

köpegeš (Dyrenkova 1941: 78, Stachowski M. 1998b: 109) ◊ *kök* ‘blue’

Similarly to Oir. *köpögöš* &c. in 2.12.3, this form is most probably a composition of *köp-* + *kök* > *köpkök* + *-aš* (diminutive, see Dyrenkova 1941: 32 (§6), ShorTS, and Stachowski M. 1998b: 78). Here, too, **kegeš* and *kögeš* do not seem to be attested, suggesting reduplication was applied before the diminutive.

Interesting, however, is the vocalism of this form and of the related *köppegeš* below, which seems to imply backward propagation of harmony from the suffix (*-aš*) to the stem (*köpkök*). Similarly in Khak. *köppeges* (see 2.9.3).

See *apagaš* above for a similar set of examples based on *ak* ‘white’.

köppegeš (Dyrenkova 1941: 78, Stachowski M. 1998b: 78) ◊ *kök* ‘blue’

Mutatis mutandis, the commentary to Oir. *köppōš* in 2.12.3 applies, plus the remark on what appears to be backward harmony in *köpegeš* above.

2.14.4 Structure

One closing consonant of *C*-type is attested in a total of seven examples derived from seven unique bases:

p: 7 examples: *ak*, *apagaš*, *kara*, *kök*, *kyzyl*, *saryg* and *tegen*.

2.14.5 Semantics

All reduplicated meanings are simple intensifications of their bases. Parts of speech other than adjectives and adverbs do not occur.

2.15 Tatar

Tatar reduplications are moderately numerous, and quite uniform.

Three closing consonants are attested, but one can only be found in one word (*pp* with *ak*), another in just three (*m* with *jäšel*, *kük* and *tügäräk*), and the remaining 55 cases, i.e. 93% of the entire stock, are all closed with *p*. See 2.15.4.

Also semantically are Tatar reduplications quite standard. Non-adjectives or non-trivial evolutions effectively do not occur. See 2.15.5.

Perhaps only the case of *čepči* ‘completely raw, ...’ and its phonetic anomaly in the reduplicated anlaut stands out from the general picture. See 2.15.3.

2.15.1 Sources

The main source of the material is TatRS which contains ca. 38 000 entries. Some attestations have also been found in Baskakov 1985: 30, Berta 1998a: 284, Pallas 1786–89, Pomorska 2004: 144, Räsänen 1957: 74, Ščerbak 1977: 120, Serebrennikov/Gadžieva 1986: 112, and in the grammars listed below.

For grammatical descriptions, Latypov 1969: 173f and TatGramm: 69 have been used.

The information in Latypov 1969: 173f is presented from the perspective of intensification in general, and the description of reduplication is reduced to a few examples and a statement that not all adjectives can be thus intensified. Interestingly, however, the process is considered *prefixation* rather than *reduplication* – which, up to a point, justifies the inclusion of *ör-jaŋa* ‘completely new’ among what are in fact reduplications; see Bshk. *örjaŋy* in 2.3.3.

In TatGramm: 69, the perspective is the same, but here, intensification is considered to be carried out *при помощи односложных усилительных слов*. Only three examples are given: *apak* ‘snow-white’, *iŋ avyr* ‘very heavy’ and *örjaŋa* ‘brand-new’. It is also noted that stress is initial in such formations.

Transcription:

Baskakov 1985, Latypov 1969, TatGramm, TatRS: $a \rightarrow \ddot{a}$ | $h \rightarrow h$ | $\eta \rightarrow \eta$ | $\theta \rightarrow \ddot{o}$ | $\gamma \rightarrow \ddot{u}$ |
 $\varkappa \rightarrow \check{z}$ | $\iota o \rightarrow ju, j\ddot{u}$ | $\pi \rightarrow ja, j\ddot{a}$
 Berta 1998a: $e \rightarrow \ddot{a}$ | $\check{e} \rightarrow e$ | $q \rightarrow k$ | $y \rightarrow j$,
 Pallas 1786–89: $x \rightarrow \chi$ | $b \rightarrow$ ignored,
 Ščerbak 1977: $\check{i} \rightarrow e$,
 Serebrennikov/Gadžieva 1986: $a^o \rightarrow a$ | $\kappa \rightarrow k$,

2.15.2 Standard cases

äče ‘bitter; sour’ \diamond **äpäče** (Ščerbak 1977: 120)

ačyk i.a. ‘light, bright’ \diamond **apačyk** (Räsänen 1957:74, TatRS)

ajaz ‘clear, serene, bright’ \diamond **apajaz** (TatRS)

ak ‘1. white; 2. grey; 3. clean, pure; 4. happy, cheerful’ (TatRS) \sim **aχ** ‘white’ (Pallas 1786–89 s.v. *бѣло*: *ахъ*) \diamond **apak** ‘very white’ (TatRS, Latypov 1969, TatGramm: 69, Berta 1998a: 284) \diamond **appaχ** (Pallas 1786–89 s.v. *бѣло*: Tobolsk *аппахъ*) \sim ***appak** (TatRS: only attested in *appagym* ‘my little white one’)

See *appagym* in 2.15.3 below.

aryk ‘very slim, skinny’ \diamond **aparyk** (TatRS)

az ‘little, few’ \diamond **apaz** (TatRS)

buš ‘empty, free’ \diamond **bupbuš** (TatRS)

či ‘1. raw, half-baked; 2. round, complete, true; 3. *dial.* immature, unripe’ \diamond **čepči** ‘1. *intens.*; 2. \equiv ; 3. inveterate, double-dyed, genuine’ (TatRS)

See *čepči* in 2.15.3 below.

čibär ‘beautiful’ \diamond **čipčibär** (TatRS)

čista ‘clean’ \diamond **čipčista** (TatRS)

čuar ‘motley’ \diamond **čupčuar** (TatRS)

čyn ‘true, faithful, accurate, certain’ \diamond **čypčyn** (TatRS)

döres ‘true, faithful, accurate’ \diamond **döpdöres** (TatRS)

gadi ‘simple, normal, ordinary’ \diamond **gapgadi** (TatRS)

jakty ‘light(ing)’ \diamond **japjakty** ‘very light, very bright’ (TatRS)

jakyn ‘close, near’ \diamond **japjakyn** (TatRS)

jalangač ‘naked, stripped’ \diamond **japjalangač** (TatRS)

jäš ‘young’ \diamond **jäpjäš** (TatRS, Latypov 1969)

jäšel ‘green’ \diamond **jämjäšel** (TatRS, Latypov 1969, Berta 1998a: 284)

Pallas 1786–89 s.v. *зелено* lists ‘ямьяшиль’ for the Tobolsk region. The intended form is probably the same.

jomry ‘round’ \diamond **jopjomry** (TatRS)

jomšak ‘soft’ \diamond **jopjomšak** (TatRS)

juan ‘thick, stout’ \diamond **jupjuan** (TatRS)

- juaş** ‘meek, gentle, soft’ ◊ **jupjuaş** (TatRS)
jüeş ‘1. wet; 2. raw’ ◊ **jüpjüeş** (TatRS)
jumart ‘generous’ ◊ **jupjumart** (TatRS)
käkre ‘crooked’ ◊ **käpkäkre** (TatRS)
kara ‘black, dark’ ◊ **kapkara** (TatRS, Latypov 1969, Baskakov 1985: 30, Serebrennikov/Gadžieva 1986: 112)
karangy ‘dark(ness)’ ◊ **kapkarangy** (TatRS)
katy ‘hard, solid’ ◊ **kapkaty** (TatRS)
kory ‘dry’ ◊ **kopkory** (TatRS)
kük ‘blue’ ◊ **kümük** (TatRS, Ščerbak 1977: 120)
kyska ‘short’ ◊ **kypkyska** (TatRS)
kytyrşy ‘rough, coarse’ ◊ **kyptyrşy** (TatRS, Pomorska 2004: 144)
kyzyl ‘red’ ◊ **kypkyzyl** (TatRS, Latypov 1969, Pomorska 2004: 144)
naçar ‘bad’ ◊ **napnaçar** (TatRS)
näzek ‘thin’ ◊ **näpnäzek** (TatRS)
salkyn ‘cold, frosty, chilly’ ◊ **sapsalkyn** (TatRS)
šärä i.a. ‘naked, bald’ ◊ **šäpšärä** (TatRS)
sary ‘yellow’ ◊ **sapsary** (TatRS)
simez ‘fatty, greasy’ ◊ **sipsimez** (TatRS)
šoma ‘smooth’ ◊ **šopšoma** (TatRS)
sory ‘grey’ ◊ **sopsory** (TatRS)
takyr ‘smooth, even’ ◊ **taptakyr** (TatRS)
taza ‘healthy, sturdy’ ◊ **taptaza** (TatRS)
tekä ‘steep’ ◊ **teptekä** (TatRS)
tigez ‘smooth, even’ ◊ **tiptigez** (TatRS)
tiñ ‘equal, similar’ ◊ **tiptiñ** (TatRS: only attested in ~ *bulu* ‘to be identical’)
tügäräk ‘round’ ◊ **tümtügäräk** (TatRS) ◊ **tüptügäräk** (TatRS)
 See 3.1.4 on *m* as the closing consonant in Tatar.
tuly ‘full’ ◊ **tuptuly** (TatRS)
tury ‘straight, direct’ ◊ **tuptury** (TatRS, Latypov 1969)
tygyz ‘tight, narrow’ ◊ **typtygyz** (TatRS)
tymyzyk ‘quiet, peaceful’ ◊ **typtymyzyk** (TatRS)
 See 3.1.11 on possible cognates.
tyñ ‘quiet, peaceful’ ◊ **typtyñ** (TatRS)
 See 3.1.11 on possible cognates.
tynyč ‘quiet, peaceful’ ◊ **typtynyč** (TatRS)
 See 3.1.11 on possible cognates.
zänger ‘blue’ ◊ **zäpzänger** (TatRS)
žiñel ‘light’ ◊ **žipžiñel** (TatRS)
žyly ‘warm’ ◊ **žypžyly** (TatRS)

2.15.3 Special cases

appagym ‘my little white one’ (TatRS) \diamond **ak** ‘1. white; 2. grey; 3. clean, pure; 4. happy, cheerful’

Beside the concrete meanings ‘white’ and ‘grey’, *ak* also has a number of figurative ones. It also has two separate reduplications, *apak* and *appak*, and it appears that each only intensifies one aspect of its semantics: *p* is literal, and *pp* is figurative.

See 3.1.2 for similar examples in other languages.

čepči ‘1. *intens.*; 2. \equiv ; 3. inveterate, double-dyed, genuine’ (TatRS) \diamond **či** ‘1. raw, half-baked; 2. round, complete, true; 3. *dial.* immature, unripe’

The shape *čep* seems to only be attested in this form, and in *čep-čep* (= *čip-čip*) ‘chuck chuck (a noise made when calling chicken)’. On the other hand, *či* is the literary shape which exists along the dialectal variant *čěj* (so transcribed by Berta 1989: 267f). The reason behind the difference in the vowels is not clear to me. It is, however, beyond the scope of the present work, and so it will be ignored here, and the word included in further considerations.

See also 3.1.19 for other reduplications with anlaut not matching the base.

2.15.4 Structure

Three closing consonants of *C*-type are attested in a total of 59 cases derived from 57 unique bases, in a rather one-sided distribution:

m: 3 examples: *jäšel*, *kük*, and *tügäräk*,

p: 55 examples: *äče*, *ačyk*, *ajaz*, *ak*, *aryk*, *az*, *buš*, *či*, *čibär*, *čista*, *čuar*, *čyn*, *döres*, *gadi*, *jakty*, *jakyn*, *jalangač*, *jäš*, *jomry*, *jomšak*, *juan*, *juaš*, *jüeš*, *jumart*, *käkre*, *kara*, *karangy*, *katy*, *kory*, *kyska*, *kytyršy*, *kyzyl*, *načar*, *näzek*, *salkyn*, *šärä*, *sary*, *simez*, *šoma*, *sory*, *takyr*, *taza*, *tekä*, *tigez*, *tiŋ*, *tügäräk*, *tuly*, *tury*, *tygyz*, *tymzyk*, *tyn*, *tynyč*, *zänger*, *žiŋel*, *žyly*, and

pp: 1 example: *ak*.

Two words have more than one closer possible: *ak* \sim *aχ* (*p* and *pp*) and *tügäräk* (*m* and *p*).

Perhaps noteworthy is also the pair *jäš* \rightarrow *jäpjäš*, *jäšel* \rightarrow *jämjäšel*, where the closing consonant appears particularly clearly to be entirely independent from the phonetic shape of the rest of the word.

2.15.5 Semantics

In the great majority of cases, the reduplicated meaning is a simple intensification of the base meaning, or apparently the same.

As far as the distinction between adjectives and other nomina can be drawn in Turkic, all examples are of quite distinctly adjectival character.

2.16 Turkish

C-type reduplications are much more numerous in Turkish than in any other Turkic language. In fact, unrealistically so; see 3.2.2. They are also quite diversified, second only to the neighbouring Azeri – in which, however, the examples are fewer by almost two thirds – and to Yakut, where they are fewer by almost a half.

Five closing consonants are attested. With 101 examples, *p* is decidedly the most common, followed by *s* in 38 words, *m* in 29, and *r* in just 8, and *š* in one, which is *bešbeter* and rather unclear. In as many as seventeen examples, more than one closing consonant is possible. The long vowel in the stem is regularly shortened in the reduplication. See 2.16.4.

Semantic evolutions are very rare and very moderate in Turkish reduplications. Apart from a handful of half-substantival bases, only adjectives and adverbs can be reduplicated. See 2.16.5.

Special cases are neither numerous nor intricate. This might surprise, seeing how many examples there are in total. The most interesting ones are perhaps *köskütük* ‘completely drunk’, *perperışan* ‘in utter disarray, ...’ and *tamtakyr* ‘completely empty’. See 2.16.3.

2.16.1 Sources

The main sources of the material are Hatiboğlu 1973 and Müller 2004 (see below for the latter). Some additional attestations and new examples have also been found in later works (Balçı 2006 and Stachowski M. 2009), and two earlier ones but not included in Hatiboğlu 1973 and Müller 2004 (Kononov 1956 and Ščerbak 1977: 120).

To the best of my knowledge, Hatiboğlu 1973 was the first extensive study of Turkish *C*- and *CV*-type reduplications. Her focus was primarily on establishing synchronic, phonetic rules of distribution of closing consonants. See 1.1.2.

Müller 2004 is the second, and newest, extensive study. Based on his annotation, his main source (47% of the examples) is Steuerwald 1972 or 1974 (not clear from Müller 2004: 85 and 353). 23% of the examples are marked as coming from Hatiboğlu 1973 (in reality she lists 74.27% of the reduplications in his collection), and 14% are from Wedel 2000 and 2003?. The remaining 16% are scattered across ten sources. A list is given in Müller 2004: 353–357, but it has many gaps. More examples can be found on pp. 86f, 109, 119, 155 and 218f.

Müller considers Wedel’s examples to be unreliable (p. 218) and about this, I could not disagree with him. They are included in the list in 2.16.2 for the record but marked with a question mark and excluded from further considerations, even if some appear to be considerably more believable than others (e.g. *japjašly* ‘very old’ or *koskomik* ‘very funny’ which are apparently missing from traditional dictionaries but can be found in the TS Corpus (Sezer [draft]) and the Turkish National Corpus (Akşan et al. 2012)).

Meanings are sometimes given in a slightly particular way in Müller 2004, e.g. *čanly* ‘glockenförmig’. To avoid irrelevant digressions on German and other semantic fields, I will simply ignore them here and provide my own translations, based mostly on Alderson/İz 1959, Podolak/Nykiel 2008, and Redhouse 2001.

Finally, I added the notation of length in vowels which Müller ignores completely, based primarily on BTS.

The literature on reduplication in Turkish is relatively vast. The majority struggles to establish synchronic phonetic rules describing the choice of the closing consonant in modern Turkish. As such, it is of rather limited use here (see 3.4.1f for reasons). Müller 2004 is the newer one of the two bigger studies, and by far the most comprehensive of all. See 1.1.2 for more about this and other works.

Müller reviews many previous trials mundanely pointing out the constantly recurring weaknesses (p. 90–106, especially 96f), but focuses primarily on Hatiboğlu 1973 and Demircan 1987 and 1989. He discusses the three thoroughly and demonstrates their rules to be insufficient or imprecise (pp. 118–133). He then builds on them his own set of seven rules (p. 149f) and, perhaps not entirely realizing it, proceeds to disprove these, too (p. 251f).

His rules are purely synchronic and based solely on phonetics. This is their first weakness. Knowing that seventeen words have more than one closing consonant possible (e.g. *jepjeni* : *jesjeni* ‘brand-new’, see 2.16.4) suffices to realize even before setting to formulate the first rule that the enterprise is already doomed, and the best it can yield are *tendencies*. See 3.4.1 for more on the subject.

Secondly, although Müller goes out of his way (p. 156f) to produce a phonetic motivation for his phonetic rules, the results must be judged dubious at best simply because they so often fail to match the actual attestations (in about a fifth of the times, see argument six below).

Thirdly, information about the shortening of the reduplicated vowel (as in *tāze* → **tāptāze* → *taptāze* ‘very fresh’) is missing entirely.

Fourthly, in some cases, the rules are mutually exclusive. For example, Regel E (see appendix A) assigns *s* as a possible alternative to *p* in such words as *berrak* ‘limpid, clear’ or *dāçnyk* ‘scattered, dispersed’ but at the same time, it also assigns to them *m* “und kein anderer Laut”. (The actual reduplications are *besberrak* and *dapdāçnyk*.)

Fifthly, exceptions appear in almost all the critical points. In fact, the ‘rule’ for *r* (Regel F) is nothing but a list of exceptions. This is consistent with the evident general overfitting of the rules (see the second argument above).

Last but not least, if one counts affricates as *Zischlaute*, which Müller apparently does,²³ 21% out of the 165 reduplications he used as the base to formulate his rules, are described incorrectly. It needs to be admitted, however, that this number drops to

²³ See e.g. “[...] keine Zischlaute (von den Affrikaten *c* in [...] und *ç* in [...] und von *ş* in [...] abgesehen)” (p. 151).

mere 6% if the unfortunate restriction is removed from the rule for *m* (“*m* (und kein anderer Laut)”). But again, the rules for *m* and *s* then become just optional alternatives for *p*, and the rule for *r* already contains nothing but exceptions.

Thus, Müller’s rules can be either plain wrong (incorrect for 21% of examples), or for all practical purposes reducible to one, single rule as follows: “the closing consonant in Turkish is *p*, *m*, *s* or *r*, the latter in the following eight words only: ...”.

Altogether, Müller’s effort appears to have been severely mislaid. Undismayed, he decided to proceed to test the predictive power of his rules. The results are significant in the statistical sense of the word; for further details see appendix A. See also 1.1.2 for more commentary on Müller’s work.

To some extent, the choice of the closing consonant in Turkish reduplications can be correlated with the phonetic shape of the base, but formulation of *rules* in the proper sense of the word is not possible. See appendix A for the tendencies as described by Müller 2004 and for my own test.

Transcription:

Balcı 2006, Hatiboğlu 1973, Kononov 1956, Müller 2004, Stachowski M. 2009: $c \rightarrow \check{z} \mid \check{\zeta} \rightarrow \check{c} \mid \check{g} \rightarrow \bar{V}, j \mid \iota \rightarrow y \mid j \rightarrow \check{z} \mid k\hat{a} \rightarrow ka \mid \check{s} \rightarrow \check{s} \mid y \rightarrow j$,
 Ščerbak 1977: $\check{\varepsilon} \rightarrow e \mid \check{j} \rightarrow \bar{V}, j \mid \check{i} \rightarrow y$.

2.16.2 Standard cases

Most examples attested in Müller 2004 can be found on multiple pages. Below, only the lowest numbers are given.

ačyk ‘open, light of colour’ \diamond **apačyk** ‘intens. (Kononov 1956: 157, Müller 2004: 353); obvious, clear (Stachowski M. 2009: 120)’, Hatiboğlu 1973: no meaning given

ajdyn ‘light, bright, enlightened’ \diamond **apajdyn** (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 116)

ajdynlyk ‘light(ness), illumination’ \diamond **apajdynlyk** (Hatiboğlu 1973)

ajny ‘the same, identical’ \diamond **apajny** (Stachowski M. 2009: 116)

ajry ‘separate’ \diamond **apajry** (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 116)

ak ‘white’ \diamond **apak** (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 116)

alaža ‘motley, multicoloured’ \diamond **apalaža** (Hatiboğlu 1973, Müller 2004: 353)

ansyz(y_n) ‘sudden(ly)’ \diamond **apansyz(y_n)** (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 116)

āšīkar ‘obvious, apparent’ \diamond **apāšīkar** (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 116)

ācyr ‘heavy, serious’ \diamond **apācyr** (Hatiboğlu 1973, Müller 2004: 353)

- ažy** ‘bitter, hot, painful’ ◇ *apažy* (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 116)
- bajat** ‘stale’ ◇ *basbajat* (Müller 2004: 353)
- bajāγ** ‘common, ordinary, usual’ ◇ *basbajāγ* (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 117)
- baška** ‘different, other’ ◇ *bambaška* (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 117)
- bedāva** ‘free of charge’ ◇ *besbedāva* (Hatiboğlu 1973, Müller 2004: 353, Stachowski M. 2009: 117)
- bejaz** ‘white’ ◇ *bembejaz* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- belli** ‘certain, clear’ ◇ *besbelli* (Kononov 1956: 157, Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 353, Stachowski M. 2009: 117)
- berāber** ‘together’ ◇ *besberāber* (Hatiboğlu 1973, Müller 2004: 353)
- berrak** ‘limpid, clear’ ◇ *besberrak* (Hatiboğlu 1973, Müller 2004: 354)
- beter** ‘worse’ ◇ *besbeter* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117) ◇ *bešbeter* (Müller 2004: 354)
See 2.16.4 below.
- bež** ‘beige’ ◇ *ⁱbembež* (Müller 2004: 218 after Wedel)
- bok** ‘shit(ty), crap(py)’ ◇ *bombok* (Hatiboğlu 1973, Müller 2004: 354)
- boš** ‘empty’ ◇ *bomboš* (Kononov 1956: 157, Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 354, Stachowski M. 2009: 117)
- boz** ‘grey’ ◇ *bomboz* (Hatiboğlu 1973, Müller 2004: 354)
- bulanyk** ‘blurry, fuzzy’ ◇ *bumbulanyk* (Müller 2004: 354) ◇ *busbulanyk* (Müller 2004: 354)
- buruš** ‘wrinkle, crease’ ◇ *bumburuš* (Müller 2004: 354)
- burušuk** ‘wrinkled, creased’ ◇ *bumburušuk* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- bütün** ‘all, whole’ ◇ *büsbütün* (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- buz** ‘ice(-cold), glacial’ ◇ *bumbuz* (Müller 2004: 354)
- čabuk** ‘quick, swift’ ◇ *ⁱčapčabuk* (Müller 2004: 219 after Wedel) ◇ *čarčabuk* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- čanly** ‘having a bell’ ◇ *ⁱčapčanly* (Müller 2004: 218 after Wedel)
- čevre** ‘around, ambient’ ◇ *čepčevre* (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 120)
See *čerčeve* in 2.16.3 below.
- čij** ‘raw’ ◇ *čimčij* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117) ◇ *čipčij* (Müller 2004: 354, Stachowski M. 2009: 117)
- čilik** ‘?’ ◇ *ⁱčimčilik* (Müller 2004: 219 after Wedel)
- čirkin** ‘ugly’ ◇ *čipčirkin* (Hatiboğlu 1973, Müller 2004: 354)

- čürük** ‘rotten’ ◇ **čümčürük** (Müller 2004: 109)
- čybyl** ‘churn’ ◇ [?]**čysčybyl** ‘?’ (Müller 2004: 219 after Wedel)
- čybyldak** ‘?’ ◇ [?]**čysčybyldak** ‘?’ (Müller 2004: 219 after Wedel)
See *žysžybyldak* in 2.16.3 below.
- čybyvyk** ‘?’ ◇ [?]**čysčybyvyk** ‘?’ (Müller 2004: 219 after Wedel)
- čyplak** ‘naked, nude’ ◇ **čyrčyplak** (Kononov 1956: 157, Ščerbak 1977: 120, Müller 2004: 354, Stachowski M. 2009: 119, 120) ◇ **čysčyplak** (Stachowski M. 2009: 119, 120)
- dar** ‘narrow, tight’ ◇ **dapdar** (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
See *dap|sdaražyk* in 2.16.3.
- dā(y)nyk** ‘scattered, dispersed’ ◇ **dapdā(y)nyk** (Müller 2004: 354)
- dazlak** (Müller 2004: 354) ~ **dyzlak** (Müller 2004: 354, Stachowski M. 2009: 118)]
‘bald’ ◇ **damdazlak** (Müller 2004: 354) ~ **dymdyzlak** (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 118)
- dejirmi** ‘round’ ◇ **desdejirmi** (Hatiboğlu 1973, Müller 2004: 354)
- derin** ‘deep’ ◇ **depderin** (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- dik** ‘upright, steep, erect’ ◇ **dimdik** (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 118)
- dinč** ‘vigorous, fresh’ ◇ **dipdinč** (Hatiboğlu 1973, Müller 2004: 354)
- diri** ‘(a)live, live(ly)’ ◇ **dipdiri** (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 118)
- dolu** ‘full’ ◇ **dopdolu** (Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 86, Stachowski M. 2009: 118) ◇ **dosdolu** (Ščerbak 1977: 120)
- dōru** ‘correct, accurate’ ◇ **dosdōru** (Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 88, Stachowski M. 2009: 118)
- durgun** ‘still, stagnant’ ◇ **dupdurgun** (Hatiboğlu 1973, Müller 2004: 119)
- duru** ‘limpid, clear’ ◇ **dupduru** (Hatiboğlu 1973, Müller 2004: 86, Stachowski M. 2009: 118)
- dürü** ‘roll(ed) up’ ◇ **düpdürü** (Hatiboğlu 1973, Müller 2004: 354)
- düz** ‘straight, smooth, even’ ◇ **dümdüz** (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 119) ◇ **düpdüz** (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 119)
- düzgün** ‘smooth, shapely, regular’ ◇ **düpdüzgün** (Hatiboğlu 1973, Müller 2004: 354)
- dyzlak** see *dazlak*
- ejri** ‘crooked, bent, awry’ ◇ **epejri** (Hatiboğlu 1973, Müller 2004: 354)
- ekši** ‘sour’ ◇ **epekši** (Hatiboğlu 1973, Müller 2004: 354)
- erken** ‘early’ ◇ **eperken** (Hatiboğlu 1973, Müller 2004: 355)
- eski** ‘old’ ◇ **epeski** (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 116)

- genč** ‘young’ ◇ *gepgenč* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 119, 120)
- geniš** ‘wide’ ◇ *gepgeniš* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- gergin** ‘tense, tight, nervous’ ◇ *gepgergin* (Hatiboğlu 1973, Müller 2004: 355)
- geže** ‘(at) night’ ◇ *gepgeže* (Hatiboğlu 1973, Müller 2004: 355)
- gök** ‘1. sky; 2. blue’ ◇ *gömgök* (Hatiboğlu 1973: no meaning given, Müller 2004: 355: ‘very blue’)
- götürü** ‘(in a) lump sum’ ◇ *gösgötürü* (Hatiboğlu 1973, Müller 2004: 355)
- güdük** ‘stubby, squat’ ◇ *güsgüdük* (Müller 2004: 355)
- gündüz** ‘(at) daytime’ ◇ *güpgündüz* (Hatiboğlu 1973, Müller 2004: 355)
- gür** ‘abundant, dense, stentorian’ ◇ *güpgür* (Hatiboğlu 1973, Müller 2004: 355)
- güzel** ‘nice, beautiful’ ◇ *güpgüzel* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- hyzly** ‘fast, rapid’ ◇ *hyphyzly* (Müller 2004: 219 after Wedel)
- ibiš** ‘idiot, fool’ ◇ *ipibiš* (Müller 2004: 355)
- inže** ‘thin, fine, slim’ ◇ *ipinže* (Hatiboğlu 1973, Müller 2004: 355, Balcı 2006: 81, Stachowski M. 2009: 116)
- iri** ‘large’ ◇ *ipiri* (Hatiboğlu 1973, Müller 2004: 355)
- išsiz** ‘unemployed, jobless’ ◇ *ipišsiz* (Müller 2004: 355)
- jabanžy** ‘foreign, alien’ ◇ *japjabanžy* (Müller 2004: 357)
- jakyn** ‘close, near’ ◇ *japjakyn* (Müller 2004: 357, Stachowski M. 2009: 119)
- jakyšykly** ‘handsome’ ◇ *japjakyšykly* (Müller 2004: 219 after Wedel)
- jalnyz** ‘(a)lone, lone(ly)’ ◇ *japjalnyz* (Kononov 1956: 157, Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 357, Stachowski M. 2009: 120)
- janlyš** ‘wrong, incorrect’ ◇ *japjanlyš* (Hatiboğlu 1973, Müller 2004: 357)
- jaryk** ‘cleft, cracked’ ◇ *japjaryk* (Müller 2004: 357)
- jaš** ‘wet, humid’ ◇ *jamjaš* (Hatiboğlu 1973, Müller 2004: 357) ◇ *japjaš* (Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119)
- jašly** ‘elderly, old’ ◇ *japjašly* (Müller 2004: 219 after Wedel)
- jassy** ‘flat’ ◇ *jamjassy* (Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119) ◇ *japjassy* (Müller 2004: 357)
- jašyl** ‘green’ ◇ *japjašyl* (Müller 2004: 357)
See *ješil* above, and 3.1.11.
- javaš** ‘slow(ly), leisurely’ ◇ *japjavaš* (Müller 2004: 219 after Wedel)
- jeni** ‘new’ ◇ *jepjeni* (Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119) ◇ *jesjeni* (Müller 2004: 357, Stachowski M. 2009: 119)
- ješil** ‘green’ ◇ *jemješil* (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119) ◇ *jepješil* (Kononov 1956: 157)
See *jašyl* above, and 3.1.11.

- jō(un** ‘dense, intense’ ◇ [?]*jopjō(un* (Müller 2004: 219 after Wedel) ◇ *josjō(un* (Hatiboğlu 1973, Müller 2004: 357)
- jumru** ‘lump, knob; tuberous’ ◇ *jusjumru* (Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119)
- jumušak** ‘soft’ ◇ *jusjumušak* (Hatiboğlu 1973, Müller 2004: 357)
- juvarlak** ‘round’ ◇ *jusjuvarlak* (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119)
- kahve** ‘brown?’²⁴ ◇ [?]*kapkahve* (Müller 2004: 219 after Wedel)
- kalyn** ‘thick, dense, stout’ ◇ *kapkalyn* (Hatiboğlu 1973, Müller 2004: 355)
- kapaly** ‘closed’ ◇ [?]*kapkapaly* (Müller 2004: 219 after Wedel)
- kar** ‘snow’ ? **kar** ‘deaf’ ? **kar** ‘profit’ ◇ *kamkar* ? *kamkar* ‘?’ (Müller 2004: 219 after Wedel)
- kara** ‘black’ ◇ *kapkara* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- karanlyk** ‘dark’ ◇ *kapkaranlyk* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- katy** ‘hard, solid’ ◇ *kapkaty* (Kononov 1956: 157, Hatiboğlu 1973, Stachowski M. 2009: 118) ◇ *kaskaty* (Kononov 1956: 157, Müller 2004: 355, Stachowski M. 2009: 118)
- kel** ‘bald, hairless’ ◇ *kepkel* (Hatiboğlu 1973, Müller 2004: 355)
- kirli** ‘dirty, soiled’ ◇ *kipkirli* (Hatiboğlu 1973, Müller 2004: 355)
- koju** ‘dark; deep, thick, dense’ ◇ *kopkoju* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118) ◇ *koskoju* (Müller 2004: 355)
- kolaj** ‘easy’ ◇ *kopkolaj* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- komik** ‘funny’ ◇ [?]*koskomik* (Müller 2004: 219 after Wedel)
- kör** ‘blind’ ◇ *kömkör* (Hatiboğlu 1973, Müller 2004: 355)
- kötü** ‘bad’ ◇ *köpkötü* (Hatiboğlu 1973, Müller 2004: 355)
- kötürüm** ‘crippled, paralyzed’ ◇ *köskötürüm* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- koža** ‘old, large’ ◇ *koskoža* ‘very large’ (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- kožaman** ‘huge, large’ ◇ *koskožaman* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- küçük** ‘small’ ◇ *küpküçük* (Müller 2004: 355) ◇ [?]*küsküçük* (Müller 2004: 219 after Wedel)
- kuru** ‘dry’ ◇ *kupkuru* (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)

²⁴ Müller 2004: 219 gives the meaning ‘kaffeefarben’ but rightly noting that *kahve* is not used as colour name.

- kütük** ‘tree-stump; baulk; log’ ◊ **köskütük** ~ **küskütük** ‘completely drunk’ (Müller 2004: 88, 106f &c.)
See *köskütük* in 2.16.3.
- kyrmyzy** ‘red’ ◊ **kypkyrmyzy** (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- kysa** ‘short’ ◊ **kypkysa** (Hatiboğlu 1973, Müller 2004: 355)
- kyvrak** ‘lithe, brisk, agile, dexterous’ ◊ **kyskyorak** ‘strongly, firmly’ (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 120)
- kyzyl** ‘red’ ◊ **kypkyzyl** (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118) ◊ [?]**kyskyzyl** (Müller 2004: 219 after Wedel)
- lāživert** ‘dark blue’ ◊ **lāplāživert** (Müller 2004: 355)
- māvi** ‘blue’ ◊ **masmāvi** (Hatiboğlu 1973, Müller 2004: 355, Balcı 2006: 81, Stachowski M. 2009: 118)
- mor** ‘purple, violet’ ◊ **mosmor** (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 118)
- olgun** ‘mature, ripe’ ◊ **opolgun** (Müller 2004: 355)
- ölgün** ‘withered, lifeless’ ◊ **öpölgün** (Müller 2004: 356)
- pembe** ‘pink’ ◊ **pespembe** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118)
- perīšan** ‘scattered, disordered, distraught, miserable’ ◊ **perperīšan** (Müller 2004: 76)
See 3.1.6 on the closer being identical to *C*₂.
- pis** ‘dirty, foul’ ◊ **pimpis** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118)
- renkli** ‘coloured, colourful’ ◊ **reprenkli** (Müller 2004: 356)
- sā** ‘1. right(-hand), dexter; 2. alive, sound, safe’ ◊ **sapsā** (Müller 2004: 356)
- sāde** ‘simple, plain, pure’ ◊ **sapsāde** (Hatiboğlu 1973, Müller 2004: 356)
- salak** ‘stupid, idiot’ ◊ [?]**sapsalak** (Müller 2004: 219 after Wedel)
- sālam** ‘sound, solid, sturdy’ ◊ **sapsālam** (Kononov 1956: 157, Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 356, Stachowski M. 2009: 120)
- sary** ‘yellow’ ◊ **sapsary** (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118)
- sefil** ‘miserable, wretched’ ◊ **sersefil** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118)
- šekerli** ‘sweet(ened), sugared’ ◊ [?]**šepšekerli** (Müller 2004: 219 after Wedel)
- serin** ‘cool, chilly’ ◊ **sepserin** (Hatiboğlu 1973, Müller 2004: 356)
- sert** ‘hard, harsh’ ◊ **semsert** (Müller 2004: 356, Stachowski M. 2009: 118) ◊ [?]**sepsert** (Müller 2004: 219 after Wedel)
- sevimli** ‘cute, amiable’ ◊ [?]**sepsevimli** (Müller 2004: 219 after Wedel)
- sijah** ‘black’ ◊ **simsijah** (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118) ◊ **sipsijah** (Kononov 1956: 157, Müller 2004: 356)
- silik** ‘worn, weak, indistinct’ ◊ **sipsilik** (Hatiboğlu 1973, Müller 2004: 356)

- širin** ‘cute, lovely, sweet’ ◇ **šipširin** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 119)
See *sipsirin* in 2.16.3 below.
- širkin** ‘?’ ◇ **šypširkin** [sic] (Müller 2004: 219 after Wedel)
- sivri** ‘pointed, sharp’ ◇ **sipsivri** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118)
- sō(uk** ‘cold’ ◇ **sopsō(uk** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118)
- sulu** ‘watery, moist’ ◇ **supsulu** (Müller 2004: 219 after Wedel)
- syk** ‘frequent, dense, thick’ ◇ **symsyk** (Hatiboğlu 1973, Müller 2004: 356)
- sykkyn** ‘distressed, annoyed, troubled’ ◇ **sypsykkyn** (Müller 2004: 219 after Wedel)
- syklam** ‘wet’ ◇ **syrsyklam** (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 120)
- syky** ‘tight, firm’ ◇ **symsyky** (Hatiboğlu 1973, Müller 2004: 356, Balcı 2006: 81, Stachowski M. 2009: 118)
- syska** ‘skinny, gaunt’ ◇ **sypsyska** (Hatiboğlu 1973, Müller 2004: 356)
- syžak** ‘warm, hot’ ◇ **symsyžak** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 118) ◇ **sypsyžak** (Müller 2004: 356, Stachowski M. 2009: 119)
- tam** ‘accurate, precise’ ◇ **tastam** (Müller 2004: 356)
- tamam** ‘whole, complete, finished’ ◇ **tastamam** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 119)
- tatly** ‘sweet, tasty’ ◇ **taptatly** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 119)
- tāze** ‘fresh’ ◇ **tamtāze** (Müller 2004: 155) ◇ **taptāze** (Hatiboğlu 1973, Müller 2004: 356, Stachowski M. 2009: 116, 119)
- tekerlek** ‘1. wheel; tire; 2. round’ ◇ **testekerlek** (Hatiboğlu 1973: no meaning given, Müller 2004: 356: ‘completely round’)
- temiz** ‘clean’ ◇ **teptemiz** (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 356, Balcı 2006: 81) ◇ **tertemiz** (Müller 2004: 356, Stachowski M. 2009: 119)
- tok** ‘full, satiated’ ◇ **tomtok** (Müller 2004: 356)
- top** ‘1. ball; 2. round’ ◇ **tortop** (Müller 2004: 356, Stachowski M. 2009: 119) ◇ **tostop** (Müller 2004: 219 after Wedel)
- topač** ‘1. (spin)top, teetotum, whirligig; 2. ~ *gibi* sturdy’ ◇ **tortopač** (Müller 2004: 356) ◇ **tostopač** (Hatiboğlu 1973, Müller 2004: 356)
- toparlak** ‘round’ ◇ **tostoparlak** (Kononov 1956: 157, Hatiboğlu 1973, Ščerbak 1977: 120, Müller 2004: 356, Stachowski M. 2009: 119)
- turunžu** ‘orange’ ◇ **tupturunžu** (Hatiboğlu 1973, Müller 2004: 356)
- tuzlu** ‘salty’ ◇ **tuptuzlu** (Hatiboğlu 1973, Müller 2004: 356)
- tykyz** ‘compact, dense’ ◇ **tymtykyz** (Müller 2004: 356)
- ufak** ‘small, little’ ◇ **upufak** (Müller 2004: 357)

- ujgun** ‘convenient, appropriate’ ◇ *upujgun* (Hatiboğlu 1973, Müller 2004: 87, Stachowski M. 2009: 116)
- ujuz** ‘itch, mange, scab, mangy, scabby’ ◇ *upujuz* (Hatiboğlu 1973, Müller 2004: 357)
- ünlü** ‘famous, renowned’ ◇ *üpünlü* (Müller 2004: 357)
- ürjan** ‘naked, nude’ ◇ *üpürjan* (Müller 2004: 357 ⟨uryan, up-uryan⟩)
- uslu** ‘well-behaved’ ◇ *upuslu* (Hatiboğlu 1973, Müller 2004: 357)
- uzun** ‘long’ ◇ *upuzun* (Kononov 1956: 157, Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 116)
- użuz** ‘cheap, inexpensive’ ◇ *upużuz* (Hatiboğlu 1973, Müller 2004: 356)
- lyk** ‘lukewarm, tepid’ ◇ *ypylyk* (Hatiboğlu 1973, Müller 2004: 355)
- yarak** ‘far, distant’ ◇ *ypyarak* (Müller 2004: 355)
- yslak** ‘wet, damp’ ◇ *ypyslak* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 116)
- yssyz** ‘desert(ed), isolated’ ◇ *ypyssyz* (Hatiboğlu 1973, Müller 2004: 355, Stachowski M. 2009: 116)
- zajyf** ‘weak’ ◇ *zapzajyf* (Hatiboğlu 1973, Müller 2004: 357, Stachowski M. 2009: 119)
- zengin** ‘rich’ ◇ *zepzengin* (Müller 2004: 219 after Wedel)
- zor** ‘difficult’ ◇ *zopzor* (Müller 2004: 357)
- żanly** ‘(a)live, live(ly)’ ◇ *żapżanly* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- żavlak** ‘bald, hairless, naked’ ◇ *żasżavlak* (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 117)
- żybyl** ‘naked, nude’ ◇ *żysżybyl* (Hatiboğlu 1973, Müller 2004: 354)
- żybyldyk** ‘wet’ ? ‘naked?’ ◇ *żysżybyldyk* (Müller 2004: 219 after Wedel)
See *żysżybyldak* in 2.16.3 below.
- żylk** ‘addled, rotten’ ◇ *żymżylk* (Hatiboğlu 1973, Müller 2004: 354)
- żylyz** ‘feeble, scrawny, puny’ ◇ *żypżylyz* (Hatiboğlu 1973, Müller 2004: 354)
- żyvlyk** ‘?’ ◇ *żymżyvlyk* (Müller 2004: 219 after Wedel)
- żyvyk** ‘juicy, saucy’ ◇ *żysżyvyk* (Hatiboğlu 1973, Müller 2004: 354)

2.16.3 Special cases

Most examples attested in Müller 2004 can be found on multiple pages. Below, only the pages with commentary are given if it is available, and the lowest ones if it is not.

čerčeve ‘frame(work)’ (TS, Stachowski S. 1998: [48], Nişanyan ÇTES) ~ *čerčive* (Stachowski S. 1998: [48])

This word is suspiciously similar both in form and meaning to *čevre* ‘around, ambient’ (see 2.16.2 above) and to Turkish reduplications in general.

However, this is not the case. Stachowski S. 1998: [48], Nişanyan ÇTES and other sources unanimously derive it from Pers. چار چوبه *čār čube* (< چهار *čahār* ‘four’)

‘(wooden) frame’, lit. ‘four stick’, and this etymology does not seem to be lacking in any way.

dap|sdaražyk (Hatiboğlu 1973, Müller 2004: 354, Stachowski M. 2009: 122) ~ **das** ~~sdaražyk~~ (Hatiboğlu 1973, Müller 2004: 354) \diamond **daražyk** ‘very narrow, skintight’ It is not possible to determine whether it was reduplication or diminutive that was first here. Both scenarios seem to be equally plausible: *dap|sdar* → *dap|sdara* ~~žyk~~, and *daražyk* → *dap|sdaražyk*. The fact that *dasdar* is only attested dialectally can hardly be viewed as conclusive. See also *žysžybyldak* below.

However, the *-(y)žyk* diminutive suffix is even nowadays very productive (see e.g. Stachowski M. 2006: 122), while reduplication is no longer so. It will be assumed in the present work, therefore, that it was diminutive that was added later, and that *dap-* and *dasdaražyk* are not separate reduplications.

Both forms are doubly intensified. See 3.1.13 for more examples of this phenomenon.

epej ~ **epeji** ~ **epejže** (Hatiboğlu 1973, Müller 2004: 108, BTS) ~ **epejiže** (BTS) ‘quite, fairly’ \diamond Ott. *eji* ‘good’

The modern phonetic shape of the base is *iji*. The four forms here reduplicate the now-obsolete, Ottoman sounding with an *e-*. They will be included in the list for Ottoman (2.13.2), and omitted from the list for Turkish (2.16.2 above).

See also 3.1.15 for other cases of the base becoming obsolete, and 3.1.19 for other reduplications with anlaut not matching the base.

köskütük ~ **küskütük** ‘completely drunk’ (Müller 2004: 106f) \diamond **kütük** ‘tree-stump; baulk; log’

Müller 2004: 106f considers *kö|üskütük* together with *kör kütük* ~ *körkütük* id., and concludes that they are “keine Intensiv-Adjektive”. It is not clear to me what exactly brought him to this opinion.

Kör, *kös* and *küs* are all independent words: *kör* means ‘blind’, *kös* means ‘big drum’ (also in ~ *jürüme* or ~ ~ ‘to walk slowly, pensively, heavily’), and *küs* means ‘sulky, offended’.

The alleged base word, *kütük*, means ‘tree-stump; baulk; log’. A comparison of an intoxicated person to a tree, or generally something numb and stuporous, is quite possible; see e.g. Tksh. *kütük gibi sarhoş* lit. ‘drunk as a tree-stump, baulk, log’, but also Pol. *pijany jak bela* lit. ‘drunk as a log’ or Russ. *пьяны как дрoва* lit. ‘drunk as firewood’, and Engl. *stoned* and *dead drunk*. Blindness (as with *kör* ‘blind’) is also a conceivable intensification; see e.g. Engl. *blind drunk* or Nor. *blind drukket*.

Big drum would be a less obvious choice, albeit not entirely impossible because in this field, creativity seems to be bottomless; see e.g. Germ. *betrunken wie eine Strandhaubitze* lit. ‘drunk as a beach howitzer’ or Pol. *narq̄bany jak Messerschnitt* lit. ‘smashed like a Messerschmitt’.

Only *küs* ‘sulky, offended’ appears to be somewhat out of line.

In itself, *kütük* does not appear to be attested in the meaning ‘drunk’, but this needs not be considered an obstacle to deriving *köskütük* &c. from it, whether as a reduplication or as a composition.

The shape *küskütük* is phonetically most likely of the three to be a reduplication as it is the only one where the alleged reduplicated vowel and the vowel of the base actually match. At the same time, it is semantically least likely to be a composition. In the present work, it will be considered a reduplication.

On the other hand, *körkütük* is a semantically likely composition and a phonetically unlikely reduplication. Also, it is the only one of the three words that is also attested spelt separately. In the present work, it will be considered a composition.

Finally, *köskütük* is neither likely as a composition nor as a reduplication. Perhaps the most plausible explanation is that it is a variant of the reduplication *küskütük* created by analogy to the composition *körkütük*, and also to the reduplication *köskötürüm* ‘crippled, paralyzed’ – since it is probably the inertia and numbness of the drunk that motivated the comparison to a log.

sap saman ‘the whole crop’ (Marchand 1952: 62)

Although this form has the phonetic and semantic appearance of a reduplication, and reduplications of nouns are quite possible in the closely related Azeri, it seems that Marchand 1952: 56 is probably right in interpreting it as a composition of *sap* ‘stalk, stem’ + *saman* ‘straw; hay’. See also 3.1.3 for more examples of apparent reduplications.

sipsirin (Müller 2004: 87 &c.) ◊ **sirin** ‘cute’

In Müller 2004: 87 &c., the words (*sip*)*sirin* ‘(very) cute’ are mentioned, but with only a question mark instead of the source. I could not find such forms in Alderson/İz 1959, BTS, GTS, TS, or TurRS. Most probably, they are phantom words brought to life by Müller’s erroneous reading of *şirin* ‘sweet; cute’ and its reduplication *şipşirin* – which, incidentally, is also included in his book.

See also *sersebil* below.

sersebil ‘?’ (Müller 2004: 87 &c.) ◊ ? **sebil** ‘1. road; 2. public fountain’

This form is given in Müller 2004: 87 &c. without a source or meaning. It is indeed an exciting challenge to imagine what the intensification of ‘road’ or ‘public fountain’ could be.

serseri ‘vagabond, tramp’ (Müller 2004: 107) ◊ **seri** ‘quick, swift’

Müller 2004: 107 rightly notes that *serseri* is a loanword from Persian and thus not a reduplication. The Persian source is *sarsar*(ī) (سرسر(ی)) ‘careless(ness), inattentive(ness), foolish(ness)’; see e.g. Nişanyan ÇTES.

sersem ‘stunned, bewildered; scatter-brained, foolish’ (Müller 2004: 106)

Although the semantic shift *sem* ‘poison’ → ‘stunned, bewildered’ → ‘scatter-brained, foolish’ seems quite plausible, Müller 2004: 106 is right to note after Stachowski S. 1998: [191] that the word is in fact a loanword from MPers. *sarsām* lit. ‘head inflammation’ → ‘(delusional) fever; delirium; meningitis’, and therefore merely an apparent reduplication; see 3.1.3 for more examples.

tamtakyr ‘completely empty’ (Hatiboğlu 1973, Müller 2004: 100f, 138f) ♦ **takyr** ‘salt lake’, dial. ‘clog’, ‘drying and cracking of soil’, ‘hernia’, ‘tool for making ropes and reins’ • ~ ~ ‘1. dry and hard; 2. knocking, tapping noise like that of hooves’ and ~ *tukur* ‘knocking, tapping noise like that of hooves’

Hatiboğlu 1973 only lists the word among her examples and devotes no more thought to it. Müller 2004: 100f and 139 notes the oddity of semantics and hesitates whether the word is a reduplication, but does not reach any specific conclusion.

One argument for reduplication which Müller seems to have overseen, is the initial accent (*'tamtakyr*). But the word is an emphatic one, and this invites in a natural way a shift of the accent to the initial syllable; see e.g. *'tamters* ‘the exact opposite’ < *tam* ‘perfect; exact’ + *ters* ‘back, reverse, opposite’.

All the different meanings associated with the phonetic shape *takyr* can actually be connected if one sets off from either ‘tapping, knocking noise’ or ‘devoid, destitute’, but it is not possible without a good pinch of imagination. In the present work, I choose to stay on the safe side and exclude the word from further considerations.

žysžybyldak (Hatiboğlu 1973) ♦ **žybyldak** ‘naked, nude’

Both *žybyldak* and its reduplication are dialectal. Hatiboğlu 1973 includes them in the list for what I understand to be literary Turkish, but in the present work it will be excluded. Turkish dialects are a plentiful source of very interestingly diversified reduplications, and they deserve a study in their own right.

The forms *čysčybyldak* and *žysžybyldyk*, which Müller 2004: 219 quotes after Wedel, must either be phonetic variants of *žysžybyldak*, or Wedel’s errors.

2.16.4 Structure

Four closing consonants of *C*-type are attested in a surprisingly high total of 177 examples (see below) derived from 160 unique bases, in a quite uneven distribution:

m: 29 examples: *baška, bejaz, bok, boš, boz, bulanyk, buruš, burušuk, buz, čij, čürük, dazlak, dik, düz, gök, jaš, jassy, ješil, kör, pis, sert, sijah, syk, syky, syžak, tāze, tok, tykyz, and žylk,*

p: 101 examples: *ačyk, ajdyn, ajdynlyk, ajny, ajry, ak, alaža, ansyz(yn), āšīkar, ā(yr, ažy, čevre, čij, čirkin, dar, dā(ynyk, derin, dinč, diri, dolu, durgun, dürü, duru,*

düz, düzgün, ejri, ekşi, erken, eski, genç, geniş, gergin, geže, gündüz, güre, güzel, ibiř, inže, iri, iřsiz, jabanžy, jakyn, jalnyz, janlyř, jaryk, jař, jassy, jařyl, jeni, jeřil, kalyn, kara, karanlyk, katy, kel, kirli, koju, kolaj, kötü, küçük, kuru, kyrmyzy, kysa, kyzyl, lāživert, olgun, ölgün, renkli, sā, sāde, sālam, sary, serin, sijah, silik, řirin, sıvri, sō(uk, syska, syžak, tatly, tāze, temiz, turunžu, tuzlu, ufak, ugun, ujuz, ünlü, urjan, uslu, uzun, užuz, ylyk, yrak, yslak, yssyz, zajyf, zor, žanly, and žylyz,

r: 8 examples: *čabuk, čyplak, perišan, sefil, syklam, temiz, top,* and *topaç,*

s: 38 examples: *bajat, bajā(y, bedāva, belli, berāber, berrak, beter, bulanyk, bütün, čyplak, dejirmi, dolu, dōru, götürü, güdük, jeni, jō(un, jumru, jumuřak, juvarlak, katy, koju, kötürüm, koža, kožaman, kütük, kyvrak, māvi, mor, pembe, tam, tamam, tekerlek, topaç, toparlak, žavlak, žybyl, and žyvyk,* and

ř: 1 example: *beter.*

Seventeen words have more than one closing consonant possible: *beter* (*s* and *ř*), *buřlanyk* (*m* and *s*), *čij* (*m* and *p*), *čyplak* (*r* and *s*), *dolu* (*p* and *s*), *düz* (*m* and *p*), *jař* (*m* and *p*), *jassy* (*m* and *p*), *jeni* (*p* and *s*), *jeřil* (*m* and *p*), *katy* (*p* and *s*), *koju* (*p* and *s*), *sijah* (*m* and *p*), *syžak* (*m* and *p*), *tāze* (*m* and *p*), *temiz* (*p* and *r*), and *topaç* (*r* and *s*).

The combinations are: *m* and *p* (8 words), *p* and *s* (4 words), *r* and *s* (2 words), and *m-s*, *p-r* and *s-ř* (1 word each). No word has more than two closing consonants possible. The most common is *p* (13 words), followed by *m* (9 words), *s* (8 words), and finally *r* (3 words) and *ř* (1 word).

The *ř* in *beřbeter* is highly surprising. It does not appear anywhere else as a closing consonant, except for in Azeri and Kumyk – in both in the very same word; see *betār* in 2.2.2 and *beter* in 2.11.2.

Also unusual is the shape *perperišan*; see 3.1.6 on other reduplications where the closing consonant is the same as *C*₂.

In twelve words, the first vowel is long: *āřikar* ‘obvious, apparent’, *āyr* ‘heavy, serious’, *dāynyk* ‘scattered, dispersed’, *dōru* ‘correct, accurate’, *jōun* ‘dense, intense’, *lāživert* ‘dark blue’, *māvi* ‘blue’, *sā* ‘1. right, ...; 2. alive, ...’, *sāde* ‘simple, ...’, *sālam* ‘sound, solid, ...’, *sōuk* ‘cold’, and *tāze* ‘fresh’. In all it is shortened in the reduplication. Note that the length is secondary in seven cases, and original in five: *āřikar*, *lāživert*, *māvi*, *sāde*, and *tāze* (see Niřanyan ÇTES and Stachowski S. 1998). See also 3.1.20.

Overall, the number of reduplications is unusually high in Turkish. The language with the second most numerous collection is Kazakh with 108 examples, and the gap between the two is the only one so big across all the Turkic languages. Müller’s 2004 questionnaire (see appendix A) showed that on average, Turkish students only knew 84.2% of a hundred reduplications they were presented with. Combined, these two pieces of information allow to presume that only a part of the 177 examples listed here are actually commonly known and used across the entire Turkish territory, and

to estimate this part to be about a half of the collection – or a little less, taking into account that it did not always had to be the same 84.2% that the students knew.

This is not to say that the attestations of the other half can be deemed untrue. Rather, they are simply very rare or geographically limited forms, which only found their way to linguistic descriptions because of the relatively much attention that has been paid to Turkish reduplications. See also 1.2.1 on the general availability of sources, and 3.2.2 for a comparison of Turkish with other languages.

It should be also noted that thanks to this observation, it is to some point possible to explain the large gap between the 177-strong Turkish collection and the humble Ottoman set of 58 reduplications.

See 3.4.6 for another argument supporting these conclusions.

2.16.5 Semantics

In almost all the cases, the reduplicated meaning is a simple intensification of the base meaning, or apparently the same. As for the latter, finer distinctions might have been lost in translation, or the strength of the reduplication might have somewhat faded over the years due to overuse.

Apart from *köskütük* &c. (see 2.16.3 above), the observed semantic shifts are few and simple, as e.g. in *kyorak*, where ‘lithe, brisk, agile, dexterous’ → ‘strongly, firmly’.

Even Wedel’s examples excluded, almost all cases have a primarily adjectival or adverbial character, as far as the distinction is valid in Turkic. The strongest substantival component can be observed in: *ajdynlyk* ‘light(ness), illumination’, *buruş* ‘wrinkle, crease’ (along *buruşuk* ‘wrinkled, creased’), *geže* ‘(at) night’, *gündüz* ‘(at) daytime’, *kütük* ‘tree-stump; baulk; log’, *top* ‘1. ball; 2. round’, *topaç* ‘1. (spin)top, teetotum, whirligig; 2. ~ *gibi* sturdy’, and *ujuz* ‘itch, mange, scab, mangy, scabby’.

2.17 Turkmen

For an Oghuz language, Turkmen *C*-type reduplications are neither particularly numerous nor diversified. Only Gagauz has less examples, but then the data available for Gagauz are also much more modest, which likely distorts the picture.

Three closing consonants are attested. The domination of *p* is less pronounced than in the majority of languages. Interestingly, all words which have more than one closing consonant possible, always only have *p* and *s* for alternatives. See 2.17.4.

Semantically, Turkmen reduplications are very standard with effectively no non-adjectives and no far or unusual semantic evolutions. See 2.17.5.

2.17.1 Sources

The main source of the material is TrkmRS which contains ca. 40 000 entries. Some attestations have also been found in Ščerbak 1977: 120, Schönig 1998c: 264, Zeynalov 1993: 149f, and the two grammars listed below. Long vowels are marked based primarily on TrkmRS.

For a grammatical description, Clark 1998: 150, 510 has been used, as Hanser 1977: 176 and 2003: 176 are effectively reduced to a short list of examples.

Reduplication is discussed in two places in Clark 1998. On p. 150, only a brief mention can be found along with three examples. The body of the description is on p. 510, where more examples are given, and the information that the closing consonant is *p*. This must be an overlooking, as *dosdogry* is among the examples on p. 150; reduplications closed by *m* do not seem to be included in the book. Also not mentioned in the commentary, but evident from examples is the shortening of the reduplicated vowel. The next section (p. 511) is perhaps more intriguing as it discusses what apparently are severed and emancipated reduplicated anlauts (see 3.1.10).

Transcription:

Clark 1998: $V: \rightarrow \bar{V} \mid \delta \rightarrow z \mid \iota \rightarrow y \mid \check{j} \rightarrow \check{z} \mid \theta \rightarrow s \mid y \rightarrow j$,
 Hanser 1977: $V: \rightarrow \bar{V} \mid \vartheta \rightarrow e \mid \check{g} \rightarrow g \mid \theta \rightarrow \ddot{o} \mid \theta \rightarrow s$,
 Hanser 2003: $V: \rightarrow \bar{V} \mid \varsigma \rightarrow \check{c} \mid \vartheta \rightarrow e \mid \check{g} \rightarrow g \mid \theta \rightarrow \ddot{o} \mid \varsigma \rightarrow \check{s} \mid x \rightarrow h \mid w \rightarrow v$,
 Ščerbak 1977: $\bar{y} \rightarrow g \mid i \rightarrow y \mid \check{j} \rightarrow \check{z} \mid \theta \rightarrow s \mid \check{y}\check{j} \rightarrow \bar{u}$,
 TrkmRS: $\vartheta \rightarrow \ddot{a} \mid \eta \rightarrow \eta \mid \theta \rightarrow \ddot{o} \mid \gamma \rightarrow \bar{u} \mid \gamma\check{u} \rightarrow \bar{u} \mid \varkappa \rightarrow \check{z}$,
 Zeynalov 1993: $c \rightarrow \check{z} \mid \varsigma \rightarrow \check{c} \mid \iota \rightarrow y \mid \varsigma \rightarrow \check{s} \mid \check{u}y \rightarrow \bar{u} \mid y \rightarrow j$.

2.17.2 Standard cases

ājdyň i.a. ‘moonlit, light, bright’ \diamond **apājdyň** ‘very light, very bright’ (TrkmRS)

āk ‘white’ \diamond **apāk** (TrkmRS, Zeynalov 1993: 150, Clark 1998: 150, 510)

See *apāk* in 2.17.3 below, and also 3.1.10 on other intensifications of the word.

āňsat ‘light, easy’ \diamond **apāňsat** (TrkmRS)

arassa ‘unspotted, clean, neat’ \diamond **aparassa** (TrkmRS)

arkajyn ‘1. quiet, peaceful; 2. free, unimpeded’ \diamond **aparkajyn** ‘1. very confidently, very surely; 2. very quietly, very peacefully’ (TrkmRS)

āžy ‘bitter’ \diamond **apāžy** (TrkmRS, Ščerbak 1977: 120, Zeynalov 1993: 150)

belli i.a. ‘clear, obvious’ \diamond **besbelli** ‘completely clear, completely obvious, absolutely certain’ (TrkmRS)

beter ‘particularly, more so’ \diamond **besbeter** ‘1. more so, very; 2. worst of all’ (TrkmRS)

bütin ‘all, whole’ \diamond **büsbütin** ‘completely, entirely’ (TrkmRS)

čāl ‘grey’ \diamond ***čapčāl** (Zeynalov 1993: 150: <çapçal>)

See **čapčāl* in 2.17.3 below.

- dajav** ‘strong, sturdy, healthy’ ◇ *dapdajav* ‘very hefty’ (TrkmRS)
den ‘equal, identical’ ◇ *depden* (TrkmRS) ◇ *desden* (TrkmRS)
dik ‘steep’ ◇ *dimdik* (Zeynalov 1993: 150, TrkmRS)
dīri ‘(a)live’ ◇ *disdīri* (TrkmRS)
dogry ‘1. smooth, straight; 2. accurate, just’ ◇ *dopdogry* (TrkmRS) ◇ *dosdogry* (TrkmRS, Hanser 1977: 176, Clark 1998: 150, Hanser 2003: 176)
dōly ‘full’ ◇ *dosdōly* (TrkmRS)
dury ‘clean, clear’ ◇ *dupdury* (TrkmRS)
düz i.a. ‘1. smooth, straight; 2. true, faithful, accurate’ ◇ *dupdüz* (TrkmRS)
esli ‘considerable, significant’ ◇ *epesli* (TrkmRS, Hanser 1977: 176, 2003: 176²⁵)
gadyrly ‘dear, close, respected’ ◇ *gapgadyrly* ‘dear, cordial’ (TrkmRS)
gara ‘black’ ◇ *gapgara* (TrkmRS, Zeynalov 1993: 150, Clark 1998: 510)
 See also 3.1.10 on other intensifications of the word.
garangky ‘dark(ness)’ ◇ *gapgarangky* (TrkmRS, Clark 1998: 510)
 See also 3.1.10 on other intensifications of the word.
geñsi ‘remarkable, good’ ◇ *gepgeñsi* (TrkmRS)
geñsilik ‘beaut(ifull)y’ ◇ *gepgeñsilik* ‘great, very well, wonderfully’ (TrkmRS)
gīñ ‘wide’ ◇ *gipgīñ* (TrkmRS)
gök ‘1. blue; 2. green’ ◇ *gömgök* ‘1. very blue; 2. light blue; 3. very green (TrkmRS); completely green (Schönig 1998c: 264); very blue (Hanser 1977: 176, 2003)’
göni ‘straight, direct’ ◇ *gösgöni* (TrkmRS, Zeynalov 1993: 150)
govy ‘good’ ◇ *gopgovy* (TrkmRS)
gūry ‘dry’ ◇ *gupgūry* (TrkmRS) ◇ *gusgūry* (TrkmRS)
gyrmyzy ‘red’ ◇ *gyppgyrmyzy* (TrkmRS, Clark 1998: 510)
jagty ‘light, bright’ ◇ *japjagty* (TrkmRS)
jany ‘recently, only just’ ◇ *japjany* (TrkmRS) ◇ *jasjany* (TrkmRS)
jāşyl ‘green’ ◇ *japjāşyl* (Clark 1998: 510)
jeñil ‘light’ ◇ *jepjeñil* (TrkmRS, Clark 1998: 510)
sāry ‘yellow’ ◇ *sapsāry* (TrkmRS, Hanser 1977: 176, Clark 1998: 150, 510, Hanser 2003: 176)
semiz ‘fatty, greasy’ ◇ *sepsemiz* (TrkmRS)
sovuk ‘cold, frosty, chilly’ ◇ *sopsovuk* (TrkmRS)
sūži ‘sweet’ ◇ *süpsüži* (TrkmRS, Ščerbak 1977: 120, Zeynalov 1993: 150, Clark 1998: 510)
tajjar ‘ready, completed’ ◇ *taptajjar* (TrkmRS)
takyr i.a. ‘smooth, bare, blank, empty’ ◇ *taptakyr* (TrkmRS, Clark 1998: 510)
tämiz ‘clean’ ◇ *täptämiz* (TrkmRS)

²⁵ In Hanser 1977: 176, the meaning is given as ‘very long’ on p. 176 and as ‘very much, considerably’ on p. 261. The former must be mistake and is corrected in Hanser 2003: 176 with the following comment: “Hanser, „çok uzun (very long) anlamını vermiş, ama sözlüklerde sadece „çok” anlamı var”.

täze ‘new’ ◊ **täptäze** (TrkmRS)

tegelek ‘round’ ◊ **testegelek** (STrkmJa, TrkmRS)

See **teptegelek* in 2.17.3 below.

tekiz ‘smooth, even’ ◊ **teptekiz** (STrkmJa, TrkmRS)

See **teptegelek* in 2.17.3 below.

togalak ‘round’ ◊ **tostogalak** (TrkmRS, Ščerbak 1977: 120)

turşy ‘sour, tart’ ◊ **tupturşy** (Zeynalov 1993: 150)

ullakān ‘big, huge’ ◊ **upullakān** (TrkmRS)

uzyn ‘long’ ◊ **upuzyn** (TrkmRS, Zeynalov 1993: 150: *upuzun*)

yssy ‘hot, sultry’ ◊ **ypyssy** (Zeynalov 1993: 150)

2.17.3 Special cases

āpāk (Hanser 1977: 176, 2003: 176) ◊ **āk** ‘white’

In both editions of Hanser’s grammar (1977 and 2003), *āpāk* is given with the initial, reduplicated *ā* not shortened. This is rather suprising as it goes against not only the attestations in Clark 1998: 150 and 510 (⟨ap-a:k⟩) and the remaining twelve examples in Turkmen, but also against the general Turkic rule, undisputable exceptions to which can only be found in Dolgan and Yakut (see 3.1.20). Unfortunately, the word seems to be missing, or is given without indication of vowel length, from STrkmJa and TrkmRS.

In the present work, it will be assumed that the shapes given by Hanser are misprints.

***čapčāl** (Zeynalov 1993: 150: ⟨çapçal⟩) ◊ **čāl** ‘grey’

Unfortunately, Zeynalov 1993 does not mark vowel length, and TrkmRS only attests the base word but not the reduplication. Shortening of the reduplicated vowel is more likely than not as in eleven out of the twelve remaining examples it has been shortened, and the last one is probably a misprint (*āpāk* above). See 2.17.4 below and also 3.1.20 on vowel shortening in general.

čypčynym ‘absolute truth, sheer truth’ (TrkmRS, Clark 1998: 510) ◊ ? **čyn** ‘truth, true’

The shape **čynym* is missing as a separate word from STrkmJa, TrkmRS and TrkmTS. It makes the impression of being a composition of *čyn* ‘truth, true’ with a Px1Sg, but in that case, the function of this alleged suffix would be incomprehensible to me.

Also, if this word is a reduplication, it is the only one of a noun in Turkmen.

The form is not clear and will be omitted from further considerations so as to not point to false conclusions.

Interestingly, *čyp* can also be used as an intensifying particle in a composition with *jalaŋač* ‘naked, bare’ (see 3.1.10).

gumgūkluk ‘complete silence’ (TrkmRS)

The base *gūk* seems to only be attested in the phrases *gūk bermek* and *gūk dijmek* ‘подавать голос, откликаться (при игре в прятки); 2. перен[осное значение] сообщать, давать знать о чем-л.; 3. перен. заходить к кому-л., куда-л., навешать кого-л. (изредко)’ (TrkmRS). The semantic shift from these meanings to ‘complete silence’ is definitely not trivial, but possibly not without a parallel example in Tuv. *šipšimēn* (see 2.18.4).

***teptegelek** (? TrkmRS) \diamond **tegelek** ‘round’

With two bases, *tegelek* ‘round’, and *tekiz* ‘smooth, even’, the wording in TrkmRS is quite unclear, and I am not certain how many reduplications either of them has:

ТЕП- см. **тес-**; \sim -**текиз** ровный-преровный.

ТЕС- (**теп-**) препозитивная усил[ительная] частица, присоединяемая к словам, начинающимся на букву «т»: \sim -**тегелек** круглый-прекруглый, совершенно круглый; [the entry continues but without mentioning *tekiz*]

Two forms are explicitly attested, then, *teptekez* and *teptegelek*, and both are confirmed by STrkmJa. Two more can be suspected, *testekiz* and *teptegelek*, but as they are uncertain, and also apparently missing from STrkmJa, they will be ignored in further considerations.

2.17.4 Structure

Three closing consonants of *C*-type are attested in a total of 53 examples derived from 48 unique bases, in a relatively uneven distribution:

m: 2 examples: *dik*, and *gōk*,

p: 39 examples: *ājdyn*, *āk*, *āṅsat*, *arassa*, *arkajyn*, *āžy*, *čāl*, *dajav*, *deṅ*, *dogry*, *dury*, *düz*, *esli*, *gadyrly*, *gara*, *garangky*, *geṅsi*, *geṅsilik*, *gīṅ*, *govy*, *gūry*, *gyrmyzy*, *jagty*, *jaṅy*, *jāšyl*, *jeṅil*, *sāry*, *semiz*, *sovuk*, *sūži*, *tajjar*, *takyr*, *tāmiz*, *tāze*, *tekiz*, *turšy*, *ullakān*, *uzyn*, *yssy*, and

s: 12 examples: *belli*, *beter*, *bütin*, *deṅ*, *dīri*, *dogry*, *dōly*, *gōni*, *gūry*, *jaṅy*, *tegelek*, and *togalak*.

Four words have more than one closer possible: *deṅ* ‘equal, identical’, *dogry* ‘1. smooth, ...; 2. accurate, ...’, *gūry* ‘dry’, and *jaṅy* ‘recently, only just’, and possibly also *tegelek* ‘round’ and *tekiz* ‘smooth, even’ belong here (see **teptegelek* in 2.17.3 above). In all cases, the alternatives are *p* and *s*.

In thirteen cases, the first vowel of the base is long: *ājdyn* i.a. ‘moonlit, ...’, *āk* ‘white’, *āṅsat* ‘light, easy’, *āžy* ‘bitter’, *čāl* ‘grey’, *dīri* ‘(a)live’, *dōly* ‘full’, *gīṅ* ‘wide’, *gōk* ‘1. blue; 2. green’, *gūry* ‘dry’, *jāšyl* ‘green’, *sāry* ‘yellow’, and *sūži* ‘sweet’. In eleven, the reduplicated vowel has been shortened; *āk* and *čāl* are not entirely clear (see 2.17.3). See also 3.1.20.

2.17.5 Semantics

Almost all reduplicated meanings are simple intensifications or the same as their respective base meanings. Perhaps only *arkajyn* shows a slightly further development: ‘1. спокойный, тихий, лишённый тревог; 2. свободный, беспрепятственный’ → ‘1. очень уверенно; 2. совершенно спокойно’ (TrkmRS).

As far as such a statement can be made with reference to a Turkic language, all examples are adjectives.

2.18 Tuvinian

Compared to most South Siberian Turkic languages, Tuvinian *C*-type reduplications appear to be relatively numerous, on par only with Khakas. Indeed, K.D. Harrison goes as far as to declare full productivity of reduplication across several parts of speech. This claim seems exaggerated; see 2.18.2.

Only *p* is attested as a closing consonant. Harrison’s examples aside, all words are clear adjectives except for one, which has in addition a substantival meaning. See 2.18.5 and 2.18.6.

Beside the standard ‘pitch-black’, *kara* also reduplicates to ‘dear’ and possibly ‘world’, and can be found in the rather unusually built *kap-la kara*. See 2.18.4.

2.18.1 Sources

The main source of the material is TuvRS which contains ca. 22 000 entries. Some attestations have also been found in Krueger 1997: 77, Li et al. 2007, Ölmez 2007, Pormorska 2004: 144, Ščerbak 1977: 120, Schönig 1998b: 408, Serebrennikov/Gadžieva 1986: 112, and the grammatical descriptions listed below. The data provided by K.D. Harrison are dealt with separately in 2.18.2 below.

For grammatical descriptions, Palmbah 1955: 656, Ishakov/Palmbah 1961: 187, Èrgil-ool 1993: 109 and Takashima 2008 have been used. See also 2.18.2 below on Harrison’s data.

Palmbah 1955: 656 is rather brief on reduplication. Ishakov/Palmbah 1961: 187 are less so and, along with some examples, they provide a description of the process and the statement that the closing consonant is *p*.

Likewise, Èrgil-ool 1993: 109 and Takashima 2008: 83f, 102 are quite condensed, and limited effectively to several examples and a brief description of the process with the information that reduplicated anlauts are closed by *p*.

Transcription:

Krueger 1997: $ɳ \rightarrow \eta \mid VV \rightarrow \bar{V}$,

Harrison 2000, 2004: $\dot{i} \rightarrow y \mid u \rightarrow u \mid q \rightarrow k \mid x \rightarrow h$,

Harrison/Raimy 2004: $\gamma \rightarrow g \mid \dot{i} \rightarrow y \mid t\dot{f} \rightarrow \check{c} \mid q \rightarrow k \mid VV \rightarrow \bar{V} \mid x \rightarrow h$,
 Ščerbak 1977: $\bar{h} \rightarrow g \mid \dot{i} \rightarrow y$,
 Schönig 1998b: $q \rightarrow k$,
 Serebrennikov/Gadžieva 1986: $\varepsilon \rightarrow g$,²⁶
 TuvRS: $\mu \rightarrow \eta \mid \theta \rightarrow \ddot{o}$.

2.18.2 K.D. Harrison

Altogether, three works authored or co-authored by K.D. Harrison have been considered: Harrison 2000 and 2004, and Harrison/Raimy 2004. Harrison's data are not universally noted for fidelity and credibility; see e.g. Salminen 2006 or Pomorska 2010, also Helimski 2001. I have adopted here an accordingly cautious attitude. In fact, since much of it is rather sensational, and extraordinary claims require extraordinary evidence, I generally chose to ignore it in lack of the latter. Exceptions have only been made for four much more standard cases.

The two most important claims made by K.D. Harrison are that: 1. Tuvian verbs can be reduplicated, and 2. reduplication is “fully productive across several word classes in Tuvan”.

The claim of verb reduplication is thus formulated in Harrison 2000: 158:

In Tuvan, unlike in any other Turkic language, CVp- reduplication has become fully productive for verbs as an aspectual marker. This reduplication may apply to various tenses to add an aspectual connotation of rapid or intense action [...]. It may also indicate unexpected or sudden action [...]. In other instances, it simply adds a meaning of strong assertion to the verb [...].

A partially compatible description is given in Harrison/Raimy 2004, where the semantics are defined as “‘emphatic’ for modifiers, and ‘emphatic’, ‘intensifying’ or ‘iterative’ for verbs”.

I will forgo here the terminological discussion on Harrison's understanding of the term *aspect* and instead, will look in a little more detail into semantics. Despite the advertised full productivity, he gives no more than three examples of reduplicated verbs:

haly- ‘to run’ \diamond ***haphaly-** only attested in oblique forms (Harrison 2000: 158, Harrison/Raimy 2004) • *haphalān* ‘ran really fast’ (Harrison/Raimy 2004), *haphalān men* ‘I was driving fast’, *haphaladym* ‘I drove fast’ (Harrison 2000: 158)
kör- ‘to see’ \diamond ***köpkör-** only attested in oblique forms (Harrison 2000: 158, Harrison/Raimy 2004) • *köpkörbēn* ‘did not see at all’ (Harrison/Raimy 2004). *köpkörbēn men* ‘I have never seen at all’, *köpkördüm* ‘I saw! *unexpectedly/suddenly*’ (Harrison 2000: 158)

²⁶ Only in *saryg* \rightarrow ‘yellow’.

sag- ‘to milk’ ◊ ***sapsag-** only attested in oblique forms (Harrison 2000: 158, Harrison/Raimy 2004) • *inekti sapsār men* ‘I will definitely milk the cow’, *inekti sapsagbas men* ‘I will definitely not milk the cow’ (Harrison 2000: 158), *sapsagbas* ‘will definitely not milk’, *sapsār* ‘will definitely milk’ (Harrison/Raimy 2004)

Thus, the semantic changes are as follows ‘to run’ → ‘to run fast’, ‘to see’ → ‘to see unexpectedly or suddenly’, ‘to not see’ → ‘to not see at all’, and ‘to milk’ → ‘to definitely milk’. They match very closely the description given in Harrison 2000: 158 (quoted above). One might be even surprised that it was possible to find examples illustrating so accurately and fully the general meaning of reduplication of Tuvinian verbs. The 2004 description is clearly an abstraction, albeit apparently based on a partially different set of examples. Unfortunately, I was not able to find any more than the three listed above.

The possibility of verb reduplication does not seem to be mentioned in any of the grammars used here.

The claim of full productivity is particularly difficult to evaluate. The number of examples that can be found in TuvRS seems to speak against it. A dictionary of that size, however, cannot be expected to contain very many more. Nevertheless, the claim is rather unusual for a Turkic language, and I expect that at least one grammar book should like to make a note of it. As it seems, neither of the three employed here did.

Summing up, if Harrison’s data were accurate, they would constitute a very clear evidence of the very independent and unique path that the evolution of reduplication has taken in Tuvinian. For them to be believable, however, more evidence needs to be published, and preferably, by more than one author. I think that even the modest collection extracted mainly from TuvRS and presented here, shows sufficiently the autonomy of Tuvinian reduplication.

2.18.3 Standard cases

ak ‘white’ ◊ **apak** (Paľmbah 1955: 656, Ishakov/Paľmbah 1961: 187, Takashima 2008: 83)

borbak i.a. ‘clot, ball; round, spherical’ ◊ **bopborbak** (Paľmbah 1955: 656, TuvRS, Èrgil-ool 1993: 109, Harrison 2004: 201, Harrison/Raimy 2004, Takashima 2008: 83)

čā ‘1. new; 2. recently, only just’ ◊ **čapčā** (Paľmbah 1955: 656, Ishakov/Paľmbah 1961: 187, TuvRS, Ščerbak 1977: 120, Pomorska 2004: 144, Li et al. 2007)

čiŋge ‘thin’ ◊ **čipčiŋge** (Harrison 2000: 158, Harrison/Raimy 2004)

deŋ ‘equal, identical’ ◊ **depdeŋ** (TuvRS)

deski ‘smooth, even’ ◊ **depdeski** (TuvRS)

dorān ‘at once, immediately’ ◊ **dopdorān** (Paľmbah 1955: 656, TuvRS, Ölmez 2007, Takashima 2008: 102)

- henerten** ‘suddenly, unexpectedly’ ◇ **hēphenerten** (Paľmbah 1955: 656, TuvRS)
- kara** ‘black’ ◇ **kapkara** ‘1. *intens.* (Paľmbah 1955: 656, Ishakov/Paľmbah 1961: 187, TuvRS, Schönig 1998b: 408, Harrison/Raimy 2004, Ölmez 2007, Takashima 2008: 83); 2. dear (TuvRS)’
See *kap-la kara*, and *kapkara* ‘dear’ and ?‘world’ in 2.18.4 below.
- kök** ‘blue’ ◇ **köpök** (Takashima 2008: 83)
- kurug** ‘empty’ ◇ **kupkurug** (Takashima 2008: 84)
- kyzyl** ‘red’ ◇ **kypkyzyl** ‘bright red’ (Paľmbah 1955: 656, Ishakov/Paľmbah 1961: 187, TuvRS, Èrgil-ool 1993: 109, Krueger 1997: 77, Harrison 2000: 158, Harrison/Raimy 2004, Takashima 2008: 83)
- nogān** ‘green’ ◇ **nopnogān** (Harrison/Raimy 2004)
- saryg** ‘yellow’ ◇ **sapsaryg** ‘very yellow, bright yellow’ (Paľmbah 1955: 656, Ishakov/Paľmbah 1961: 187, TuvRS, Ščerbak 1977: 120, Serebrennikov/Gadžieva 1986: 112, Èrgil-ool 1993: 109, Takashima 2008: 83)
- tögerik** ‘round’ ◇ **töptögerik** (TuvRS)
- tura** participle of *turar* ‘to stand up’ (TuvRS: only attested in ~ *halȳr* ‘to jump, to spring, to start’) ◇ **tuptura** (TuvRS: only attested in ~ *halȳr* ‘to immediately get on one’s feet’)
- türgen** ‘quickly’ ◇ **tüptürgen** (Harrison 2000: 158, Harrison/Raimy 2004)
- uzun** ‘long’ ◇ **upuzun** (Harrison 2000: 158, Harrison/Raimy 2004)

2.18.4 Special cases

kapkara ‘dear’ (TuvRS) ◇ **kara** ‘black’

The shift ‘black’ > ‘dear’ is rare and might surprise, but it is not impossible. Laude-Cirtautas 1961: 20 collects examples from different Turkic languages where phrases such as ‘black-eyed’ and similar are used with positive meanings (‘beautiful, attractive’, ‘my dear’ when with a Px1Sg), which she explains by that black eyes, especially in girls, are often considered to be particularly beautiful. An omission of the word ‘eye’ from the phrase seems to be quite conceivable in a clear (cultural) context, and is probably what happened in our Tuvinian example.

See also *kap-la kara* below.

kapkara ?‘world’ (Ölmez 2007) ◇ ? **kara** ‘black’

The entry in Ölmez 2007 is not entirely clear:

kap-kara → **kara-kara kaptagay** dünya, evren [...] Etim.: Mo[ngolisch].
ķabtağay “düz; tahta; yassı düzey” [...]

It is my understanding that *kapkara* can be used interchangeably with *kara-kara* as an attributive for *kaptagaj*. Laude-Cirtautas 1961: 34 lists a series of examples where *kara* acts as an intensifier with effectively no reference to ‘blackness’ whatsoever, e.g. Oir. *kara janyš* ~ Sag., Shor *kara čagyš* ‘completely (a)lone’,

Tat. *kara karşy* ‘exactly opposite’,²⁷ etc. Perhaps this form is also to be understood in the same way?

kap-la kara ‘dear’ (TuvRS) \diamond **kara** ‘black’

Beside the standard *kapkara*, TuvRS also attests the phrase *kap-la kara meŋ ežim* ‘my dear friend’. While multiple intensification is not a rarity in itself (see 3.1.13 for more examples), this is the only example of it being created in this way. Perhaps a contraction of **ka-p.yl-a-kara* (as in Tksh. *čy-r.yl-čyplak* ‘stark-naked’), combined with a CV-type reduplication as in *gü-p.e-gündüz* ‘in broad daylight’? Or possibly a Mongolian influence, see *dala dalbagaar* &c. in 3.4.3?

See also *kapkara* ‘dear’ above.

sapsajtyk \sim **sypsyjtyk** ‘1. imitation of a gopher’s scream; 2. a chirring sound made by a rabbit or hare; 3. chirping or calling of a magpie; 4. squeaking of a mouse’ (Harrison 2004: 201)

TuvRS only attests *sajt*, and only in \sim *dēr* ‘to chirr of a magpie’. The reduplication is therefore plausible, but as the base does not seem to be attested on its own, the whole must be considered a special case here and ignored in further considerations.

šipšimēn ‘silence, peace’ (TuvRS) \diamond ? **šimēn** ‘noise, clamour’

If *šipšimēn* really is the reduplication of *šimēn*, then it has undergone a peculiar semantic evolution which I am not able to explain. Possibly, however, a parallel example can be found in Trkm. *gumgūkluk* (see 2.17.3).

šypšyk ‘most’²⁸ (TuvRS, Ölmez 2007)

The form and meaning of this word strongly resemble a reduplication. The shape *šyk*, however, does not seem to be attested other than with the meanings ‘1. wet, moisture; 2. meadow, pasture’, which hardly can be considered a match here. Perhaps merely an apparent reduplication (see 3.1.3 for other examples)?

2.18.5 Structure

Only one closing consonant of C-type is attested in a total of 18 examples derived from 18 unique stems. See 2.18.2 above for claims of verb reduplication and of full productivity of reduplication.

p: 18 examples: *ak, borbak, čā, činge, deŋ, deski, dorān, henerten, kara, kök, kurug, kyzyl, nogān, saryg, tögerik, tura, türgen, and uzun.*

²⁷ This example is less clear. In theory, the form could also be a CV-type reduplication with the closing consonant identical to C_2 . This occurs sometimes in C-type reduplications (see 3.1.6), but I am only aware of one such example in the CV-type, namely Yak. *sürüsürdēχ* ‘absolutely: fearsome, dreadful’ (Pekarskij 1907–30).

²⁸ TuvRS: ‘самый (о высшей точке чего-л.)’, Ölmez 2007: ‘en, pek (birşeyin en yüksek noktası için) || sehr (die Spitze von irgend etwas)’. TuvRS also gives two examples: *dagnyŋ* \sim *bažy* ‘the very top of a mountain’, and *yjaštyŋ* \sim *bažy* ‘the very top of a tree’.

In one word, *čā* ‘1. new; 2. recently, only just’, the first vowel of the stem is long. It has been shortened in the reduplication. See also 3.1.20 on the phenomenon in general.

2.18.6 Semantics

Most reduplicated meanings are, predictably, simple intensifications or the same as their respective base meanings. Two points, however, are interesting about *kara* ‘black’, see *kapkara* ‘dear’ and [?]‘world’ in 2.18.4 above.

All words are quite clear adjectives, only *borbak* additionally has a stronger substantival component in its meaning (‘clot, ball; round, spherical’). K.D. Harrison believes reduplication is fully productive across several parts of speech in Tuviniian, including even verbs; see 2.18.2.

2.19 Uighur

The present section only collects modern Uighur material; the few reduplications attested in Old Uighur are listed in 2.1. The modern examples of the *C*-type are neither particularly numerous nor diversified. They are fewer than in Uzbek and decidedly more uniform, a fact which cannot be blamed on insufficient sources.

Technically, three closing consonants are attested: *p* in 33 examples, *pp* in one, and *v* in one. The latter is dialectal, and obviously a phonetic variant of *p*. A noteworthy characteristic is that reduplicated is the pre-umlaut shape of the vowel, so that e.g. *ješil* ‘green’ → *japješil* instead of **japjašil*. See 2.19.4.

Semantically, Uighur reduplications are quite standard. They are all nomina and only rarely non-adjectives. The reduplicated meanings are generally simple intensification, with perhaps the only notable exception being *aq* ‘white’. See 2.19.5.

2.19.1 Sources

The main source of the material is UjgRS which contains ca. 33 000 entries. Some attestations have also been found in Baskakov 1978, Jarring 1964, Malov 1954, Menges [1936], Ščerbak 1977: 120, and in the grammars listed below.

For grammatical descriptions, Sadvakasov 1976: 156, Kaşgarlı 1992: 126, Öztürk 1994: 58, Friederich/Yakup 2002: 31, Tömür 2003: 110f, De Jong 2007: 60 and Tursun 2007: xviii have been used.

All are imprecise in that they speak of the reduplication of the (entire) first syllable, or even of the root (Tursun 2007: xviii). Kaşgarlı 1992: 126 and De Jong 2007: 60 report that the closing consonant *p* is added to the reduplicated syllable which ends in a vowel; neither specifies what happens when it ends in a consonant. Also Öztürk 1994: 58 and Tömür 2003: 110f mention that the closing consonant is *p*.

De Jong 2007: 60 and Tömür 2003: 110f limit the scope of the phenomenon to adjectives expressing colour and outward appearance. The latter also mentions that reduplication is not as productive as the Decreasing Degree of the adjective. Sadvakasov 1976: 156 effectively does not go beyond listing a few examples.

Only Friederich/Yakup 2002: 31 devote a little more space to reduplication. Surprisingly, they do not mention the term but rather speak of a prefix which “ähneln in seiner Lautung immer dem eigentlichen Adjektiv”. One important piece of information they provide is that the vowel of the reduplicated anlaut reflects a pre-umlaut vowel of the root, as can be observed in *sapseriq* ‘very yellow’.

Transcription:

Baskakov 1978: $\kappa \rightarrow q$,

Jarring: the focus of the transcription in Jarring 1933 and 1964 is on phonetics rather than phonology; all references are given here together with the original notation,

Kaşgarlı 1992: $\zeta \rightarrow \check{c} \mid e \rightarrow \check{a} \mid \acute{e} \rightarrow e \mid \check{k} \rightarrow q \mid \check{s} \rightarrow \check{s}$,

Malov 1954: $\ddot{a} \rightarrow \check{a} \mid j \rightarrow j \mid k \rightarrow k \mid \kappa \rightarrow q \mid l, \lambda \rightarrow l$,

Öztürk 1994: $\zeta \rightarrow \check{c} \mid e \rightarrow \check{a} \mid \acute{e} \rightarrow e \mid \acute{g} \rightarrow \gamma \mid \check{k} \rightarrow q \mid \check{n} \rightarrow \eta \mid \check{s} \rightarrow \check{s} \mid y \rightarrow j$,

Sadvakasov 1976: $\kappa \rightarrow q$,

Tömür 2003: $\text{خ} \rightarrow \chi \mid \text{ه} \rightarrow h$,

UjgRS: $e \text{ ي} \rightarrow e \mid \text{ا} \text{ ئ} \rightarrow \check{a} \mid \text{ا} \text{ غ} \rightarrow \gamma \mid x \text{ خ} \rightarrow h \mid \kappa \text{ ك} \rightarrow k \mid \kappa \text{ ق} \rightarrow q \mid \eta \text{ گ} \rightarrow \eta \mid o \text{ و} \rightarrow o \mid \text{ه} \text{ و} \rightarrow \check{o} \mid y \text{ و} \rightarrow u \mid y \text{ و} \rightarrow \check{u} \mid \text{ж} \text{ ج} \rightarrow \check{z}$.

2.19.2 Standard cases

The list below contains many dialectal variants. Where the differences between them are small, the main entry is the same as in UjgRS, and only the alternative reduplications are all listed after the \diamond symbol (so e.g. in *jeŋi* ‘new’ or *ješil* ‘green’). Where the differences are greater, the forms are broken up into separate entries (so e.g. in *teč* : *tinč* ‘quiet, peaceful’).

aq ‘white’ \diamond **apaq** ‘1. intens. (Jarring 1964: ⟨ap aq, âp aq, ap aχ⟩, Tömür 2003: 111); 2. address to a respected woman (Malov 1954)’ \diamond **appaq** ‘1. intens. (Malov 1954, UjgRS, Sadvakasov 1976, Baskakov 1978, Kaşgarlı 1992: 126²⁹, Öztürk 1994: 58); 2. Frau der kleinen *Bäge* und der Groß-*Mullā*’ (Menges [1936])’ • *appiyim* ‘my dear, my darling’ (UjgRS) ~ *appyyym* ‘my little white one’ (Malov 1954)
See ⟨apáq⟩ in 2.19.3 below.

aškara ‘clear, obvious, bright, open’ \diamond **apaškara** (Öztürk 1994: 58)

baravär ‘equal(ly), identical(ly)’ \diamond **bapbaravär** (UjgRS, Kaşgarlı 1992: 126, Tömür 2003: 110)

²⁹ Besides *appaq* (⟨appaq⟩), Kaşgarlı 1992: 126 also gives *appak* (⟨appak⟩), which is probably a misprint. Similar mistakes might be suspected in the case of *ješil* and *saq*.

- boš** ‘empty’ ◇ *bopboš* (UjgRS, Kaşgarlı 1992: 126, Öztürk 1994: 58)
- jeñi** ‘new’ ◇ *jepjeñi* (Tömür 2003: 110) ~ *jipjiñi* (Kaşgarlı 1992: 126)
- jaš** ‘young’ ◇ *japjaš* (Sadvakasov 1976: 156)
- ješil** ‘green’ ◇ *japješil* (Jarring 1964: ⟨jâp ~ jep + ješil ~ ješil ~ jišil, Kaşgarlı 1992: 126³⁰, Tömür 2003: 111) ~ *jäpješil* (Öztürk 1994: 58)
See 2.19.4 below.
- joruq** ‘light, bright’ ◇ *jopjoruq* (UjgRS)
- jumulaq** ‘spherical, round’ ◇ *jopjumulaq* (UjgRS)
See 2.19.4 below.
- kök** ‘blue’ ◇ *köpkök* (Kaşgarlı 1992: 126, Öztürk 1994: 58, De Jong 2007: 60) ~ *küpkük* (Tömür 2003: 111)
- kökläk** ‘green field’ ◇ *köpköklä | ik* ‘rich verdure’ (Jarring 1964: ⟨köp kökläk, köp köklik⟩)
- köklük** see *kökläk*
- küçlük** ‘powerful, strong’ ◇ *küpküçlük* (Ščerbak 1977: 120, Öztürk 1994: 58)
- kündüz** ‘by day’ ◇ *küpkündüz* (UjgRS s.v. كۈنچۈدە içidä: only attested in ~ içidä ‘in broad daylight’)
- oçuq** ‘clear, distinct’ ◇ *apačyq* (Sadvakasov 1976: 156) ~ *opočuq* (UjgRS)
- qara** ‘black, dark’ ◇ *kapkara* (Malov 1954) ~ *qapqara* (Malov 1954, Jarring 1964: ⟨qap ~ qâp + qara ~ qara ~ qaa, UjgRS, Tömür 2003: 111)
- qaranıy** ‘dark, gloomy’ ◇ *qapqaranıy* (Kaşgarlı 1992: 126, Öztürk 1994: 58)
- qizil** ‘red, ruddy’ ◇ *kypkyzyl* (Malov 1954) ~ *qipqizil* (Jarring 1964: ⟨qip qizil, UjgRS: only attested in ~ *jalyanči* ‘shameless liar’, Kaşgarlı 1992: 126, Öztürk 1994: 58, Tömür 2003: 110, De Jong 2007: 60, Tursun 2007: xviii)
- saq** ‘healthy, happy’ ◇ *sapsay* (UjgRS) ~ **sapsaq* ‘sound, healthy’ (Kaşgarlı 1992: 126)³¹
- seriq** ‘yellow’ ◇ *sapsერიq* (Jarring 1964: ⟨sâp sâriy: ‘golden yellow’, Friederich/Yakup 2002: 31) ~ *sepseriq* (Jarring 1964: ⟨sâp sâriy: ‘golden yellow’, UjgRS)
See ⟨sâpsâp⟩ in 2.19.3, and 2.19.4 below.
- šük** ‘silence’ ◇ *šüpšük* (UjgRS)
- süzük** ‘clear, transparent, bright’ ◇ *süpsüzük* (UjgRS, Kaşgarlı 1992: 126, Tömür 2003: 110, De Jong 2007: 60, Tursun 2007: xviii)
- tän** ‘equal, identical’ ◇ *täptän* (UjgRS)
- taqir** ‘smooth, bare’ ◇ *taptaqir* ‘intens.’ ~ *tapteqir* ‘1. intens.; 2. complete absence of something’ (UjgRS)
See 2.19.4 below.

³⁰ Besides *japješil* (⟨yapyéšil⟩), Kaşgarlı 1992: 126 also gives *japjäšil* (⟨yapyéšil⟩), which is probably a misprint. Similar mistakes might be suspected in the case of *aq* and *saq*.

³¹ The original spelling in Kaşgarlı 1992: 126 is ⟨sap, sak⟩, which is probably a misprint for **sapsaq* (⟨sapsaq⟩). Similar mistakes might be suspected in the case of *aq* and *ješil*.

taza ‘clean’ ◇ *taptaza* (UjgRS)

taziliq ‘cleanness’ ◇ *taptaziliq* (UjgRS)

teč ‘quiet, peaceful’ ◇ *tipteč* ‘complete peace’ (UjgRS)

See 2.19.4 below, and also 3.1.11 on possible cognates.

tekis|z ‘smooth, even’ ◇ *teptekiz* ~ *tüptekis* (UjgRS)

tin i.a. ‘silence, peace’ ◇ *tiptin* (UjgRS: only attested in *ätrap* ~ ‘complete silence around’.)

See 3.1.11 on possible cognates.

tinč ‘quiet, still, peaceful’ ◇ *tiptinč* (Kaşgarlı 1992: 126, Tömür 2003: 110)

See 3.1.11 on possible cognates.

tiniq ‘clear, transparent’ ◇ *tiptiniq* (UjgRS)

See 3.1.11 on possible cognates.

toyra i.a. ‘true, accurate, just’ ◇ *toptoyra* (UjgRS)

tüz ‘smooth, even, straight, direct’ ◇ *tüptüz* (UjgRS)

uzun ‘long’ ◇ *upuzun* (Ščerbak 1977: 120)

2.19.3 Special cases

⟨āpáq⟩ ‘wife of an *aq saqal* or a *dāroya*’ (Menges [1936] after Le Coq (exact source not specified))

With respect to meaning, this form corresponds quite well with *apaq* i.a. ‘address to a respected woman’ (Malov 1954), *appaq* i.a. ‘Frau der kleinen *Bäge* und der Groß-*Mullā*’ (Menges [1936]), and also with *appiyim* ~ *appyyym* (with a Px1Sg) ‘my dear, darling; my little white one’ (UjgRS and Malov 1954, respectively).

Menges [1936] does not specify the exact source in Le Coq’s works. If, however, the notation ⟨āpáq⟩ is to be read as /ā’paq/, then the form of this word is rather extraordinary, as the first vowel seems to always be short in reduplications, and the accent initial (see 3.1.20 on the shortening).

It is maybe for this reason, that Menges [1936] suggests that the word might be a borrowing from Mo. *abahaj*. Lessing et al. 1960 attests *ᠠᠪᠠᠬᠠᠵᠢ abayai* with the meaning i.a. ‘wife of a prince; lady’. The phonetic side of this proposition might require a more detailed investigation. Perhaps a contamination with the native *apaq* ~ *appaq* ‘1. snow-white; 2. address to a respected woman, ...’, facilitated by its use with a Px1Sg in the meaning ‘my dear, my darling’ (see 3.1.2 on alternative closers intensifying different components of the base meaning)?

In the present work, Menges’s suggestion will be accepted and ⟨āpáq⟩ excluded from further considerations.

čöp čā ‘motley’ (Jarring 1964: ⟨čöp ča:⟩)

If this word were a reduplication, an unusual change **a > ö* would need to be assumed in its reduplicated anlaut. More likely, it is a composition of the same kind as e.g. Tksh. *güčlü kuvvetli* lit. ‘strong powerful’ = ‘very strong’.

As for the would-be base *čā*, Jarring 1964 attests it (⟨ča:⟩) with the meaning ‘spotted’. For the shape *čöp* he lists three meanings: ‘pasture’, ‘pieces of dough [...]’ (see *čöpčürä* below), and ‘reinforcing word (in *čöp čā*)’. The first one is not clear to me; the other two may well be brought down to a single source.

Stachowski M. [in print] derives Tksh. *žibre* ‘Ausgepreßtes, Trester, Treber’ together with related forms in other languages ≪ general Tk. * $\sqrt{\check{c}öp}$ ‘Überbleibsel, Ausgepreßtes’ (> Tksh. *čöp* ‘garbage’). The semantic relation between ‘pomace’ and ‘spots’ might not be immediately obvious. A surprisingly accurate parallel can be found in Polish which, incidentally, is even phonetically similar: *ćap.áč* ‘to hit water, to bath splashing about’ → *ćap.ka* which denotes the result of this splashing, i.e. ‘spot, dot, splash’, and at the same time ‘soft and dense mass, mash, pulp’ (Boryś 2005).

The whole would be then a composition of **čöp* ‘1. spot, dot, splash; 2. soft and dense mass, mash, pulp’ + *čā* ‘spotted’ → *čöp čā* lit. ‘spot-spotted’ = ‘all in dots, motley’. The first meaning of *čöp* was probably not preserved outside of this phrase, and this fact resulted in the meaning of ‘reinforcing word’ that Jarring 1964 assigned to it, and made *čöp* resemble severed reduplicated anlauts (see 3.1.10).

čöpčürä ‘a kind of noodles’? (Jarring 1964 s.v. *čöčüre* &c.)

This form is recorded in Menges [1936] with the meaning ‘eine Mehlspeise ähnlich dem *пельмень* der Russen’ but, as Jarring reports, “Katanov hat fast überall zuerst *čöpčürä* geschrieben dann das *p* weggestrichen”.

Jarring himself lists ⟨*čöčüre* ~ *čöčüre* ~ *čöčüri*⟩ ‘a dish, described in [...]’ and, after UjgRS39, ⟨*cöcyrə*⟩ ‘small noodles, boiled in bouillon’ (possibly related to Tksh. *žibre* ‘Ausgepreßtes, Trester, Treber’ &c., see Stachowski M. [in print], and *čöp čā* ‘motley’ above.), and below, ⟨*čöp*⟩ ‘2. pieces of dough boiled in water, a kind of noodles; [...]’.

Reduplication can be used with nouns and, in fact, denote plural, but the only examples known to me come from Azeri, where reduplication is incomparably more common and more diversified than in Uighur (see 2.2.5, but also *kim kiček* ‘garment, clothes’ below). Also, the vowel in the reduplicated anlaut can fail to match the vowel of the stem, as a number of Uighur examples attests but, phonetic details aside, the difference lies almost always in the Uighur umlaut (see 2.19.4).

Neither of the words mentioned above is really clear, but despite that, it seems relatively certain that whatever they are, it is not a C-type reduplication.

See also *čöp čā* above.

kim kiček &c. ‘garment, clothes’ (Jarring 1964: ⟨*kim kiček* ~ *kim keček* ~ *keimkiček*⟩) Jarring 1964 considers these forms to be a reduplication, but it appears to me that a composition is a more likely interpretation.

Only the initial part of this form seems to be attested independently: ⟨*ki:m* ~ *kim* ~ *kim* ~ *keim*⟩ ‘garment, clothes, dress’. Most probably, the word is related

to Tksh. *kežim* ~ *kičim* ‘(horse) armour’ < Mo. *kežim* ‘saddlecloth, caparison’ << Tkc. **käδ.im* (>> Tksh. *gijim* ‘clothing’), see Stachowski M. [in print]. The final part, *kiček* &c., although less clear, can probably be eventually reduced to the same root.

Our word would be then a composition of two synonyms, much like Tksh. *güčlü kuvvetli* lit. ‘strong powerful’ and many similar.

See also 3.1.3 for other examples of apparent reduplications.

⟨*sæpsær*⟩ (Jarring 1964) ◊ ⟨*sæjïγ*⟩ ~ ⟨*særiγ*⟩ ~ ⟨*særiq*⟩ ~ ⟨*sijïγ*⟩ ~ ⟨*siriγ*⟩ ‘yellow’ (Jarring 1964)

The base **sær* does not seem to be attested with the appropriate semantics. The word is probably a truncated form of ⟨*sæp særiγ*⟩, but how exactly it came about is not clear to me.

žimžit (Jarring 1964: ⟨*dzimdzit*⟩ ‘sudden silence’, UjgRs: ‘completely silent, completely mute’)

It is not absolutely clear whether this form is actually a reduplication. The base **žit* does not seem to be attested, whereas the would-be reduplicated anlaut *žim* is. Its meaning is ‘quietly, calmly; silently, tacitly’, as in *žim (žim)!* ‘be quiet, shut up!’, *žim turmaq* ‘to keep quiet’ or *žim bolmaq* ‘to quiet down, to grow silent’. Accordingly, UjgRS classifies *žit* as “парное к  [ЖИМ]”.

However, Turkic echo words and echoic compounds are typically built on rhyme and consonance, as in Uigh. *nan-pan* ‘bread, all sorts of pastry’, *opul-topul* ‘quickly, hastily’ or *taraq-turuq* ‘noise, crash, crack’, while *žimžit* is clearly a case of alliteration.

Further, *m* would have been a highly unusual closing consonant for Uighur, where all the 35 examples are closed with *p* or its derivatives (see 2.19.4 below).

More probably, *žim* is the reduplicated anlaut of *žimžit* which was severed and promoted to an independent word, while *žit* belongs to a small family of words across several languages. See 3.1.10 for emancipated anlauts, and 3.1.15 for *čyrt* ‘silence’.

As an alternative – or perhaps complementary? – explanation, a connection with Bshk. *šym* ‘quiet’ can be offered (see 2.3.3), and maybe also with two somewhat mysterious intensifiers with unclear semantics: Trkm. *čym* and, attested in the oldest documents, *čim* [?] *čym* (see 3.1.10).

See also *žimžitliq* below.

žimžitliq ‘complete silence’ (UjgRS)

This word is obviously related to *žimžit* above. Given that neither **žit* nor **žitliq* seems to be attested, it has to be interpreted not as a reduplication of a derivative, but rather a derivative from a reduplication – if it really is a reduplication. See *žimžit* above.

žuyžemi ‘stock and block’ (Jarring 1964: ⟨dzuγ dzæmi⟩)

Jarring 1964 supposes that the word might be a reduplication, but composition appears to be a more likely explanation.

The second part of this form is attested independently with the meaning ‘all, whole’ (Jarring 1964: ⟨dzæmi ~ dzemi ~ dzemε ~ dzemi⟩). The first part does not seem to be so, but see ⟨dzuγla- ~ dzuγla-⟩ ‘to gather, to collect’.

As a reduplication, *žuyžemi* would have been the only one across all of the Turkic languages closed by *γ*. Should this have evolved from **r*, it would have been the only reduplication in Uighur closed by that consonant. There is also no parallel for the *u – e* correspondence in the reduplicated anlaut and the base.

See also 3.1.3 for other examples of apparent reduplications.

2.19.4 Structure

Technically, two closing consonants of *C*-type are attested in a total of 34 examples derived from 33 unique bases, in a rather even distribution:

p: 33 examples: *aq*, *aškara*, *baravär*, *boš*, *jeji*, *jaš*, *ješil*, *joruq*, *jumulaq*, *kök*, *kökläk*, *küčlük*, *kündüz*, *očuq*, *qara*, *qaraγyu*, *qizil*, *saq*, *seriq*, *šük*, *süzük*, *täj*, *taqir*, *taza*, *taziliq*, *teč*, *tekis* | *z*, *tin*, *tinč*, *tiniq*, *toyra*, *tüz*, and *uzun*,

pp: 1 example: *aq*, and

Only one word has more than one closing consonant possible: *aq* with *p* and *pp*.

Dialectally, *p* can be spirantized in the reduplication of the literary *ješil* ‘green’ → *javjäšel* (Malov 1954). See 3.1.21 on spirantization of the closing consonant.

In six words, the reduplicated vowel does not match its original in the root.

The cases of 1. *ješil* ‘green’ → *japješil*, *javjäšel*, 2. *seriq* ‘yellow’ → *sapseriq* and 3. *taqir* ‘smooth, bare’ → *taptaqir*, *tapteqir* are explained by the information in Frießerich/Yakup 2002: 31 that it is the pre-umlaut vowel that becomes reduplicated (see 2.19.1). The shapes *jäpješil* and *sepseriq* apparently reflect the two successive stages of secondary unification.

Next, *tipteč* ‘complete peace’ is possibly a result of contamination: *teč* ~ *tič* → **tepteč* ∅ **tiptič* > *tipteč*. The shape **teptič* does not appear to be attested. It might also be the same case with *jopjumulaq* ‘completely round’: *domalaq* ~ *jumulaq* ~ *žumulaq* → **dopdomalaq* ∅ **jupjumulaq* ∅ **župžumalaq* > *jopjumulaq*.

And finally, *tüptekis* ‘absolutely: smooth, even’ (beside the regular *teptekiz* id.) is not at all clear, as neither **tükis* seems to be attested, nor any other similar shape with a *ü* in the first syllable.

See also 3.1.19 for other reduplications with anlaut not matching the base.

2.19.5 Semantics

In the majority of cases, the reduplicated meaning is a simple intensification of the base meaning.

A more distant evolution is only to be observed in *apaq* and *appaq* ← *aq* ‘1. white; 2. bright; 3. grey; 4. white of an egg; 5. corneal ulcer; ...’ (UjgRS), which beside the straightforward ‘snow-white’ can also be used with a Px1Sg in the meaning ‘my dear, my darling’, or without it to address a respected woman. The latter might have arisen under the influence of ⟨āpāq⟩, probably a borrowing from Mo. *abayaj* i.a. ‘the wife of a prince; lady’; see 2.19.3.

A not entirely trivial evolution is also attested in *taqir*, where ‘smooth, bare’ → ‘1. intens.; 2. complete absence of something’. There has been a slight semantic shift in *kökläk*, too, where ‘green field’ → ‘rich verdure’.

As far as parts of speech are concerned, are Uighur reduplications fairly standard. The great majority of examples are adjectives or adverbs. Three words, *kökläk* ‘green field’, *šük* ‘silence’ and *tin* ‘silence, peace’, have a substantival character, and in one case, *teč* ‘quiet, peaceful’ → *tipteč* ‘complete peace’, an adjectival base reduplicates to a noun.

2.20 Uzbek

Uzbek C-type reduplications are relatively few, moderately diversified, and seem to quite accurately mirror the composite character of the language.

Four closing consonants are attested, but *p* or *pp* are the only possibility for 80.43% of the bases. Both numbers fit precisely between the typical values found in the Oghuz languages on one hand, and the Karakhanid and Kipchak on the other, and as such, are characteristic of none but Uzbek. See 2.20.4.

Non-trivial semantic evolutions or reduplications of non-adjectives are less than rare, but there are two rather unusual exceptions, the reduplicated verbs *qipqizarmâq* ‘to turn very red; to blush strongly’, and *qâpqârajmâq* ‘to turn completely black’. See 2.20.5, and also 3.1.22 on reduplication of verbs in general.

2.20.1 Sources

The main source of the material is UzbRS59 which contains ca. 40 000 entries. Some attestations have also been found in Doerfer 1967: 52, von Gabain 1945: 49, Harrison 2004: 201, Johanson 1998: 39, Räsänen 1957: 74, Ščerbak 1977: 120, Zeynalov 1993: 149f, and in the grammars listed below.

For grammatical descriptions, Wurm 1945: 46f, Borovkov 1959: 695, Kononov 1960: 161f, Sjoberg 1962: 65, Kissen 1975: 20 and Bodrogligeti 2003: 351f have been used.

Wurm 1945: 46f deals briefly with reduplication. He mentions that the possible closing consonants are *m*, *p*, *r* and *s*. Unfortunately, the only example provided is *kök* ‘1. blue; 2. green’ → ⟨kømkök⟩. I could not find any examples for *r* in other sources.

Borovkov 1959: 695, interestingly, only lists *m* and *p* as the possible closing consonants, although *büsbütün* ‘absolutely all’ and *tostopâlan* ‘disorder, ...’ are attested in UzbRS59, of which Borovkov 1959 is a part. It is not clear whether these examples have been discarded for some reason, or simply overlooked. Double *pp* is not listed either, even though *âppâq* ‘snow-white’ is among the examples that follow. Perhaps Borovkov considered it a simple phonetic variant of *p*. Directly after the examples, nevertheless, the *CV*-type with double *pp* is discussed (e.g. *sâppasây* ‘completely healthy’, *çippaçin* ‘absolutely: true, genuine’), and the lengthening of the closing consonant is explicitly mentioned. Borovkov 1959: 695 also points out that reduplications can [only] be formed from adjectives *обозначающих цвета и некоторые свойства*.

Kononov 1960: 161f gives an unnecessarily complicated description of the process of reduplication and lists *m*, *p*, and *s* as the possible closing consonants. Some examples follow, including *âppâq* ‘snow-white’ which is explained as “(⟨оп-ок⟩)”. The lack of asterisk might suggest that *âpâq* does exist in parallel but since the form is apparently not attested in any other source, I believe that it is merely an imperfection of the notation. Also, it is noted that stress in reduplications is initial.

Sjoberg 1963: 65 is very brief. He describes the structure, lists most of the possible closers (*p* ~ *b*, *m*, and *s*), mentions the initial stress, and gives several examples.

Kissen 1975: 20 effectively limits his description to three examples and the statement that reduplications can only be formed from certain adjectives, mainly names of colours.

Bodrogligeti 2003: 351 lists the closing consonants (*p*, *m*, and *s*), mentions the initial stress, and that it is primarily adjectives denoting colours and physical characteristics that are reduplicated. The description is completed with a list of examples, also as used in a sentence.

Transcription:

Bodrogligeti 2003: 351f, Kissen 1975: 20, Kononov 1960, UzbRS59: 161fə- → e- | ε → γ | κ → q | o → â | y → u, ü³² | ŷ → o, ö³³ | x → χ | x → h | ж → ž,
 von Gabain 1945: 49: a → a, â³⁴ | ə → a | oɣ → γ | ø → ö | ş → š | y → ü | v → i |
 small caps ignored,
 Harrison 2004: *uu* → *i*,

³² Based on the etymology: *u* in *bu*, *jâruγ*, *jumšâq*, *juvmâq*, *quruq*, *quš*, *šu*, *suv*, *toyru*, *tuhmat*, *uçmâq*, *ustida*, *u* and *Uzbekistân*; *ü* in *bütün*, *jüzlar*, *süpürülmâq*, *tün* and *üj*.

³³ Based on the etymology: *o* in *bojin*, *boš*, *jol*, *orin*, *ot*, *toyri*|*u*, *tola*, *topâlân*, *tosatdan* and *tosindan*; *ö* in *çötir*, *köjlak*, *kök*, *körmâq* and *öz*.

³⁴ Based on UzbRS59: *a* → in *qâra* → and *tola*; *â* → in *âq*, *jâlyiz* → and *qâra*.

Johanson 1998: 39: $e \rightarrow a^{35}$

Ščerbak 1977: $\kappa \rightarrow q \mid \jmath \rightarrow \text{á} \mid \acute{y} \rightarrow o$,

Sjöberg 1963: $\jmath \rightarrow \text{á} \mid w \rightarrow v$,

Wurm 1945: $o, \acute{o} \rightarrow \acute{o}$,

Zeynalov 1993: $\zeta \rightarrow \check{c} \mid \iota \rightarrow i \mid k \rightarrow q \mid y \rightarrow j$.

2.20.2 Standard cases

áčiq ‘1. open; 2. clear, fair *of weather*; 3. nice, clear, legible; 4. open, frank’ \diamond **ápáčiq** (UzbRS59, Zeynalov 1993: 150)

âq ‘white’ \diamond **ápâq** (Ščerbak 1977: 120) \diamond **âppâq** (von Gabain 1945: 49, Borovkov 1959: 695, UzbRS59, Kononov 1960: 162, Ščerbak 1977: 120, Bodrogligeti 2003: 352)

See **âppâyim** in 2.20.3 below.

âsân ‘light, easy’ \diamond **âpâsân** ‘simple, plain’ (UzbRS59)

baravar ‘same, equal’ \diamond **bapbaravar** (UzbRS41: «баб-баравар», Sjöberg 1963: 65: «bábbarawar»)

boš ‘empty’ \diamond **bomboš** (UzbRS59, Ščerbak 1977: 120, Bodrogligeti 2003: 352) \diamond **boš** (von Gabain 1945: 49, UzbRS59: «бӱб-бӱш».)

See 3.1.4.

bütün ‘all, whole, complete’ \diamond **büsbütün** (von Gabain 1945: 49, UzbRS59, Kononov 1960: 162, Bodrogligeti 2003: 352) \diamond **bütbütün** (Kononov 1960: 162, Bodrogligeti 2003: 352)

See 3.1.6.

čirâjli ‘beautiful, pretty’ \diamond **čipčirâjli** (UzbRS59)

čötir ‘1. pockmarked; 2. rough, uneven, tuberous’ \diamond **čöpčötir** (UzbRS59)

dumalâq ‘round, circular’ (DUzbFr) \diamond **dumdumalâq** (Kononov 1960: 162, Bodrogligeti 2003: 352, Pomorska 2004: 144)

See 3.1.6.

jajdâq ‘bare, devoid of vegetation’ \diamond **japjajdâq** ‘1. bareback; 2. unbridled’ (UzbRS59)

jalanyâč ‘naked, bare, bald, stripped’ \diamond **japjalanyâč** (UzbRS59)

jâlyiz ‘(a)lone, lone(ly)’ \diamond **jâpjâlyiz** (von Gabain 1945: 49, UzbRS59)

jangi ‘1. new, fresh, young; 2. recently’ \diamond **japjangi** (Borovkov 1959: 695, UzbRS59, Bodrogligeti 2003: 352)

japalâq ‘flat’ \diamond **japjapalâq** (Borovkov 1959: 695)

See 3.1.6.

jâruy ‘light, bright’ \diamond **jâpjâruy** (UzbRS59)

jašil ‘green’ \diamond **jamjašil** (von Gabain 1945: 49, UzbRS59, Kononov 1960: 162, Zeynalov 1993: 150, Johanson 1998: 39, Bodrogligeti 2003: 352) \diamond **japjašil** (von Gabain 1945: 49)

³⁵ In the only example, *jamjašil*.

- jengil** ‘light’ ◇ *jepjengil* (UzbRS59)
- katta** ‘big, huge’ ◇ *kapkatta* (UzbRS59, Bodrogligeti 2003: 352f)
- kök** ‘1. blue; 2. green’ ◇ *kömkök* ‘1. very green (von Gabain 1945: 49, UzbRS59, Kononov 1960: 162, Kissen 1975: 20); 2. very green or very blue (Wurm 1945: 47)³⁶; 3. very blue (Borovkov 1959: 695, UzbRS59, Kononov 1960: 162, Bodrogligeti 2003: 352)’
- lâjij** ‘1. worthy, deserving; 2. of appropriate size’ ◇ *lâplâjij* (UzbRS59)
- qâra** ‘black’ ◇ *qâpqâra* (von Gabain 1945: 49, Räsänen 1957: 74: ⟨kap-kara⟩, Borovkov 1959: 695, UzbRS59, Kononov 1960: 162, Sjöberg 1963: 65, Doerfer 1967: 52, Zeynalov 1993: 150, Bodrogligeti 2003: 352f)
See also 3.1.10 on other intensifications of the word.
- qârajmâq** ‘to blacken, to turn black’ ◇ *qâpqârajmâq* (UzbRS59)
- qârânyî** ‘dark, gloomy’ ◇ *qâpqârânyî* (UzbRS59)
- qizarmâq** ‘to redden, to turn red, to flush’ ◇ *qipqizarmâq* (von Gabain 1945: 49, UzbRS59)
- qizil** ‘red, ruddy’ ◇ *qipqizil* ‘1. *intens.* (von Gabain 1945: 49, Borovkov 1959: 695, UzbRS59, Kononov 1960: 162, Sjöberg 1963: 65, Kissen 1975: 20, Ščerbak 1977: 120, Zeynalov 1993: 150, Bodrogligeti 2003: 352); 2. present, explicit, uniform (UzbRS59)’
See 2.20.5 below.
- quruq** ‘1. dry; 2. empty; 3. vain’ ◇ *qupquruq* (UzbRS59)
- sariq** ‘yellow’ ◇ *sapsariq* (Borovkov 1959: 695, UzbRS59, Kononov 1960: 162, Bodrogligeti 2003: 352)
- sijâ(h)** i.a. ‘ink’ ◇ *simsijâ(h)* (UzbRS59) ‘completely dark’
See 2.20.5 below.
- šijdam** ‘naked, bare, empty, clean (*e.g. rob*)’ (Magrufov et al. 1981) ◇ *šipšijdam* (UzbRS59)
- silliq** ‘smooth’ ◇ *sipsilliq* (Borovkov 1959: 695, UzbRS59, Bodrogligeti 2003: 352)
- tajin** ‘(well-)known, certain’ ◇ *taptajin* (UzbRS59)
- tajjâr** ‘ready, prepared’ ◇ *taptajjâr* ‘1. *intens.* (UzbRS59, Bodrogligeti 2003: 352); 2. very definitely (UzbRS59)’
- taqir** ‘smooth, bare, bald’ ◇ *taptaqir* ‘1. *intens.*; 2. devoid’ (UzbRS59, Harrison 2004: 201)
- tâza** ‘clean, neat’ ◇ *tâptâza* (UzbRS59, Bodrogligeti 2003: 352f)
- tekis** ‘smooth, even’ ◇ *teptekis* (UzbRS59, Bodrogligeti 2003: 352) ‘1. *intens.*; 2. united, harmonious’
- tikka** ‘straight, direct’ ◇ *tiptikka* (von Gabain 1945: 49)
- tin** ‘still, steady’ ◇ *timtin* (von Gabain 1945: 49, UzbRS41)
- tola** ‘full, whole, complete, all’ ◇ *toptola* (von Gabain 1945: 49, UzbRS59)
See also 3.1.10 on other intensifications of the word.

³⁶ Wurm 1945: 47: “*komkök* = ‘sehr grün’ (eigentlich: sehr blau)”.

topâlân ‘turmoil, commotion, uproar’ ◊ **tostopâlân** ‘disorder, chaos, turmoil’ (UzbRS59)
tosatdan ‘suddenly, unexpectedly’ ◊ **toptosatdan** (UzbRS59)
tosindan ‘suddenly, unexpectedly’ ◊ **toptosindan** (UzbRS59)
toyri ‘1. straight, direct; 2. true, faithful, accurate’ ◊ **toptoyri** (UzbRS59, Kissen 1975: 20)
 ~ **toptoyru** (von Gabain 1945: 49)

2.20.3 Special cases

âppâyim ‘my dear, my darling’ (UzbRS59) ◊ **âq** ‘white’

Similarly to the state in Bashkir, Kirghiz, Tatar and possibly Uighur, Uzb. *âq* has two different reduplications, each of which intensifies a different component of its base meaning. See 3.1.2 on other such cases.

pakpakana (Bodrogligeti 2003: 352) ◊ **pakana** ‘short of a person’

This word is mentioned by Bodrogligeti 2003: 352 together with *bütbütün*, *dum≠dumalâq*, and *japjapalâq* as “adjectives [...] [that] use in the reduplication the consonant that follows the first syllable”.

However, *pakpakana* is a different case than the other three, because here the closing consonant would have to be *k*, and this does not occur except for one example in Yakut (*maŋan* ‘white’, see 2.21.4). Perhaps, the word really is a germ of a separate type of reduplication, as Bodrogligeti apparently wants it?

See 3.1.6 for other examples with the closing consonant identical to C_2 .

žimžit ‘completely silent, completely mute’ (UzbRS59)

The base *žit does not seem to be attested in itself, but apparently cognate forms can be found in different languages, see 3.1.15.

2.20.4 Structure

Four closing consonants of *C*-type are attested in a total of 46 examples derived from 42 unique bases, in a relatively even distribution:

m: 6 examples: *boš*, *dumalâq*, *jašil*, *kök*, *sijâ(h)* and *tin*,

p: 36 examples: *âčiq*, *âq*, *âsân*, *baravar*, *boš*, *čirâjli*, *čötir*, *jajdâq*, *jalanyâč*, *jâlyiz*, *jangi*, *japalâq*, *jâruy*, *jašil*, *jengil*, *katta*, *lâjiq*, *qâra*, *qârajmâq*, *qârânyî*, *qizarmâq*, *qizil*, *quruq*, *sariq*, *šijdam*, *silliq*, *tajin*, *tajjâr*, *taqir*, *tâza*, *tekis*, *tikka*, *toyri*, *tola*, *tosatdan*, and *tosindan*,

pp: 1 example: *âq*,

s: 2 examples: *bütün*, *topâlân*, and

t: 1 example: *bütün*.

Four words have more than one closing consonant possible: *âq* (*p* and *pp*; see *âppâyim* in 2.20.3 above), *boš* (*m* and *p*), *bütün* (*s* and *t*), and *jašil* (*m* and *p*).

The forms *bütbütün* ‘absolutely all’, *dumdumalâq* ‘completely round’, *jarjaralâq* ‘completely flat’, and *pakpakana* ‘very short of a person’ are quite specific because they all use their C_2 for the closer; see 3.1.6 for more examples and a discussion.

Derivatives and non-adjectives are not unusually numerous, with the notable exception of two verbs: *qipqizarmâq* and *qâpqrâjmâq*; see 2.20.5 below.

2.20.5 Semantics

The majority of reduplicated meanings are simple intensifications or apparently the same as their respective base meanings.

The reduplicated meaning departs a little further away in five cases: *âsân* where ‘light, easy’ → ‘simple, plain’, *qizil* where ‘red’ → i.a. ‘present, explicit, uniform’, *sijâ(h)* where i.a. ‘ink’ → ‘completely dark’, *tajjâr* where ‘ready, prepared’ → i.a. ‘very definitely’, and *tekis* where ‘smooth, even’ → i.a. ‘united, harmonious’.

The first of two nontrivial cases among them is *qizil*. It is not clear to me exactly which component of its meaning could have given rise to the evolution. Magrufov et al. 1981 only defines the word as ‘1. қон рангидаги; қирмизи; 2. юзнинг, баданнинг шу тусга мойил ранги; қизиллик; 3. революцион ҳаракатта, совет социалистик тузумига, Қизил Армияга оид’, and neither of these seems a likely starting point for intensification to ‘present, explicit, uniform’. Perhaps it is a mistake in UzbRS59? (Note that no other source appears to confirm such meaning of the reduplication.)

Also unclear is *sijâ(h)* ‘1. чернила; 2. вид, внешность, наружность, облик, лицо; 3. ...’, as the word eventually stems from Pers. سیاه *sijâh*, primarily ‘black; dark’. Traces of these meanings can still be found when it is used with *dil* ‘1. heart; 2. soul’, e.g. in *sijâ dil* ‘regretful, resentful, sad’ and similar, see UzbRS59 s.v. *cuë*. However, since the reduplicated meaning ‘completely dark’ can be easily derived from ‘ink’, the history behind the meaning ‘form, appearance’ needs not concern us here.

The majority of examples are quite clearly adjectives, to the degree allowed by the structure of Turkic in general. However, one example is a noun, and two, rather surprisingly, are verbs.

First, the noun. UzbRS59 translates *töpâlân* as ‘хаотический беспорядок, хаос, суматоха, сумятица’, and its reduplication appears to be no more adjectival in character: ‘1. суматоха, сумятица, шум, скандал; 2. буйство; 3. озорство, шалость’. See more on reduplications of nouns in 3.1.14.

But the most interesting are the two verbs: *qâpqrâjmâq* ‘to blacken, ...’ and *qipqizarmâq* ‘to redden, ...’. Both *qâra* ‘black’ and *qizil* ‘red’ are among the most common bases for reduplication in all the Turkic languages. Reduplications of verbs, however, are more than rare. The pair at hand appears then to be an Uzbek innovation. No evidence seems to exist to suggest whether it is a vestige of a failed attempt to broaden the

scope of reduplication onto verbs in Uzbek, or two isolate cases of analogy to *qipqizil* and *qâpqâra*. See more on reduplications of verbs in 3.1.22.

2.21 Yakut

Reduplications in general, and *C*-type in particular, have prospered in Yakut. They are quite numerous and rather diversified. Interestingly, this is strictly unlike Dolgan (see 2.4).

Although most can be reduced to the standard four, as many as ten closing consonants are attested, which is almost twice as many as in the second Azeri. Interestingly, *m* is not one of them. As long as only *C*-type reduplications are considered, words with multiple closing consonants are relatively few. The reduplicated long vowel or diphthong is not always shortened; in fact, in four or more cases a short vowel of the stem has been actually lengthened in the reduplication. See 2.21.4.

The semantic side is much more temperate. Reduplicated are primarily adjectives and adverbs, and only rarely words of a more definite substantival character. The reduplicated meaning never departs too far from the base. See 2.21.5.

Special cases are surprisingly few, given the general bloom of reduplications in Yakut. Three words cause a certain difficulty with categorization as their reduplicated anlaut is exactly the same as the initial syllable of the base (*bosbosxo* ‘absolutely straight, ...’, *čepčepčeki* ‘very cheap’, and *tastastyŋ* ‘completely outer, ...’), one is not clear because it has its reduplicated anlaut closed by *pp*, normally only reserved for *ak* ‘white’ (*üppürüŋ*, also ‘snow-white’), and one has evolved into an entire family of twelve words (*soččoyotox* ‘very lone(ly)’ &c.). Also rather unusual is the form *ūnutyary* ‘completely opposite, ...’. See 2.21.3.

Although beyond the primary scope of this work, it ought to be noted here that beside *C*-type reduplications, Yakut has a large number of examples in several other types such as *ah.ys.ahy* ‘very bitter’, *bö.rü.böyö* ‘very strong, sturdy’, or *di.bis.diriŋ* ‘very deep’. They will not be discussed in the present work, but see 3.4.3 for similarities with the Mongolic reduplications.

2.21.1 Sources

The main source of the material is Pekarskij 1907–30 which contains ca. 40 000 entries. Some attestations have also been found in Ščerbak 1977: 120, Serebrennikov/Gadžieva 1986: 112, and the grammars listed below.

For grammatical descriptions, Korkina/Ubrjatova et al. 1982: 159 and Kirişcioğlu 1999: 84 have been used.

Korkina/Ubrjatova et al. 1982: 159 provide a slightly random description of the process, and list a few examples of different types (*C*, *CV*, *CVC*). They clearly treat

the forms closed by *p* separately from those closed by *č*, *n* and *s* but without explaining why. Sadly, they also fail to even remark upon the peculiarities in the reduplicated anlaut (e.g. the long vowels in *būsbutün* ‘absolutely all’ vs. *upūllayas* ‘completely thawed’).

Kirişçioğlu 1999: 84 gives a one sentence description, which is imprecise to the point of being useless ([...] pekiştirme değişik şekillerde yapılabilir: [...] sıfatın ilk sesinin veya ilk hecesinin ekli tekrarlanmasıyla), and adds a few examples but does not remark upon their phonetic or other peculiarities (e.g. in *küpküöç*, *suossoyotoç* or *ybysyrās*).

Transcription:

Kirişçioğlu 1999: 84: $\check{g} \rightarrow \gamma$ | $h \rightarrow \acute{h}$ | $\iota \rightarrow y$ | $\tilde{n} \rightarrow \eta$ | $x \rightarrow \chi$ | $VV \rightarrow \bar{V}$,

Korkina/Ubrjatova et al. 1982: 159: $h \rightarrow \acute{h}$ | $VV \rightarrow \bar{V}$ | $x \rightarrow \chi$,

Pekarskij 1907–30: $\ddot{a} \rightarrow e$ | $\bar{h} \rightarrow \gamma$ | $i \rightarrow i, Vj$ | $j \rightarrow j$ | $l \rightarrow l$ | $h \rightarrow \eta$ | $hj \rightarrow \acute{h}$ | $hh \rightarrow \eta\eta$ |

$\ddot{o} \rightarrow \ddot{o}$ | $VcV \rightarrow \acute{h}$ | $x \rightarrow \chi$ | $\check{y} \rightarrow \ddot{u}$ | $\check{u} \rightarrow \check{z}$,

Ščerbak 1977: $h \rightarrow \acute{h}$ | $\bar{i} \rightarrow \bar{y}$, and

Serebrennikov/Gadžieva 1986: $\kappa \rightarrow k$.

2.21.2 Standard cases

añy ‘1. bitter; 2. sour; 3. oversalted’ \diamond **apafñy** ‘1. very bitter (Pekarskij 1907–30, Ščerbak 1977: 120, Korkina/Ubrjatova et al. 1982: 57, 159, Kirişçioğlu 1999: 84); 2. very sour; 3. very oversalted (Korkina/Ubrjatova et al. 1982: 159)’

arayas ‘1. (light) yellow; 2. reddish’ \diamond **aparayas** (Korkina/Ubrjatova et al. 1982: 159)

bosxo i.a. ‘straight, upright’ \diamond **borbosxo** (Pekarskij 1907–30) \diamond **bosbosxo** (Pekarskij 1907–30)

See *čepčepčeki* in 2.21.3 below.

budān ‘misty, foggy’ \diamond **busbudān** (Pekarskij 1907–30)

bütün ‘all, whole’ \diamond **büsbütün** (Pekarskij 1907–30) \diamond **būsbutün** (Korkina/Ubrjatova et al. 1982: 159)

See 2.21.4 below.

byrtaç ‘unclean, icky, foul, vile’ \diamond **bysbyrtaç** ‘lousy, crappy’ (Pekarskij 1907–30)

čaçarçaj i.a. ‘with auburn, gingery hair of people’ \diamond **čapčaçarçaj** ‘fulvous, chestnut, blond’ (Pekarskij 1907–30)

čarās ‘thin’ \diamond **čapčarās** (Pekarskij 1907–30)

čegejikēn ‘with head high on a thin neck and wide eyes’ (Pekarskij 1907–30) \diamond **čepčēgejikēn** (Pekarskij 1907–30: only attested in the riddle ~ *ojun* ‘мутъ [?]–шаманъ (о клеѣ въ клеянкѣ)’)

čepčeki i.a. ‘inexpensive, cheap’ \diamond **čepčepčeki** (Pekarskij 1907–30)

See *čepčepčeki* in 2.21.3 below.

čugas ‘close, near’ \diamond **čupčugas** (Pekarskij 1907–30)

čuoyur ‘motley’ \diamond **čuopčuoyur** (Pekarskij 1907–30)

čyčās ‘small, shallow, light, modest’ \diamond **čypčyčās** (Pekarskij 1907–30)

- čyχ|ka atyn** ‘completely other, completely different’ ◇ **čyрчyχ|ka atyn** (Pekarskij 1907–30)
- č|sylās** ‘warm’ ◇ **čyрčylās** ~ **sypsylās** (Pekarskij 1907–30)
On č ~ s, see *soččoyotoχ* in 2.21.3.
- delej** ‘plentiful, affluent’ ◇ **depdelej** (Pekarskij 1907–30)
- den** ‘unexpectedness, unexpected event’ ◇ **depden** ‘completely by accident’ (Pekarskij 1907–30)
- dirin** ‘deep’ ◇ **dipdirin** (Pekarskij 1907–30)
- doyolon** ‘lame, limping’ ◇ **dopdoyolon** (Pekarskij 1907–30)
- eder** ‘young’ ◇ **epeder** (Pekarskij 1907–30)
- erien** ‘motley’ ◇ **eperien** (Pekarskij 1907–30)
- χara** ‘black’ ◇ **χapχara** (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 57, 159, Kiriščioğlu 1999: 84)
- χobū** ‘slander, calumny’ ◇ **χopχobū** (Pekarskij 1907–30)
- χojū** ‘thick, dense’ ◇ **χopχojū** (Korkina/Ubrjatova et al. 1982: 57, 159)
- χožoγor** ‘slender, tall, svelte’ ◇ **χopχožoγor** (Pekarskij 1907–30)
- itī** ‘hot, torrid’ ◇ **ipitī** (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 159)
- ketit** ‘wide, broad, vast’ ◇ **kepketit** (Pekarskij 1907–30)
- kien** ‘wide, broad, spacious, vast’ ◇ **kipkien** (Pekarskij 1907–30)
- kilekij** ‘glossy’ ◇ **kipkilekij** (Pekarskij 1907–30)
- kiligir** ‘polished, smooth’ ◇ **kipkiligir** (Pekarskij 1907–30)
- könö** ‘smooth, straight, right, proper’ ◇ **köpkönö** (Pekarskij 1907–30)
- könüllük** ‘freely’ ◇ **köpkönüllük** (Pekarskij 1907–30)
- kugas** ‘red, rufous’ ◇ **kupkugas** ‘completely red-haired’ (Pekarskij 1907–30)
- kufiayan** ‘thin, bad, unfit, rubbishy’ ◇ **kupkufiayan** ‘very bad, very thin’ (Pekarskij 1907–30)
- küöχ** ‘1. blue; 2. green’ ◇ **küpküöχ** ‘1. very blue (Kiriščioğlu 1999: 84); 2. very green (Pekarskij 1907–30)’
- kütür** i.a. ‘wicked, rabid, fierce’ ◇ **küpkütür** (Pekarskij 1907–30)
- kyfiyl** ‘red’ ◇ **kypkyfiyl** (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 159, Seř rebrennikov/Gadžieva 1986: 112)
- kyra** ‘small, tiny, fine’ ◇ **kypkyra** (Pekarskij 1907–30)
- kyžy** ‘passion to act contrary to the usual way’ (Pekarskij 1907–30) (= ‘perversity, conkynkyžy (Pekarskij 1907–30: only attested in: ~ *byhylāχ* ‘perverse, contrary of a person; capricious, unruly of a horse’, and ~ *majgylāχ* ‘bad-tempered’) ◇ **kyηkyžy** (Pekarskij 1907–30: only attested in: ~ *syrylāχ* ‘perverse, contrary’, and ~ *majgylāχ* ‘unruly of cattle which does not walk on the road’ ◇ **kypkyžy** (Pekarskij 1907–30: only attested in ~ *syrylāχ* ‘perverse, contrary’)
- maᅇan** ‘white’ ◇ **makmaᅇan** (Pekarskij 1907–30) ◇ **matmaᅇan** (Pekarskij 1907–30)
See 2.21.5 below.

- naryn** ‘fine, trim, thin, subtle’ ◊ *napnaryn* ‘very tiny’ (Pekarskij 1907–30)
- neg|nej** ‘unfit, improper, bad’ ◊ *nepneg|nej* (Pekarskij 1907–30)
- nenej** see *neg|nej*
- ń|ženkir** ‘transparent, clear of rock, water’ ◊ *ńepńenkir* ~ *ženženkir* (Pekarskij 1907–30)
- ńiččeyej** ~ **nilčeyej** ‘wet’ ◊ *ńipńiččeyej* ~ *nipnilčeyej* (Pekarskij 1907–30)
- nilčeyej** see *ńiččeyej*
- ninšik** ‘stale of food and bread, musty’ ◊ *nipninšik* (Pekarskij 1907–30)
- ńulun** ‘savourless, insipid’ ◊ *ńupńulun* (Pekarskij 1907–30)
- oččuguj** ‘small, tiny’ ◊ *opoččuguj* (Pekarskij 1907–30)
- saharχaj** ‘(reddish) yellow, rufous, bay, brown’ ◊ *sapsaharχaj* ‘very yellow’ (Pekarskij 1907–30)
- salan** ‘generous, very’ ◊ *sapsalan* (Pekarskij 1907–30: only attested in ~ *tymn̄y* ‘terrible cold’)
- seber** ‘clean, neat’ ◊ *sepseber* (Pekarskij 1907–30)
- sibetiej** ‘holy, saint’ ◊ *sipsibetiej* (Pekarskij 1907–30)
- sibilgin** ‘now, this instant’ ◊ *sipsibilgin* (Pekarskij 1907–30)
- sikej** ‘crude, raw’ ◊ *sinsikej* (Pekarskij 1907–30) ◊ *sipsikej* (Pekarskij 1907–30)
- simigir** ‘narrow of eyes, purblind’ ◊ *sipsimigir* ‘very narrow-eyed’ (Pekarskij 1907–30)
- sińiges** ‘thin of round objects, narrow’ ◊ *sipsińiges* (Pekarskij 1907–30)
- sīnes** ‘thin’ ◊ *sipsīnes* (Pekarskij 1907–30)
- soyotoχ** ‘lone(ly)’ ◊ *sosoyotoχ* (Pekarskij 1907–30) ◊ *suočsoyotoχ* (Korkina/Ubrjačova et al. 1982: 159) ◊ *suossoyotoχ* (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 159, Kiriščioğlu 1999: 84)
See *soččoyotoχ* in 2.21.2 below.
- sońū** ‘a cry, weep’ ◊ *sojsońū* ‘a great sob’ (Pekarskij 1907–30)
- sotoru** i.a. ‘now, immediately’ ◊ *sopsotoru* ‘very soon’ (Pekarskij 1907–30)
See 2.21.5 below.
- suon** ‘thick, fat, corpulent, stout’ ◊ *supsuon* (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 159)
- sür** ‘fear, horror’ ◊ *süpsür* ‘dreadful’ (Pekarskij 1907–30)
- sygyńaχ** ‘naked, nude’ ◊ *sypsygyńaχ* (Pekarskij 1907–30)
- sylās** see *č|sylās*
- symnaγas** ‘soft’ ◊ *sypsymnaγas* (Pekarskij 1907–30)
- symyja** ‘a lie’ ◊ *sypsymyja* ‘pure lie’ (Pekarskij 1907–30)
- syrdyk** ‘light, bright, enlightened’ ◊ *sypsyrdyk* (Pekarskij 1907–30)
- syt̄y** ‘sharp(-sighted), acute’ ◊ *sypsytȳy* (Pekarskij 1907–30)
- sytygan** ‘rotten, foul, fetid’ ◊ *sypsytygan* ‘fetid, rancid’ (Pekarskij 1907–30)
- talarχaj** ‘excessively long of a house’ ◊ *taptalarχaj* (Pekarskij 1907–30)
- tarajaj** ‘bald, hairless’ ◊ *taptarajaj* (Pekarskij 1907–30)

- tastyn** ‘outer, outsider’ ◇ *tastastyn* (Pekarskij 1907–30)
See *čepčepčeki* in 2.21.3 below.
- tefiheyes** ‘holey, leaky of utensils, bags’ ◇ *tepteheges*³⁷ (Pekarskij 1907–30)
- tenigir** ‘wide, vast, loose’ ◇ *teptenigir* (Pekarskij 1907–30)
- teređer** ‘1. with a broad edge and a narrow base; 2. split, forked’ ◇ *tepteređer* (Pekarskij 1907–30)
- tetekej** ‘pink’ ◇ *teptetekej* (Pekarskij 1907–30)
- tiere** i.a. ‘contrariwise, the other way round, inside out’ ◇ *tiptiere* (Pekarskij 1907–30)
- tögürük** ‘round’ ◇ *töptögürük* (Pekarskij 1907–30, Ščerbak 1977: 120, Korkina/Ubrjačova et al. 1982: 159)
- toku** ‘amiss, awkward’ ◇ *toptoku* ‘very awkward, very odd, very clumsy’ (Pekarskij 1907–30)
- toloru** ‘full’ ◇ *toptoloru* (Pekarskij 1907–30: only attested in ~ *žollōχ* ‘blissful’ (see *žollōχ* below))
- tolōs** ‘amiss, awkward’ ◇ *toptolōs* ‘very awkward, very odd, very clumsy’ (Pekarskij 1907–30)
- tüökün** i.a. ‘swindler, rogue, thief’ ◇ *tüöptüökün* ‘big cheater’ (Pekarskij 1907–30)
- türgennik** ‘soon’ ◇ *tüptürgennik* (Pekarskij 1907–30)
- tüstāχ** ‘salted’ ◇ *tuptüstāχ* ‘oversalted’ (Pekarskij 1907–30)
- tymn̄y** ‘cold, frosty’ ◇ *typtymn̄y* (Pekarskij 1907–30)
- ufun** ‘long’ ◇ *upufun* (Pekarskij 1907–30, Ščerbak 1977: 120f, Korkina/Ubrjatova et al. 1982: 57)
- ūllayas** ‘thawed’ ◇ *upūllayas* (Korkina/Ubrjatova et al. 1982: 159)
- ürdük** ‘high, tall’ ◇ *üpürdük* (Pekarskij 1907–30)
- ürün** ‘white’ ◇ *üpürün* ‘very white (Pekarskij 1907–30), brightness, illumination (Kiriščioğlu 1999: 84)’ ◇ *üppürün* (Pekarskij 1907–30)
See *üppürün* in 2.21.3 below.
- ürünñük** ‘whiteness’ ◇ *üpürünñük* (Pekarskij 1907–30)
- utary** ‘across, opposite’ ◇ *ūnutary* (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 159)
See *ūnutary* in 2.21.3 below.
- yaraxan** ‘heavy’ ◇ *ypyaraxan* ‘very heavy of smell (Pekarskij 1907–30), very heavy (Korkina/Ubrjatova et al. 1982: 159, Kiriščioğlu 1999: 84)’
- yksary** i.a. ‘tightly’ ◇ *ypyksary* (Pekarskij 1907–30)
- yrās** ‘clean, pure, flawless, saint’ ◇ *pyyrās* (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 57, 159)
- ytyk** i.a. ‘respected, honourable’ ◇ *ypytyk* (Pekarskij 1907–30)
- žadany** ‘poor, destitute’ ◇ *žapžadany* (Pekarskij 1907–30)
- ženkir** see *n|ženkir*

³⁷ Possibly a misprint for **tepteheyes*.

žikti ‘unusual, extraordinary, wonderful’ ◊ **žipžikti** ‘wonderful, marvelous’ (Pekarskij 1907–30)

žollōχ ‘happy, blessed, blissful’ ◊ **žopžollōχ** (Pekarskij 1907–30)

žülej ‘deaf’ ◊ **žüpžülej** (Pekarskij 1907–30)

2.21.3 Special cases

bas battax ‘here and there, every which way, aimlessly, pointlessly, in vain’ (Pekarskij 1907–30)

The base *battax* does not seem to be attested independently in an at least remotely matching meaning. Pekarskij 1907–30 explains it as a derivative of a pronominal *ba*, refers to *antax*, *χajtax* and *bettex*, and gives two examples: *bas battax* and *sir battax*, both with the same meaning.

The word is not clear and will be excluded from further considerations.

bosbosχo see *čepčepčeki* below

čepčepčeki (Pekarskij 1907–30) ◊ *čepčeki* i.a. ‘inexpensive, cheap’

There is a minor uncertainty as to the exact shape of this word as Pekarskij 1907–30 s.v. *čän* actually gives the form ⟨чән чәнчәки⟩. On the next page, however, the base is already spelt ⟨чәнчәки⟩, suggesting that the disharmonic variant is a mere misprint.

More interestingly, this word together with *bosbosχo* ‘absolutely straight, ...’ and *tastastyj* ‘completely outer, ...’ appear as if they had their full initial syllables reduplicated. There are more examples like this in other languages, and their classification is not entirely clear, see 3.1.6.

Possibly, one could also see in *čepčepčeki* a double reduplication, but this is not the case. The would-be base **čeki* does not seem to exist, and *čepčeki* as a whole is in fact a Mongolian loanword (see WMo. *seb ki-*, *sebki-*, *segki-* ‘to restore strength, to recover, to relax’, &c. in Kałuzyński 1979 s.v. *čäpčä-*, supported in DW s.v. *čäbäki*).

čuoččoyotoγun ~ **sotčoyotoγun** ~ **suoččoyotoγun** ~ **totčoyotoγun** (Pekarskij 1907–30) ◊ *soyotoγun* ~ **čoyotoγun* ‘alone, individually’

See *soččoyotoχ* below.

χōn χotojon see *mōnmötöj* below

mōn mötöj (Pekarskij 1907–30: only attested in *ān atajan tijen kellim ebe-et*, ~ *tijen kellim ebe-et* ‘and know that I came propitiated and satisfied’ $\stackrel{?}{=}$ ‘with my chest puffed out’) ◊ **mōnmötöj** ‘to buckle, to bulge, to puff out of chest’

Pekarskij 1907–30 suspects *mōn* could be an “alliterative syllable” and compares it to *ān* and *χōn*.

The first is certainly not limited to words beginning with *an-* or *ān-*, see e.g. *küden* ‘(light) fog’ : *ān küden* ‘dense fog’. (In itself, this does not outrule

a reduplicative origin; see 3.1.10 on severed reduplicated anlauts promoted to (relatively) independent intensifiers.)

The second, $\chi\bar{o}n$, is only attested in two phrases (s.v. ³*xomoi*: $\chi\bar{o}n \chiotojbut \chi\bar{o}n$ ‘an impressive household’, and $\bar{a}n atajan \chi\bar{o}n \chiotojon oloror$ ‘he lives contentedly, richly and importantly’). The semantics is not clear here as \chiotoj appears to only be attested with the meanings ‘1. eagle; 2. tortuous; 3. to not support the weight, to bend, to cave in, to sag’ (see JakRS and Pekarskij 1907–30).

One intriguing thing to note is that both $m\bar{o}n$ in its only attestation and $\chi\bar{o}n$ in one of its two, form highly rhythmic wholes. This suggests that they could be in fact ad hoc poetic creations rather than established reduplications.

Semantic difficulties, reinforced by this supposition, render the two words unclear and unsuitable for drawing conclusions from in the further part of the present work.

Nonetheless, a note needs to be made of the fact that the ‘reduplicated’ vowel is long in these two examples, while the first vowel of the would-be base is short. Most likely, this is due to emphatic lengthening, see 3.1.12.

soččoyotoχ, **sotčoyotoχ**, **suoččoyotoχ** (Pekarskij 1907–30) \diamond *soyotoχ* \sim *čoyotoχ ‘lone(ly)’

These words need to be considered together with two groups of reduplications based on derivatives of *soyotoχ*, listed under *čuoččoyotoχun* above and *soččoyotoχto* below. What sets them apart from the regular reduplications *sos-*, *suoč-* and *suos.soyotoχ* in 2.21.2 above, is the anlaut of the base. (The word *regular* is used rather loosely here. In light of Yakut reduplications as a whole, lengthening of the reduplicated vowel and use of č as the closer appear to be relatively standard compared to modification of the initial consonant of the base.) See tab. 2.1.

| Base | Regular | Irregular |
|------------------|------------------------------|--|
| <i>soyotoχ</i> | <i>sos.s-</i> <i>suoč.s-</i> | <i>soč.č-</i> <i>sot.č-</i> <i>suoč.č-</i> |
| <i>soyotoχto</i> | | <i>soč.č-</i> <i>sot.č-</i> <i>suoč.č-</i> |
| <i>soyotoχun</i> | <i>čuoč.č-</i> | <i>sot.č-</i> <i>suoč.č-</i> <i>tot.č-</i> |

Table 2.1: A comparison of regular and irregular reduplications of Yak. *soyotoχ* ‘lone(ly)’ &c.

Three factors can be adduced to explain it:

1. the Siberia-wide alternation $s \sim t$ which in Yakut additionally extends to č (e.g. *yršaj-* \sim *yrtaj-* \sim *yržaj-* ‘to protrude *especially of bones*, to put forward, to show one’s ivories, to smile’ (ĖSTJa s.v. *ыржаһ*, Pekarskij 1907–30), see also *čylās* \sim *sylās* in 2.21.2 above),

2. contemporary Yakut consonant assimilations on morpheme boundary where \neq by $-t + s- \rightarrow -čč-$ (see Jastremskij 1938: 21), and
3. historical Yakut consonant assimilations whereby $-s + s- > -čč-$ if the final $-s$ stems from $*-č$ (see Ubrjatova 1985a: 38).

In the forms with $-tč-$, it would only be option 1 (alternation) that needs to be taken into account as option 2 ($-t + s- > -tč-$) seems impossible, and the conditions are not met for option 3 ($-s < *-č$).

Thus, a $*čoyotoχ$ needs to be assumed. Such a form is also suggested by $čuoččoyotoyūn$, but it does not seem to be attested in its own right, not at least in the literary language. This renders the $-čč-$ variants unclear.

In theory, any kind of mutual influence between all these related forms is conceivable. In fact, the only *C*-type reduplication with $č$ for a closing consonant, are all members of this family. It appears to be quite likely that they result from such mutual contamination rather than are genuine reduplications. For the forms in $-čč-$, perhaps even a double descent might be imagined, from two separate (dialectal?) forms created one by alternation and the other by assimilation, which both yielded the exactly same shape.

See also Dolg. *suotčogotok* in 2.4.3, where a possibility of explaining the word as a nominal composition, is presented, and 3.1.11 where other large families of related reduplications are collected.

soččoyotoχto , **sočoyotoχto**, **suoččoyotoχto** (Pekarskij 1907–30) \diamond *soyotoχto* \sim $*čoyotoχto$ ‘once, one time’

See *soččoyotoχ* above.

sočoyotoχto see *soččoyotoχto* above

sočoyotoyūn see *čuoččoyotoyūn* above

suoččoyotoχto see *soččoyotoχto* above

suoččoyotoyūn see *čuoččoyotoyūn* above

tastastyn see *čepčepčeki* above

totčoyotoyūn see *čuoččoyotoyūn* above

üppürün (Pekarskij 1907–30) \diamond *ürün* ‘white’

C-type reduplications closed by a double *pp* almost never occur with bases other than *ak* ‘white’. The only exceptions known to me are Kar.SW *appačyk* ‘wide open, completely open’ and Yak. *üppürün* ‘pure white’. The semantic connection is clear, especially that the stem *ak* is actually missing altogether from Yakut, and *ačyk* is very often used in the meaning ‘light of colours’. However, the similarity is probably incidental. See also 3.1.8 on the use of double *pp*, and 3.1.24 on reduplications of ‘white’.

ūnutary (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982: 159) \diamond *utary* ‘across, opposite’

This form is unusual in two ways. Firstly, it is the only base beginning with a vowel that has its reduplication closed by a consonant other than *p*; see 3.1.23.

Secondly, its reduplicated vowel is not only long in itself, but it is in fact long *while* its corresponding vowel in the base is short. This is also a very rare situation, but parallel examples exist; see 3.1.20.

ypyččary ‘very full’ (Pekarskij 1907–30)

The base *yččary does not seem to be attested. The word is unclear.

2.21.4 Structure

As many as ten closing consonants of *C*-type are attested in a total of 106 examples derived from 97 unique bases, in a quite uneven distribution:

č: 1 example: *soyotoχ*,

j: 1 example: *sojū*,

k: 1 example: *maŋan*,

n: 3 examples: *kyžy*, *sikej*, and *utary*,

ŋ: 1 example: *kyžy*,

p: 88 examples: *aŋȳ*, *arayas*, *čačarχaj*, *čarās*, *čegejikēn*, *čepčeki*, *čugas*, *čuoyur*, *čyčās*, *čyχ|ka atyn*, *č|sylās*, *delej*, *deŋ*, *diriŋ*, *doyoloŋ*, *eder*, *erien*, *χara*, *χobū*, *χojū*, *χožoyor*, *itī*, *ketit*, *kieŋ*, *kilekij*, *kiligir*, *könö*, *köŋüllük*, *kugas*, *kuhayan*, *küöχ*, *kütür*, *kyhyl*, *kyra*, *kyžy*, *naryn*, *neg|nej*, *niččeŋej*, *niŋsik*, *ńulun*, *ń|žeŋkir*, *oččuguj*, *saharχaj*, *salan*, *seber*, *sibetiej*, *sibilgin*, *sikej*, *simigir*, *sińiges*, *sīŋes*, *sotoru*, *suon*, *sür*, *sygyńaχ*, *symnayas*, *symyja*, *syrdyk*, *syt̄y*, *sytygan*, *talarchaj*, *tarayaj*, *teheŋes*, *tenigir*, *tereger*, *tetekej*, *tiere*, *tögürük*, *toku*, *toloru*, *tolōs*, *tüökün*, *türgennik*, *tüstāχ*, *tymn̄y*, *uhun*, *ullayas*, *ürdük*, *ürün*, *ürünŋük*, *yaraχan*, *yksary*, *yrās*, *ytyk*, *žadany*, *žikti*, *žollōχ*, and *žülej*,

pp: 1 example: *ürünŋ*,

r: 1 example: *bosχo*,

s: 8 examples: *bosχo*, *budān*, *bütün* (twice: *büs-* and *būs-*), *byrtaχ*, *soyotoχ* (twice: *sos-* and *suos-*), *tastyŋ*, and

t: 1 example: *maŋan*.

In fact, only six of these consonants can be regarded as independent: *ŋ* is surely a pre-*k* variant of *n*, double *pp* appears to be, as in all the other languages, a variant of *p*, and *s* alternates with *t* as a part of a Siberia-wide phenomenon which in Yakut extends also over *č*, and whose exact mechanism unfortunately remains unclear (see *soččoyotoχ* in 2.21.3).

Interestingly, *m* as a closing consonant, despite being quite common in the other Turkic languages, is missing from Yakut.

Six words have more than one closing consonant possible: *bosχo* (*r* and *s*; see 2.21.3), *kyžy* (*n*, *ŋ*, and *p*), *maŋan* (*k* and *t*), *sikej* (*n* and *p*), *soyotoχ* (*č* and *s*), and *ürünŋ* (*p* and *pp*).

If we assume that *n* and *ŋ* are derivatives of *m*, and *č* and *t* are derivatives of *s* (through the $\check{c} \sim s \sim t$ alternation), then all these alternatives can be reduced to the basic set of general Turkic closing consonants, *p* and *s*. Note *kyžy* which is one of just two examples in this work, which have more than two closing consonants possible; see 3.1.1.

In twelve cases, the first vowel of the base is long or a diphthong. The four long vowels have all been shortened in the reduplication (*sinsikej* ~ *sipsikej* ‘completely raw, ...’, *sipsīnes* ‘very thin’, *tuptūstaχ* ‘oversalted’, and *upūllayas* ‘completely thawed’). In five of the seven cases with diphthongs, only the initial, high vowel of the diphthong has been reduplicated (*kipkieŋ* ‘very wide, ...’, *kūpkūöχ* ‘very: 1. blue; 2. green’, *supsuon* ‘very thick, ...’, *tiptiere* ‘completely the other way round’, and *pyparaχan* ‘very heavy’). In the remaining two, the entire diphthong has been copied (*čuopčuoyur* ‘very motley’ and *tüöptüökün* ‘big cheater’). Interestingly, in further three or more cases, the original short vowel of the base has been lengthened in the reduplication (*būsbutün* ‘absolutely all’, *suočsoyotoχ* ‘very lone(ly)’ (see 2.21.3 above), and *ūnutary* ‘completely opposite, ...’).

Shortening of the originally long vowel is the standard Turkic practice. Neither retaining length, nor the more actually adding it, seems to occur anywhere else than in these six cases in Yakut. No ready explanation presents itself. The similarity to the exclamative intonation, which is expressed by lengthening of the last vowel of the word, must be considered coincidental as with it, high vowels would have been expected to diphthongize (**büösbütün* rather than the attested *būsbutün*). Yet, it is both structurally and semantically the closest regular phenomenon in Yakut. See 3.1.12 for an alternative explanation, and 3.1.20 on shortening in general.

Structural peculiarities of Yakut *C*-type reduplications include also three words which could possibly be reclassified as a separate type of reduplications (*bosbosχo* ‘absolutely straight, ...’, *čepčepčeki* ‘very cheap’, and *tastastyŋ* ‘completely outer, ...’), and a reduplication of a word other than *ak*, which is nevertheless closed by *pp* (*üppürüŋ* ‘snow-white’). See 2.21.3 above.

2.21.5 Semantics

In almost all cases, the reduplicated meaning is a simple intensification or apparently the same as the base meaning. The only slightly less trivial evolution is to be observed in *bysbyrtaχ*, where ‘unclean, icky, foul, vile’ → ‘lousy, crappy’.

A simple, as it seems, but nonetheless noteworthy case is *sotoru*, where ‘now, immediately’ → ‘very soon’ (‘сейчасъ, тотчасъ, немедленно, въ скорости, скоро, очень скоро, вскорѣ, поспѣшно; въ тотъ же мигъ [...]’ → ‘скоренько’) which, on the face, appears to be in fact a deintensification. Most probably, both words have in reality a very similar scope of use, and it is only the brevity of the definition that creates the false impression.

The great majority of words are adjectives and adverbs, as far as the build of the Turkic languages allows.

Nine words have a pronounced substantival character: *deŋ* ‘unexpectedness, unexpected even’, *χobū* ‘slander, calumny’, *kyžy* ‘passion to act contrary to the usual way’, *sibilgin* ‘now, this instant’, *soŋū* ‘a cry, weep’, *sür* ‘fear, horror’, *symyja* ‘a lie’, *tüökün* i.a. ‘swindler, rogue, thief’, and *ürüŋŋük* ‘whiteness’.

Chapter 3

Analysis

This chapter analyses the data presented in chapter 2. First, more than twenty recurring peculiarities and phenomena are discussed, and preliminary conclusions drawn (3.1). Next, the structure and semantics of reduplications are examined (3.2 and 3.3), and finally, suppositions are made about the origins and evolution of the phenomenon (3.4).

The many conclusions that are scattered across this chapter can be found collected and summarized in chapter 4.

3.1 Recurring peculiarities

This section summarizes more than twenty peculiarities and phenomena occurring in different languages. The focus is primarily on summarizing but several conclusions are also drawn. For easier guidance, a mini table of contents is given below with characteristic examples where applicable.

| | | |
|--------|--|-----|
| 3.1.1 | Alternative closing consonants (<i>dümdüz</i> : <i>düpdüz</i>) | 158 |
| 3.1.2 | Alternative closing consonants with different meanings (<i>apak</i> ‘snow-white’ : <i>appayym</i> ‘my dear’) | 159 |
| 3.1.3 | Apparent reduplications (<i>sersem</i>) | 160 |
| 3.1.4 | Closing consonant homolocal with C_1 (<i>bapbalaža</i>) | 161 |
| 3.1.5 | Closing consonant homolocal with C_2 (<i>büsbütün</i>) | 165 |
| 3.1.6 | Closing consonant identical to C_2 (<i>japjapalâq</i>) | 167 |
| 3.1.7 | Diminutive (<i>apagaš</i>) | 168 |
| 3.1.8 | Double <i>pp</i> (<i>appak</i>) | 169 |
| 3.1.9 | Double reduplication (<i>apappak</i>) | 171 |
| 3.1.10 | Emancipated reduplicated anlaut (<i>kyp žylarjač</i>) | 171 |
| 3.1.11 | Families (<i>köpkök</i> , <i>köpögöš</i> , <i>köppöš</i> &c.) | 173 |
| 3.1.12 | Lengthening (<i>čipčikke</i>) | 174 |
| 3.1.13 | Multiple intensification (<i>apapakaj</i>) | 175 |

| | | |
|--------|---|-----|
| 3.1.14 | Nouns (<i>karajy</i>) | 176 |
| 3.1.15 | Obsolete base (<i>epeji</i>) | 176 |
| 3.1.16 | Prepended base (<i>aγappag</i>) | 177 |
| 3.1.17 | Pronouns (<i>kapkačan</i>) | 178 |
| 3.1.18 | Reduplicated binom (<i>öpörjany</i>) | 178 |
| 3.1.19 | Reduplicated anlaut not matching the base (<i>žäpžakšy</i>) | 178 |
| 3.1.20 | Shortened vowel (<i>göpgök</i>) | 179 |
| 3.1.21 | Spirantized closing consonant (<i>gafgara</i>) | 180 |
| 3.1.22 | Verbs (<i>qäpčärajmâq</i>) | 181 |
| 3.1.23 | Vocalic anlaut (<i>apak</i>) | 181 |
| 3.1.24 | ‘White’ (<i>appak</i> &c., <i>üppürün</i>) | 181 |

3.1.1 Alternative closing consonants

One base can have alternative reduplications with different closing consonants.

This is in fact the typical situation for almost all the thirteen languages which allow closers other than *p* and *pp*. See tab. 3.1 and the “Structure” subsections in the respective sections in chapter 2.

Two languages are exceptional. One is Southwestern Karaim where there are ten reduplications in *p(p)* and one in *m* (*ješil* ‘green’), and the other is Kumyk where, against 24 reduplications closed by *p(p)*, there is one closed by *s* (*bütün* ‘whole’). Both are in all probability loanwords; however, structural uniqueness being the only argument to support this supposition at my disposal at the moment, they must be counted as valid exceptions to the general Turkic tendency. See 2.6.4 for Karaim and 2.11.4 for Kumyk, and also Kklp., Kmk. *beter* ‘worse, ...’ (2.7.3 and 2.11.3, respectively), both of which were discarded as loanwords because additional arguments were available in favour of such interpretation.

A genealogical-geographical pattern emerges from tab. 3.1: apart from Ottoman, for which the data are probably severely incomplete, the variety is the greatest at the Oghuz/westernmost end, it then disappears towards the Kipchak/central area, and increases again in the Yakut/easternmost corner. This is consistent with the general geographical picture, and the distinction between “*p*-languages” and “*mprs*-languages”, which was made primarily for the purpose of subsections 3.1.4 and 3.1.5 below.

A mention must be made here about two words, Kar.E *sylak* &c. ‘wet’, and Yak. *kyžy* ‘passion to act contrary to the usual way’, which are the only ones in the present collection, whose reduplications can be closed with more than two different consonants. In the case of *sylak*, it is *m*, *p* and *r*, and in the case of *kyžy* it is *n*, *ŋ* and *p*. The Karaim forms result probably from external influence (see 2.6.2), and the Yakut ones should perhaps be attributed to phonetic assimilation. Thus both are in a certain way characteristic of their respective languages in general.

See also 3.1.2 for cases of reduplications with alternative closing consonants having different meanings.

| Language | Reduplications | |
|---------------------|----------------|--------------------------|
| | Total | With alternative closers |
| Turkish | 177 | 17 ($\approx 9.60\%$) |
| Yakut | 106 | 6 ($\approx 5.66\%$) |
| Turkmen | 53 | 4 ($\approx 7.55\%$) |
| Uzbek | 46 | 3 ($\approx 6.52\%$) |
| Eastern Karaim | 51 | 3 ($\approx 5.88\%$) |
| Azeri | 70 | 3 ($\approx 4.29\%$) |
| Gagauz | 43 | 2 ($\approx 4.65\%$) |
| Ottoman | 51 | 2 ($\approx 3.92\%$) |
| Northeastern Karaim | 12 | 1 ($\approx 8.33\%$) |
| Tatar | 59 | 1 ($\approx 1.69\%$) |
| Bashkir | 93 | 1 ($\approx 1.08\%$) |

Table 3.1: Reduplications of one base with alternative closing consonants (*p* and *pp* not considered separate).

3.1.2 Alternative closing consonants with different meanings

One base can have alternative reduplications with different closing consonants, each intensifying a different component of the base meaning.

The examples fall essentially into two groups. One collects reduplications of *ak* in a literal and figurative meaning, and the other a single exceptional case.

The primary meaning of *ak* is ‘white’ but in many languages it has developed additional, figurative and generally positive ones, e.g. ‘good, kind’ in Bashkir, ‘clean, pure; happy, cheerful’ in Tatar, ‘innocent, spotless, unblemished’ in Uighur &c.³⁸ In general, the word has proven to be unusually intensifiable (see 3.1.24), and in particular by what appears to be emphatic lengthening of the closing consonant (see 3.1.8). Thirteen such formation can be found in the present collection, but three or more are also special from the point of view of semantics.

In Bashkir, Tatar and Uzbek, *ak* reduplicates to both *apak* and *appak*. The form with a single *p* intensifies the literal meaning (\rightarrow ‘snow-white; very white; all white’ &c.), and the one with a double *pp*, which seems to only be attested with a Px1Sg, intensifies the figurative component (\rightarrow ‘my dear, my darling; my little white one’). See the “Special cases” subsections in the respective sections in chapter 2.

A more complicated case is that of Uigh. *aq*, *apaq*, $\langle \text{āpāq} \rangle$ and *appiyim* \sim *appyyym*, as it seems that native words have there mixed with Mo. *abahaj* ‘wife of a prince; lady’.

³⁸ Perversely, Tuv. *kara* ‘black’ also reduplicates to *kapkara* ‘1. pitch-black; 2. dear’; see 2.18.4.

Probably the native state was rather like that in Bashkir, Tatar and Uzbek, but the Mongolic admixture has made it difficult to state so with certainty. See 2.19.3.

Further, in Kirghiz as many as five intensifications of *ak* make use of reduplication: *apak*, *apakaj*, *apapakaj*, *appak*, and *apappak*. Of these only *apapakaj* seems to carry the figurative meaning ('very nice, very good'); see 2.10.2 and 2.10.3.

And lastly, the second group is in fact just one unusual case: Gag. *düz* 'smooth, even, straight' with its two reduplications: *dümdüz* '1. very smooth, very even; 2. openly, frankly', and *düpdüz* 'openly, frankly'. It seems as if it almost copied the division of Bashkir &c. *apak* : *appak*, but it must not be forgotten that the documentation of Gagauz is rather incomplete, and the similarity might prove illusory. See 2.5.5.

3.1.3 Apparent reduplications

Different methods can produce forms very similar to reduplications.

C-type reduplications have essentially two distinctive features in common: *se* mantics suggestive of intensification and, simultaneously, a phonetic shape such that the initial consonant, if present, is the same as the third one, and the first two vowels are the same. However, exceptions occur; see e.g. 3.1.19 for when the vowels do not match. The closing consonant is typically expected to be one of *m*, *p*, *r* or *s* but here, too, other sounds may occur. As a result, some forms which probably are not reduplications by origin appear nevertheless very much alike. They are generally omitted from the present work, but a few will be mentioned here in the way of illustration of the phenomenon.

Some cases are easier to recognize. For example, Az. *dördölçülü* (AzRS) becomes transparent as soon as the meaning is revealed: 'four volume *adj.*', Bshk. *fosfor* (BškRS58) is difficult to mistake even if it were defined as 'brilliantly burning material', and Tksh. *ebevejn* 'parents' is obviously a loanword from Arabic (أبوين *'abawajn* (Nişanyan ÇTES)). Examples can be multiplied: Az. *apatiya* 'apathy' (AzRS), Kirg. *apararat* 'apparatus' (KirgRS), Tat. *komkor* 'corps commander' (Russ. *командир корпуса*) or *pampaslar* 'pampas' (both TatRS), Uigh. *tez telegram* 'urgent telegram' (UjgRS; *tez* 'fast, quick'), &c.

Some are less transparent. For example, Az. *tez-täläsik* 'hurriedly, hastily' (AzRS; a binom), Gag. *jat-jaban* 'alien, foreign' (see 2.2.3 and 2.5.3, respectively), Kar.E **sav-saglam* '1. healthy; 2. whole' (see Stachowski K. 2010: 153f), Uzb. *badbaxt* 'ill-starred' (UzbRS; < Pers. بد بخت *badbaht* id.), or Tksh. *sersem* 'stunned, bewildered; scatter-brained, foolish' (see 2.16.3) can all be relatively easily misinterpreted as their meanings are no less intensive than is very often the case with actual reduplications, and their structure is absolutely conceivable since all *d* ~ *t*, *s* ~ *z*, *r*, and *v* are used to close the reduplicated anlaut.

Finally, some cases are unclear, e.g. Kar.E *komkos* 'very stupid', *tentek* 'very stupid, very sloven', Kklp. *kumkuwyt* 'excitation, commotion, agitation', Tuv. *šypšyk* 'most',

and also Čigil *symsymrak* ‘dish of meat cut up small’. They might be reduplications of obsolete bases (see 3.1.15 for more examples), or perhaps anything else. See the “Special cases” subsections in the respective sections in chapter 2.

3.1.4 Closing consonant homolocal with C_1

Note 1: This subsection only deals with words with a consonantal anlaut, hence the notation $\langle C_1 \rangle$ ‘the first consonant in a word’ is here equal to $\langle C \rightarrow$ ‘the initial consonant of a word’.

Note 2: It is assumed in this subsection that p was the original closing consonant; see 3.4.4.

The closing consonant can be homolocal with the first consonant of the base.³⁹

According to Hatiboğlu 1973: 39, “Türkçedeki teksesten kaçış ve ters orantı ilkele~~r~~rine göre “b, p” ya da “m” ünsüzleri ile başlayan sözcüklerin “p” ile pekiştirilmelerinin sakıncaları ortadadır”. This restriction has been upheld by Müller 2004: 150 in his Regel D: “ p ist als Überleitungslaut dann nicht (mehr) möglich, wenn das Adjektiv selber mit einem labialen Laut (b, p, m) beginnt. [...]”.

Neither of these statements says so explicitly but it might be guessed that the logic motivating them is this: it is against the Turkish phonaesthetics to close a redup~~l~~ication with p if its base begins with a homolocal sound. It can be expected that a phonaesthetic rule such as this will have a broader scope of application than just one sound in one language.

Let us now consider all the possible closing consonants, first the four common ones (the labials m and p , then the dentals r and s), and later the seven exceptional ones, \check{c} , j , k , n , η , \check{s} , and t .

Two abbreviations are used in this subsection: “ $mprs$ -languages” (Azeri, Eastern Karaim, Gagauz, Ottoman, Turkish, Turkmen, Uzbek and Yakut), and “ p -languages” (Bashkir, Dolgan, Khakas, Kirghiz, Karakalpak, Kumyk, Kazakh, Oiro, Shor, Tatar, Tuvian, Uighur and Western Karaim). See 3.2.1 for the rationale.

Labials

Two labial consonants can act as closers: p and m . Consonants which occur in Turkic anlaut and are homolocal with them are: b , f , m , p and v . The last one is not rep~~re~~sented in the current collection. The numbers are given in tab. 3.2, and remarks in the footnotes.

Conclusions for p as a closing consonant emerge with relative clarity: the $mprs$ -lan~~g~~uages tend to avoid it for bases which begin with a sound homolocal with it, while

³⁹ The term *homolocal* has been introduced here in the meaning ‘pronounced at the same place of articulation’ to escape the ambiguity of the term *homorganic*; see 1.1.1.

| <i>C-</i> | <i>p</i> -languages | | <i>mprs</i> -languages | | | | | |
|-----------|---------------------|----------|------------------------|----------|----------|----------|----------|-------|
| | Total | <i>p</i> | Total | <i>m</i> | <i>p</i> | <i>r</i> | <i>s</i> | other |
| <i>b-</i> | 33 | 31* | 59 | 25 | 3† | 3 | 26 | 4‡ |
| <i>f-</i> | – | – | 1 | – | – | – | 1 | – |
| <i>m-</i> | 10 | 10 | 10 | – | – | – | 8§ | 2¶ |
| <i>p-</i> | 1 | 1 | 4 | 2 | – | 1 | 1 | – |

Table 3.2: Usage of closing consonants with different labial consonants in the anlaut. The table is read as follows: in the *p*-languages, there are 33 reduplications of bases beginning with *b-*; 31 of them are closed by *p*, and the two exceptions are listed in the footnote; in the *mprs*-languages there are 59 words beginning with *b-*; and so on.

* Kar.NW *bomboš* ‘completely empty’ (along *bopboš* id.) and Kmk. *büsbütün* ‘absolutely all’. Neither can be deemed representative for its respective language.

† Az. *bapbalaža* ‘small, tiny’, and Uzb. *babbaravar* ‘absolutely the same, absolutely equal’ and *bopboš* ‘completely empty’. The first and the last also have alternative reduplications closed by *m*.

‡ Az. *bešbetär* ‘worse’, Tksh. *bešbeter* id., Gag. *bezbollahli* ‘absolutely clear, absolutely obvious’, and Uzb. *bütbütün* (along *büsbütün*) ‘quite complete’.

§ All eight are reduplications of *māvi* ‘blue’ and *mor* ‘purple’ in Eastern Karaim, Gagauz, Ottoman and Turkish. It is possible that they all go back to just two forms.

¶ Yak. *majan* ‘white’, closed by *k* and *t*.

|| Words in *p-* are surprisingly rare. See 3.1.4 below.

the *p*-languages use it indiscriminately for all bases. It is, however, surprising how few reduplications of bases beginning with *p-* there are; see the end of this subsection.

As for *m*, no such tendency can be observed. It is used freely to close reduplications of bases beginning with *b-* and *p-*. The conclusion, however, that escaping excessive phonetic similarity was precisely the reason for the introduction of *m* as an alternative to *p*, would be premature. Words beginning with *b-* and *p-* constitute less than 28% of all reduplications closed by *m*, which hardly suggests that *m*’s status is in any way special.

Dentals

Three dental consonants act as closers: *r*, *s* and *t*. Homolocal with them and appearing in Turkic anlaut are: *d*, *l*, *n*, *r*, *s*, *t*, and *z*. The data are given in tab. 3.3, and remarks in the footnotes.

Again, the *p*-languages must be put aside because, with just one exception, they simply always use *p* and, as was seen above, it is not in order to avoid homolocality.

In the *mprs*-languages, *s* is relatively often used to close reduplications of bases beginning with *d-* and *t-*. It is not attested with bases beginning with *l-*, *n-*, *r-* and *z-*,

| <i>C-</i> | <i>p</i> -languages | | <i>mprs</i> -languages | | | | | |
|-----------|---------------------|------------------|------------------------|----------|----------|----------|----------|----------------|
| | Total | <i>p</i> | Total | <i>m</i> | <i>p</i> | <i>r</i> | <i>s</i> | other |
| <i>d-</i> | 19 | 19 | 57 | 12 | 31 | 1 | 13 | – |
| <i>l-</i> | – | – | 3 | 1 | 2 | – | – | – |
| <i>n-</i> | 9 | 9 | 3 | – | 3 | – | – | – |
| <i>r-</i> | – | – | 1 | – | 1 | – | – | – |
| <i>s-</i> | 50 | 50 | 71 | 9 | 53 | 5 | 1* | 3 [†] |
| <i>t-</i> | 128 | 127 [‡] | 78 | 13 | 42 | 11 | 12 | – |
| <i>z-</i> | 2 | 2 | 2 | – | 2 | – | – | – |

Table 3.3: Usage of closing consonants with different dental consonants in the anlaut. See tab. 3.2 for an explanation how to read this table.

* Yak. *soyotoχ* ‘lone(ly)’; see 2.21.3.

† Three Yakut reduplications closed by *č*, *j*, and *n*; see 2.21.4.

‡ Tat. *tümtügäräk* ‘completely round’. An alternative reduplication closed by *p* is also attested.

but seeing how rare these are, no conclusions should be drawn from this fact. It is also practically unattested with bases beginning with *s*; there is only one exception, and it is in Yakut. To conclude, *s* as a closing consonant is immune to homolocality with the initial consonant of the base, but it is not used when that consonant is also *s*.

Finally, *r* is used as a closing consonant for bases beginning with *d* (merely one example), *s* (rarely), and *t* (much more commonly, almost on par with *m* and *s*). Similarly to *s* above, no conclusions can be drawn from its absence with other anlauts. To sum up, *r* can be, but relatively rarely is, used as a closing consonant for reduplications of bases beginning with a homolocal consonant, and it is unknown whether it could occur with a word which, too, begins with *r-*.

Exceptions

The remaining closing consonants are: *t* (used in two words, Az. *garyšyg* ‘mixed’ &c. (see 2.2.3) and Yak. *maġan* ‘white’), *š* (also in two words, Az. *betär* ‘worse’ and Tksh. *beter* id.), *č* (only in Yak. *soyotoχ* ‘lone(ly)’; see 2.21.3), *j* (only in Yak. *soġū* ‘a cry, weep’), *k* (only in Yak. *maġan* ‘white’), and *ŋ* (only in Yak. *kyžy* ‘passion to act contrary to the usual way’, where it is surely a result of a trivial phonetic assimilation).

All of them are so rare that suspicion must arise whether they are really reduplications. However, I know of no alternative explanations and am therefore forced to accept them at face value and conclude that conclusions are for the moment impossible.

Summary

The clear and relatively certain (supported by what seems to be sufficient data) tendencies are as follows:

1. *p*-languages use *p* as the closing consonant indiscriminately for all bases.
2. *mprs*-languages tend to replace *p* with *s*, *m* and rarely other consonants for bases beginning with *b*-.
3. *mprs*-languages use *s* and, less often, *r* as the closing consonant for bases beginning with *d* and *t*.
4. *mprs*-languages almost never use *s* for bases beginning with *s*-.

It seems then, that what is really being avoided in the *mprs*-languages is not so much homolocality as closing the reduplicated anlaut with the initial consonant of the base. (The combination *-pb-* also belongs here because in normal speech it must be expected to assimilate to [-bb-] or at least [-pb-].)

The observations of Hatiboğlu's and Müller's (see the beginning of this subsection) are technically correct because they are limited to Turkish. However, broadening the perspective shows that what appears to be the logic behind them was imprecise. This is only understandable given the small probe they had at their disposal.

Further, these phonaesthetic tendencies prove indeed to have a wider scope than just one sound in one language. At varying degrees of strictness, they apply to effectively all the eight *mprs*-languages: Azeri, Eastern Karaim, Gagauz, Ottoman, Turkish, Turkmen, Uzbek and Yakut.

See also 3.1.5 for a similar summary on homolocality of the closing consonant with C_2 as the conclusions from the two support each other, and for an attempt at generalization which is made there. The considerations started here are continued in 3.2.1, and concluded in 4.1.3.

Excursus: *p*-

One more remark remains to be made. Overall, the phonologies of the majority of the Turkic languages are roughly compatible. In particular, this applies to which consonants are permitted in anlaut. While it is not possible to precisely quantify the frequencies of specific sounds in the spoken language throughout the history, a crude opinion can be formed based on the knowledge of historical phonology and modern languages. Which words can be reduplicated is primarily defined by their semantics, but here 1198 forms have been collected, and I believe that this is enough for the frequencies of occurrence of different consonants in anlaut to correspond in an imperfect, roughly approximate way to the general values.

And indeed, this seems to be the case – except for *p*-. See tab. 3.4. The fact that most Turkic words in *p*- are Persian and later Russian loanwords should not be

relevant as apparently the great majority of reduplications have been formed after the first contact with Iranian languages, see 3.4.

It seems almost as if bases beginning with *p*- were purposefully avoided, at least until the emergence of alternative closing consonants. The available data are certainly not sufficient to support such hypothesis. In fact, they do not even point unanimously to *p* as the original closer (see 3.4.4). The disproportion, nonetheless, ought to be made note of.

| Consonant | Examples | Consonant | Examples | Consonant | Examples |
|------------|----------|------------|----------|-------------|----------|
| <i>b</i> - | 93 | <i>k</i> - | 159 | <i>s</i> - | 125 |
| <i>č</i> - | 47 | <i>l</i> - | 3 | <i>š</i> - | 27 |
| <i>d</i> - | 76 | <i>m</i> - | 20 | <i>t</i> - | 213 |
| <i>f</i> - | 1 | <i>n</i> - | 15 | <i>z</i> - | 4 |
| <i>g</i> - | 40 | <i>p</i> - | 5 | <i>ž</i> - | 23 |
| <i>h</i> - | 18 | <i>r</i> - | 1 | <i>ẓ̌</i> - | 46 |

Table 3.4: The number of bases beginning with different consonants. Language-specific phonetic differences such as *k* : *ḳ* : *q* or *h* : *χ* have been ignored; as a result, the numbers do not always correspond to those in tab. 3.2 and 3.3 above.

3.1.5 Closing consonant homolocal with C_2

Note 1: This subsection continues the considerations on the homolocality of the closing consonant with C_1 . See 3.1.4.

Note 2: Likewise, it only deals with words with a consonantal anlaut, and assumes that *p* was the original closing consonant.

The closing consonant can be homolocal with the second consonant of the base.

Both Hatiboğlu 1974 and Müller 2004 focus more on homolocality with C_1 than with C_2 , but they do also take into account other kinds of consonants further in the word: “[...] ya da sürekli ünsüzlerden biriyle kapanan sözcükler, “*p*” ünsüzüyle pekiştirilir” (Hatiboğlu 1973: 34, “Birinci Kural”), and “[...] und als weitere Konsonanten den Laut *k* [...] oder Zischlaute, aber nicht *m* enthalten. [...]” (Müller 2004: 151, “Regel E”, also elsewhere).

Despite Hatiboğlu’s and Müller’s efforts, they could only formulate tendencies, but not rules (see 1.1.2). Müller took more factors into account than Hatiboğlu and managed to slightly improve the accuracy. His rules cover ca. 79% of the examples and cannot help making a rather definite impression of overfitting, see appendix A.

Here, I will limit myself to discussing homolocality with C_2 , similarly to how it was done in 3.1.4 above. The numbers are given in tab. 3.5, and remarks in the footnotes.

| C_2 | p -languages | | $mprs$ -languages | | | | | |
|-------------|----------------|-----------------|-------------------|-----|-----|-----|-----|----------------|
| | Total | p | Total | m | p | r | s | other |
| b | 5 | 5 | 10 | – | 5 | 4 | 1 | – |
| f | – | – | 1 | – | – | 1 | – | – |
| m | 37 | 37 | 26 | 3* | 10 | 6 | 7 | – |
| p | 3 | 3 | 14 | 1 | 2 | 7 | 4 | – |
| v | 3 | 3 | 17 | – | 6 | – | 11 | – |
| d | 2 | 2 | 7 | – | 3 | – | 4 | – |
| δ | 4 | 4 | – | – | – | – | – | – |
| l | 44 | 44 | 56 | 5 | 39 | 2 | 10 | – |
| n | 30 | 30 | 18 | 1 | 15 | – | 2 | – |
| r | 85 | 85 | 101 | 10 | 76 | 1 | 13 | 1 [†] |
| s | 14 | 14 | 13 | 3 | 7 | 1 | 2 | – |
| t | 16 | 15 [‡] | 35 | 2 | 13 | – | 17 | 3 [§] |
| \emptyset | 1 | 1 | – | – | – | – | – | – |
| z | 32 | 32 | 36 | 13 | 19 | 4 | – | – |

Table 3.5: Usage of closing consonants with different consonants as C_2 . See tab. 3.2 for an explanation how to read this table.

* For this and other cases where the closing consonant is the same as C_2 , see 3.1.6.

[†] Az. *gatgaryşyg* ‘1. mixed, complicated, entangled; 2. scattered, disordered; 3. uncombed’, see 2.2.3.

[‡] Kmk. *büsbütün* ‘the absolute whole’. This reduplication is not at all representative for Kumyk; see 2.11.3.

[§] Az. *beşbetür* ‘worse’, Tksh. *beşbeter* id., and Uzb. *bütbütün* ‘quite complete’. For the last one, see 3.1.6.

The following tendencies can be extracted from tab. 3.5 with an acceptable degree of confidence:

1. p -languages use p as the closing consonant indiscriminately for all bases.
2. $mprs$ -languages do not avoid homolocality of the closing consonant with C_2 .
3. $mprs$ -languages tend to avoid using C_2 as the closing consonant.
4. The tendencies referring to C_2 are weaker than those which refer to C_1 .

The relation of the closing consonant with C_2 is then quite the same as with C_1 (see 3.1.4), only weaker. The general tendency in the $mprs$ -languages, therefore, appears to be rather simple: the closing consonant should not be the same as C_1 , and it is better if it is not the same as C_2 . homolocality, it seems, does not really play a role here.

But in theory, another explanation is also possible. See 3.1.6 for a discussion of special cases where the closing consonant is the same as C_2 . The considerations started here are continued in 3.2.1, and concluded in 4.1.3.

3.1.6 Closing consonant identical to C_2

Note: This subsection continues the considerations on the homolocality of the closing consonant with C_2 . See 3.1.5.

The closing consonant can be the same as the second consonant of the base.

There is a relatively numerous group of words with an intensive meaning where the first three sounds are the same as the triple following them. Three groups can be distinguished:

1. the reduplication is in a p -language, and the closing consonant and the second consonant of the base are both p : Bshk. *täptäpäš(äk)* and Kzk. *taptapal*, all ‘very low’ (note that these two words might actually be common Bashkir and Kazakh heritage rather than independent innovations),
2. the reduplication is in an $mprs$ -language, and the closing consonant and the second consonant of the base are the same and one of m , p , r , s (unlike in group three): Az. *jumjumšag* ‘very soft’, Kar.E *jumjuma(r)lak* ‘completely round’, Tksh. *perperišan* ‘completely scattered, disordered, distraught, miserable’, Uzb. *dumdu malâq* ‘completely round’, *japjapalâq* ‘completely flat’, and Yak. *bosbosxo* ‘absolutely straight, upright’, *čepčepčeki* ‘very cheap’, and *tastastyŋ* ‘completely outer, complete outsider’, and
3. the reduplication is in an $mprs$ -language, and the closing consonant is not one of m , p , r , s : Uzb. *bütbütün* ‘quite complete’ and *pakpakana* ‘very shot of a person’; maybe also Ott. *daydayan* ‘scattered, cluttered, all over the place’.

As for the first group, there seems to be actually no reason to treat it in a special way. The p -languages simply close all of their reduplications with p (see 3.1.4 and 3.1.5), and apparently also allow words with p as C_2 to be reduplicated.

The second group is more difficult to interpret. The data presented in 3.1.5 suggest that using C_2 for the closing consonant is avoided. Objective arguments seem to be missing with which to determine whether the words in this group are simply exceptions to this general tendency or perhaps examples of some separate type of reduplication where the first three sounds would be reduplicated.

The third group contains two or maybe three words which are not clear. Of these, Ott. *daydayan* can be now set aside because it is most probably not a reduplication at all, see 2.13.3. Uzb. *bütbütün* might be suspected of resulting from a trivial phonetic assimilation, but there seem to be no parallel examples to support this explanation.

The last word is *pakpakana* for which I am presently unable to offer a solution other than the hypothetical separate type of reduplication.

Lastly, the unclear word *symsymrak* ‘dish of meat cut up small’ (see 2.1.3) should be mentioned. It is not certain that it is a reduplication at all but if it is, it also belongs here.

Such a separate type of reduplication would have to involve the repetition of the first three sounds of the base and prepending them to the base with no intervening element in between. Perhaps the most likely candidate is the otherwise unclear Uzb. *pakpakana* ‘very short of a person’, but in theory, all the examples mentioned in this subsection need to be reconsidered.

Furthermore, a number of words which are not typically considered reduplications but which do nevertheless begin with two identical triples of sounds could potentially be reinterpreted as representatives of such a hypothetical new type. Examples: Bshk. *baš-baštak* ‘willful, headstrong’ (BškRS58, BškRS96), *min-minlāk* ‘conceit; selfishness’ (BškRS58), Kar.E *bok-baklavat* ~ *bok-boklavat*⁴⁰ (KRPS, RKarS-Haf), *boš-bošyna* ‘in vain’ (KRPS), *kün-kündüz* ‘in the middle of the day’ (RKarS-Haf), and potentially more.

The origin of our hypothetical new type of reduplication is a problem. The formations included in the three groups above rarely appear in more than one language. Also, their stems are essentially not among the most commonly reduplicated ones (see 3.2.4). This suggests that they are not remnants of an ancient method but rather innovations, and most likely independent of each other. It would probably have to be assumed that this new type evolved more or less spontaneously in five or six different languages.

Postulating a new type of reduplication for the formations presented here is riddled with difficulties. Considering that its motivation is effectively the fact that one word, Uzb. *pakpakana*, is unclear, the idea is perhaps better abandoned.

See 3.2.1 for further observations on the choice of the closing consonant, and 4.1.3 for a conclusions of these considerations.

3.1.7 Diminutive

Reduplication can be combined with diminutivization.

The present work collects 28 such formations in nine languages: Az. *jupjumruža*, *tärtäzäžä*, Khak. *appagas*, *köppeges*, Kirg. *apapakaj*, *kipkičinekej*, *taptatynakaj*, Kklp. and Kzk. *kipkiškentaj*, Oir. *apagaš* and five related words, *köpögöš* and four related words, OTkc. *oposalkyja*, Shor *apagaš*, *apapagaš*, *appagaš*, *köpegeš* and *köppegeš*, and Tksh. *dapdaražyk* and *dasdaražyk*. See the “Standard cases” and “Special cases” subsections in the respective sections in chapter 2.

⁴⁰ The base **ba|klavat* does not seem to be attested. Possibly, the whole was formed in the image of Tksh. *takym-taklavat* ‘all together’ which, in turn, was probably shaped by analogy to Arabic loanwords in *-(av.)at*, from *takim* ‘set, group’ (Nişanyan ÇTES).

Amounting to $\frac{2}{3}$ of all the examples, the South Siberian words belong effectively to one family of similar reduplications of just two bases, *ak* ‘white’ and *kök* ‘blue’ (see 3.1.11 on families of reduplications). Note also that Tuvian is the only South Siberian language included in the present work which has apparently no parallel forms. Moreover, Karakalpak and Kazakh share their only representative, and the same word is also present in Kirghiz. In effect, it is only Azeri, Old Turkic, Turkish and partly Kirghiz, comprising just a quarter of the words, that seem to have evolved this kind of multiple intensification independently.

In the majority of cases, it seems impossible to decide whether it was the reduplication that was enhanced with a diminutive suffix or the other way round. Only with regard to the South Siberian examples some suppositions can be made; see 2.12.3.

See 3.1.13 for other methods of strengthening intensification.

3.1.8 Double *pp*

The closing consonant *p* can be doubled to *pp*.

With at least fifteen examples from thirteen languages to its name, the phenomenon seems to be relatively common. However, thirteen of these examples are reduplications of *ak* ‘white’ (see 3.1.24 on the unusual productivity of the word, and 3.1.2 on the meaning of *apak* vs. *appak*), and the other two are unrelated singular innovations: Kar.E *appačyk* ‘wide open’, possibly also ‘very bright’, and Yak. *üppürüj* ‘snow-white’.

Its distribution follows a pattern, see map 3.1. Out of the languages included in the present work, it is present in all the Karakhanid and Kipchak ones (Uighur and Uzbek; Bashkir, Karaim, Karakalpak, Kazakh, Kirghiz, Kumyk and Tatar), and only sporadically in other groups: Oïrot, Ottoman, dialectal Turkish (TTAS), and also among the oldest attestations. See the “Standard cases” and “Special cases” subsections in the respective sections in chapter 2.

Further, four members of the South Siberian *apagaš* family might also belong to this group: Khak. *appagas*, Oir. *appāš*, *appagaš* and Shor *appagaš*, but other interpretations of these forms do not seem to be less likely; see 2.12.3, and also 3.1.11.

Double *pp* must have then arisen as a closing consonant at least thrice: probably in Old Turkic from which it was inherited to modern languages, and also later independently in Karaim, and Yakut.

The Kar.E *appačyk* ‘wide open’ might have been shaped after Kar.E *appak* ‘snow-white’, or independently. The Yak. *üppürüj* ‘snow-white’ and OTkc. *apak* id. are most probably the latter. At least three explanations seem conceivable:

1. contraction of a double reduplication,
2. contraction of a phrase with the *ma* particle, or
3. emphatic lengthening.



Figure 3.1: Geographical distribution of double *pp* as the closer. A geographical and genealogical pattern can be observed, see the main text.

A double reduplication, **ap.ap.āk* (see 3.1.9), can be easily believed to have contracted to *appāk*, the middle syllable being unstressed. However, such formations are very rare, and unlikely to occur until the reduplication had had been in use for so long as to see its emphatic load rather weakened. Knowing that *appak* had already existed in Old Turkic, this option can probably be dismissed here.

Roos 2000: 79 suggests that WYug. *appaq* ‘very white’ < *aq pa aq* < *aq ma aq*, where *ma* is a particle which, among other things, expresses emphasis or another meaning also imaginable here, ‘moreover, too, also’ (Roos 2000: 153). But our forms should require a less language-specific solution. It can be found in OTkc. *ma* ‘and’ (Erdal 2004: 347f).

A similar particle, *mi*, can be used in Turkish to intensify the meanings of adjectives, e.g. *güzel mi güzel* ‘how nice, very nice’. It is not clear whether it is etymologically identical to the interrogative particle *mi* ‘if, whether’ ≪ OTkc. *mu* (the latter is discussed in Erdal 2004: 411f and on other pages).

Overall, the second option (a phrase with a particle) appears to be too general. Such mechanism could be applied to virtually any intensifiable adjective, and it seems unlikely that its traces should only survive in those three, where the closer is double *pp*. See also 3.4.4 on the application of this idea to the very origins of reduplication.

Finally, lengthening can be employed in the Turkic languages as a means of adding emphasis. It is not a particularly frequent phenomenon but at least of the three options offered here, this seems to be the most plausible one. Double *pp* would then be a special case of lengthening, see 3.1.12 below.

Despite this, however, *pp* must be treated as a separate closer to a certain degree because there exist languages (Karakalpak and Kazakh) where the reduplication of *ak* can only be closed with it, and no other closing consonant. In this sense, double *pp* is not “just a phonetic variant” of *p* in as much as the spirantized versions (see 3.1.21).

3.1.9 Double reduplication

Words can be reduplicated more than once.

In the current collection, there are only three or four examples of this phenomenon: Kirg. *apapakaj*, *apapak* and Shor *apapagaš*, all ‘very white’ (see 2.10.3, 2.14.3 and also 3.1.7), and possibly also Bshk. *šypšym* (see 2.3.3). More can be found in dialectal Turkish and perhaps other non-literary idioms, which however lie outside of the scope of the present work, e.g. *abapačyk* or *apappažyk*. Interestingly, all are reduplications of *ak* ‘white’; see 3.1.24 for the extraordinary productivity of the word.

Note that doubly closed reduplications, such as Yak. *čybysčylās* ‘very warm’ or *tarys* ≠ *tarayaj* ‘completely bald’ (Pekarskij 1907–30) are here considered a separate, *CVC*-type rather than double reduplications.

Double reduplication might have also given rise to the forms with a double *pp* for a closer (see 3.1.8), although alternative explanations seem more plausible.

See 3.1.13 for other methods of strengthening intensification.

3.1.10 Emancipated reduplicated anlaut

The reduplicated anlaut, together with the closer, can be promoted to an independent intensifier.

In six languages, lying suspiciously in one belt from Kumyk through Uzbek to Yakut, peculiar forms exist which are typically one syllable-long and serve as intensifiers for a strictly limited number of words. At least one of them is in all probability a severed reduplicated anlaut by origin. Let us begin with this one, and then proceed eastward from Kumyk listing the mostly uncertain examples, and finally recall an interesting parallel from al-Kāšyarī’s dictionary and modern interpretations of reduplication.

It is Kirg. *kyp*, which KirgRS85 (s.v. *кын II*) defines as “усиление к словам, начинающимся на *кы*” before proceeding to give just one example, *kyp(a) žyłaŋač* ‘stark naked’. To be sure, **kylaŋač* is not only impossible as a phonetic variant of *žyłaŋač*, but also not attested as a separate and just accidentally similar root. However, *žyłaŋač* ‘naked’ is, and, more importantly, it is attested in the phrase *kypkyzyl žyłaŋač* ‘stark naked’, lit. ‘red-naked’ (KirgRS85 s.v. *кыпкызыл*). The shape *kyp* itself is also attested as a “подражательное слово” (s.v. *кын I*) in two phrases: *kyp edip* and *kyp dep*, both ‘immediately’. It is not certain whether or how *kyp I* and *II* compare. The origin of Kirg. *kyp žyłaŋač*, nevertheless, seems to be clear.

The phrase is also present in Kumyk: *kyp jalangač* id. (KmkRS), and a similar one can be found in Turkmen: *čyp jalanač* id. The latter is more mysterious; maybe related to the similarly unclear *čypčynym* ‘absolute truth, sheer truth’ (see 2.17.3)?

In Kumyk, there exist at least three more such compositions: *kap ortasy* ‘the exact middle’, *kap jartysy* ‘precisely a half’ (both KmkRS, see also Khakas below), and *šam jalangač* ‘stark naked of a person’. Two reduplications are attested in the present collection that begin with *kap-*: *kapkara* ‘pitch-black’ and *kapkarangy* ‘complete dark(ness)’, and none with *šam-*, or even closed with *m* at all.

For Turkmen, Clark 1998: 511 lists four intensifiers: *čym* (+ *āk* ‘snow-white’ and *gyzyl* ‘bright red’, see also below), *duv* (+ *āk* ‘pale white, ghost-white’),⁴¹ *dym* (+ *garaňky* ‘very dark’, see fn. 50 for parallels), *šar* (+ *gara* ‘jet-black, raven-black’), and *saň* (+ *gaty* ‘rock-hard’). With the last one, Pers. سنگ *sang* ‘stone’ is mentioned, and probably it is indeed the actual source of Trkm. *saň*. Two examples from TrkmRS can be added: *dym garaňky* ‘very dark’ (see fn. 50 for parallels), and *gap bīl* ‘the exact middle’. Of the above, only *gap* has reduplications beginning with it in the current collection: *gapgadyrly* ‘dear, cordial’, *gapgara* ‘pitch-black’, and *gapgaraňky* ‘complete dark(ness)’.

For Uzbek, Kononov 1960: 162 mentions four words: *yirt* (+ *χām* ‘completely unripe’, *jālgān* ‘a blatant lie’ (UzbRS41), and *savādsiz* ‘completely illiterate’), *lik* (+ *tola* ‘absolutely full’), *tim* (+ *qāra* ‘completely black’, see fn. 50 for parallels), and *žiqqa* (+ *hol* ‘completely wet’).⁴² Only *tim* also appears in a standard reduplication, with *tin* → ‘absolutely still, steady’.

For Kirghiz, I do not know of examples other than *kyp žylanač* discussed above.

For Uighur, too, I am only aware of one example. It is *žim*, which UjgRS classifies as an independent word with the meaning ‘quietly, calmly; silently, tacitly’, and exemplifies with the phrases *žim (žim)!* ‘be quiet, shut up!’, *žim turmaq* ‘to keep quiet’ and *žim bolmaq* ‘to quiet down, to grow silent’. However, it is fairly possible that the word was originally the reduplicated anlaut of **čymčyrt* or a similar shape, see 2.19.3.

For Khakas am I aware of one example in two phonetic variants: *haborta* and *hap orta* (see also Kumyk above). HakRS and Subrakova 1999 differ on the exact meanings but their general semantics is ‘right, exactly, precisely: half, on target, on time’. Possibly, also the shape *hap-sort* ‘кстати, как раз’ belongs here (attested in HakRS, where it is suggested to be a dialectal variant of *haborta*, and Subrakova 1999). The voiced variant is not used in standard reduplications, and *hap* is attested in *haphara* ‘jet-black’ and *hapharashy* ‘completely dark’.

⁴¹ Also with *dagyn* ‘scattered, diffused’ (TrkmRS). This combination is most probably related to Ott. *dardaŷan* ‘scattered, cluttered, all over the place’, and unclear. See 2.13.3.

⁴² This intensifier is more likely unrelated to reduplication. CV-type reduplications are particularly numerous in Uzbek (not discussed in the present work), and they do typically have their closing consonant doubled (e.g. *duppadurust* ‘very good, decent’, *jāppajālyiz* ‘completely alone’ (both UzbRS) &c.), but it is always *p*.

For Yakut, I can mention two intensifiers: *op* (+ *sollon* ‘excessive greed’), and *suot* (+ *čygyńaχ* ‘naked’, *čat̄y* ~ *sat̄y* ‘on foot’, and *symyja* ‘deceitful’). In standard examples, the first one can also be found with *oččuguj* ‘small, tiny’. The other one not as such, but see Dolg. *suotčogotok* ‘completely alone’ and a variety of similar forms in Yak. in 2.21.3, and also the reduplications of *soŋū* ‘a cry, weep’ and *sotoru* i.a. ‘now, immediately’ in 2.21.2.

Possibly, one more form should be added here. It is *čim* which al-Kāšyari defines as ‘an exaggerative particle of dampness or rawness’ and attests in two phrases: *čim jig ät* ‘very raw meat’, and *čim öl tōn* ‘a very damp garment’ (Dankoff/Kelly 1982: 267). However, Clauson 1972 mentions similar shapes: *čym* ~ *šym* with *ak* and *kara*, ‘plain white’, ‘pure black’, and *čyŋ* with *tolu* ‘full’. The word might be related to Trkm. *čym* above, and maybe also to Uigh. *žim(žit)* (above) and Bshk. *šym* ‘quiet’. See *čimjig* in 2.1.3.

It is not possible at present to determine with certitude which of all of these forms originate from reduplicated anlauts. The most likely candidates, it seems, are only Kirg. and Kmk. *kyp*; al-Kāšyari’s *čim* and Trkm. *čym* are also interesting. The phonetic similarity between the last two, Bshk. *šym*, and Uigh. *žim* is difficult to oversee, but the issue deserves a dedicated study beyond our present scope.

Note that the existence of a phrase similar to *kypkyzyl žyłaŋač*, while fortunate, is not a sine qua non. Morphemes can become independent words regardless of phrases &c. they are used in, if only their semantics is sufficiently clear, see e.g. Engl. *ish*. This is in line with al-Kāšyari’s understanding of *čim* which, incidentally, is not at all only characteristic of the 11th century as dictionaries have treated reduplicated anlauts as separate entries and defined them as ‘intensifying particles for words beginning with ...’ well into the 20th century (see 1.1.2).

In theory, the reverse process is also possible, whereby an adjective becomes so tightly bound to a specific noun that it never or almost never appears with any other, as e.g. Pol. *wierutny* ‘arrant’ with *bzdura* ‘nonsense’ and *kłamstwo* ‘lie’. Nevertheless, this is quite a rare phenomenon, and cases where the eventual combination possesses the appropriate phonetic shape of a reduplication must be most infrequent, rendering such explanation rather ill-suited to the Turkic examples above.

See also 3.4.4 for a possible use of the above to explain the origins of reduplication.

3.1.11 Families

Reduplications can grow out into entire families of related forms.

Naturally, very many reduplications in different languages can be eventually reduced to one (proto-)root and thus considered siblings in one big family (eg. Az. *gara* ‘black’, Tksh. *kara* id., Trkm. *gara* id., &c.). Two subgroups, however, are special because they have a reduplication for a parent. They are the south Siberian derivatives

of *ak* ‘white’ and *kök* ‘blue’, seventeen or eighteen forms in total (see 2.12.3, and also 3.1.24 for the unusual productivity of *ak*).

Also other sizeable subgroups can be distinguished where, however, the specific relations between their members are yet to be resolved. In particular, it is not clear in most cases whether the declensional and derivational suffix have been added to a reduplication, or the other way round, i.e. which of these forms are actual reduplications. The most prospective candidates are listed below. The exact etymologies remaining presently unknown, the numbers must be considered provisional. For details see the “Standard cases” and “Special cases” subsections in the respective sections in chapter 2.

The largest of these groups, with approximately fourteen members in six languages is the one centred around **tin* or a similar shape with the meaning ‘silent, quiet, calm, peaceful’. It comprises: Bshk. *tymyk*, *tyn*, *tynys*, Kklp. *tynyk*, Kzk. *tynyk*, *tynyš*, Tat. *tymzyk*, *tyn*, *tynyč*, Uigh. *teč*, *tin*, *tinč*, *tiniq*, and possibly also Ott. *dipdiñsüz*. See also 3.2.4.

It is followed by the North Siberian family of *soyotoχ* ‘lone(ly)’ which comprises thirteen forms in Dolgan and Yakut (see 2.21.3).

Later is the family of *čymčyrt* ‘complete silence’ which has up to eight members in seven languages. The parent *čyrt* exists still in different languages but apparently only with the meanings related to ‘tearing’, and not any more to ‘silence’. See also 3.1.15 on obsolete bases.

After it, comes the group of four reduplications of *garyš(yg)* ‘mixed’ in Azeri, another one of three reduplications of *dar(ažyk)* ‘narrow’ in Turkish, and a number of pairs such as Az. *göj(lük)* ‘1. blue; 2. green’, Az. *jalnyz(žā)* ‘lone(ly)’, and others.

Perhaps more interesting are also two pairs of phonetic variants: Az. *jalgyz* : *jalnyz* ‘lone(ly)’, and Tksh. *jašyl* : *ješil* ‘green’. The Azeri reduplications are both closed by *p*, but the Kipchak-like Tksh. *jašyl* is closed by *p*, while its Oghuz-like counterpart *ješil* can be closed by both *m* and *p*. This is consistent with the general characteristics of the two branches; see 3.2.2.

3.1.12 Lengthening

Single phonemes can be lengthened (doubled) to add emphasis.

The phenomenon is a cross-linguistic one, see Blevins 2004: 174. Turkic examples, however, are relatively few and, particularly among reduplications, uncertain.

They fall into five groups. The clearest one only contains a single example, Oir. *čike* ‘straight, right, accurate’ which, besides the regular *čipčike* ‘exactly, precisely’, also reduplicates to *čipčikke* id. The second, less clear, comprises North Siberian reduplications with long vowels (see below). The third, also uncertain, reduplications with a double *pp* (see 3.1.8), the fourth, if it belongs here, four members of the South Siberian *apagaš* family (see 3.1.8 and also 3.1.11), and finally the fifth, if it belongs here, four Ottoman shapes (see below).

The North Siberian words (Dolg. *suotčogotok* ‘completely alone’, Yak. *būsūtün* ‘absolutely all’, *čuoščuoγur* ‘very motley’, *suočsoγotoχ* ‘very lone(ly)’ &c., *tüöptüökün* ‘big cheater’, and *ūnutary* ‘completely opposite, ...’; see 2.4.4 and 2.21.4) display a conspicuously irregular long vowel in the reduplicated anlaut, including even cases where the one in the base was short. Since a regular explanation falls short (exclamative intonation, see 3.1.20), an irregular one can be proposed. Noteworthy here might be the unclear Yakut shape *büttün* ‘without exception’ (Pekarskij 1907–30), no doubt related to the regular *bütün* ‘all, whole’, and the variety of forms derived from Dolg. *sogotok* ‘alone, lone(ly)’ ~ Yak. *soγotoχ*. Possibly the unclear Yakut forms *χōn χotojon* and *mōn mötöj* (see 2.21.3) also belong here.

The Ottoman words with a long vowel in the reduplicated anlaut are: *māsmāvi* ‘very blue’, *sāmsāfi* ‘absolutely pure’, *sāpsāry* ‘very yellow’, and *tāstamām* ‘absolutely right, ...’. In the first three, the length is probably just an orthographic device, but in the last one emphatic lengthening is a more likely explanation; see 2.13.3.

Maḥmūd al-Kāšyarī adduces two shapes that are relevant here: *essiz* ‘impudent, treacherous, shameless, wicked’ and *arriy* ‘very clean’, explaining in both that “the doubling is for exaggeration” (Dankoff/Kelly 1982: 162). In addition, the Dagur alternative reduplications *xubxulaang* ~ *xuubxulaang* ‘deep red’ can be mentioned, as Tsumagari 2003: 135 considers the latter to be a case of emphatic lengthening.

Even if irregular and unpredictable, emphatic lengthening appears to be a satisfactory explanation for reduplications with a double *pp* and for the North Siberian forms, and to be acceptable as at least a working interpretation for the four members of the *apagaš* family (see 3.1.11).

3.1.13 Multiple intensification

Reduplication can be combined with different means of intensification, including itself.

It seems only natural that the emphatic load wears away after time and what used to be an intensive form becomes trite and potential candidate for intensification.

In the present collection, five types of multiple intensification can be distinguished, where reduplication is used together with: a diminutive (see 3.1.7), another reduplication (3.1.9), emphatic lengthening (3.1.12, including a special case where it is the closing consonant that is doubled), a doubled, prepended base (3.1.16), and with a binom (3.1.18).

Only the first method, combination with a diminutive suffix, is relatively numerously attested; the remaining four are at best sporadic. In general, it appears that the highest proportion of multiple intensifications can be found in Kirghiz.

In most cases the order of composition cannot be reconstructed based on the data collected here. In some words, particularly those from the SSib. *apagaš* family (3.1.11), conjectures can be made that diminutive must have been applied after reduplication, but they are merely based on the apparent lack of attestations of **agaš* &c.

A mention has to be made here about those words which amass more than two intensifying elements: Az. *aγappag*, Khak. *appagas* id., Kirg. *apapakaj*, *apappak*, Oir. *appāš*, *appagaš*, and Shor *apapagaš* and *appagaš*. All stem eventually from *ak* ‘white’, and all mean primarily ‘snow-white’. In some of them it is not certain exactly which intensifiers have been used to create them, e.g. the double *pp* in Oir. *appāš* could result from contraction (\ll **apapagaš*), emphatic lengthening (**apagaš*), or simplification (**akapagaš*), see 2.12.3, but it seems to always be the case that three intensifying elements have been used. See 3.1.24 on the unusual productivity of ‘white’, and of *ak* ‘white’ in particular.

Possibly, also Kar.E *tüztümüz* and Oir. *akpāš* accumulate three intensifying elements, see 3.1.16.

Although essentially not discussed in the current work, dialectal forms need to be mentioned because they do sometimes hoard even more than three intensifiers, e.g. in Tksh. dial. *ap.ap.p.a.žyk* ‘snow-white’ (DS) one finds two reduplications, an emphatic lengthening, and a diminutive.

3.1.14 Nouns

Nouns can be reduplicated.

Examples of reduplicated nouns seem to be quite numerous and appearing in an array of languages, but an exact specification is not possible. The boundary between adjectives and nouns is rather leaky in the Turkic languages and the classical Graeco-Roman distinction of parts of speech too often simply does not hold water.

Certain peculiar cases, such as e.g. Kirg. *ynak* ‘1. clean; 2. close friend’ → *ypynak* ‘very close friend’, or Yak. *tüökün* i.a. ‘swindler, rogue, thief’ → *tüöptüökün* ‘big cheater’, are enumerated in the “Semantics” subsections in the respective sections in chapter 2. Three cases, however, need to be mentioned here. Firstly, Az. *bäzäk* ‘decoration, decorative’ and *söküntü* ‘chip, splinter’ both have a relatively clear substantival character which also unusually manifests itself in reduplication, and thus effectively changes its function from the intensifying to the pluralizing; see 2.2.5. Secondly, al-Kāšyarī attests the pair *jazi* ‘steppe, ...’ : *japjazi* ‘wide open space’ which, although recored also with an adjectival meaning in other sources, is the earliest example of a reduplicated noun that I am aware of.

See also 3.1.17 and 3.1.22 for reduplications of pronouns and verbs, respectively.

3.1.15 Obsolete base

Reduplications can outlive their bases.

Two examples of this phenomenon can be found in the present collection, illustrating two different ways for a base to become obsolete.

One is modern Tksh. *epeji* &c. ‘quite, fairly’ << Ott. *epi* ‘good’ > Tksh. *iji* id., for which see 2.16.3. Here, the base did not go out of use as such, but its phonetics has changed and the reduplicated anlaut did not follow suit, yielding in effect a form that from the current synchronic perspective can no longer be considered a reduplication.

The other one is a family of cognate words scattered across several languages (see also 3.1.11 for other families), which pivots around *čymčyrt* ‘complete silence’. Its members are: Kar.E *čymčyrt* ‘complete silence’, Kklp. *žymžyrt* ‘quiet(ly), silence’, Kzk. *žymžyrt* ‘complete silence’, Uigh. *žimžit* ‘sudden silence; completely silent, completely mute’ (see also 3.1.10) and *žimžitliq* ‘complete silence’, Uzb. *žimžit* ‘completely silent, completely mute’, and possibly also Khak. *symsyryh* ‘silence, quiet(ly)’, Kirg. *tyntyrs* ‘complete silence’, and Kzk. *tyntyrys* ‘(in) complete silence, completely silent, completely mute’. See the “Special cases” subsections in the respective sections in chapter 2.

The base *čyrt or similar does not seem to be attested in any of these languages with a meaning closely resembling ‘silence’. However, there exists in different languages a verbal root *čyrt* &c. with the meanings ‘to tear, to rip, to cleave &c.’ (see VEWT s.v. *jyrt*), and also Tksh. *žyrt* ‘the sound of tearing paper, cloth &c.’.

This last word allows a semantic bridge to be imagined: ‘tearing’ → ‘the sound of tearing paper’ → ‘a quiet sound’ → ‘silence’. Should it prove true, it would also allow to connect the seemingly unrelated meanings of Khak. *syryh*: ‘fine snow; drizzle’, ‘arrow with a tetrahedral head’, and ‘bone plate used for whistling arrows; whistling arrow’ (‘shred to pieces’ : ‘cutting arrow’ : ‘the sound of ripping through the air’).

Overall, the family seems to deserve a separate study beyond the scope of the present work. Here, it must only be noted that if the first syllables of *čymčyrt* &c. are really reduplicated anlauts, then they are closed by *m*, which suggests an Oghuz origin (see 3.2.2) and thus further complicates the issue.

See also 3.1.3 for more possible cases of obsolete bases.

3.1.16 Prepended base

The base word might be prepended again to the reduplication.

In the core data analysed in the present work, there seem to only be three examples of this phenomenon, and neither is absolutely certain. The least questionable is Az. *ayappag* ‘very white’, then Kar.E *tüztümüz* ‘very straight’, and finally the most uncertain Oir. *akpāš* ‘very white’. It is always phonetics that raises suspicions, regarding the base in the Azeri and Karaim forms, and the reduplicated anlaut in the Oirot word. See the “Special cases” subsections in the respective sections in chapter 2.

Clearer cases, it seems, can be found in Turkish dialects, but as they lie outside of the primary scope of the present work (see 1.2.1), I will only limit myself here to listing a few forms in the way of parallel examples.

All words that I am aware of are eventually reduplications of *ak* ‘white’, but in Turkish dialects, they are aplenty. DS s.v. *ağabbak* and *akabbak* lists no less than 33 shapes,

including some relatively clear ones such as ⟨ağappağ⟩, ⟨ahampak⟩, ⟨akabbacık⟩ or simply ⟨akappak⟩, and a few more exotic and possibly uncertain ones such as ⟨akbacak⟩, ⟨akcacık⟩ or ⟨akpacık⟩ (? < *ak* ‘white’ + *pak* ‘clean, pure’ + *-çyk* DIMIN.).

As far as the origin of this method is concerned, the influence of Russian can be suspected since compositions such as *белый-пребелый* lit. ‘white – very white’ &c. are fairly common in it. But Russian would only be a likely source of the Azeri, Karaim and Oïrot forms, and much less so of the dialectal Turkish ones. Native innovation can be offered as an unprovable, yet plausible alternative explanation.

See 3.1.13 for other methods of strengthening intensification.

3.1.17 Pronouns

Pronouns can be reduplicated.

There is, however, only one example to support this claim, Kirg. *kačan* ‘when?’ which reduplicates to *kapkačan* ‘(very) long ago’; see 2.10.3. It is noteworthy that the only two depronominal adjectives in the present collection, *kačanky* ‘1. related to what time? 2. old, past, earlier’ and *kajdagy* ‘located at any place’ are both Kirghiz words, too.

See also 3.1.14 and 3.1.22 for reduplications of nouns and verbs, respectively.

3.1.18 Reduplicated binom

Compositions (hendiadyses) can be reduplicated.

There is only one example of this phenomenon in the present collection. It is Bshk. *ör-janğy* and it is not, in fact, entirely clear. Probably, it stems from the composition of **ör* (: Yak. &c. *ür-* in *ürüñ* ‘white’) with *janğy* ‘new’, yielding *‘white-new’ = ‘brand-new’. The whole is reduplicated to *öpörjanğy* ‘brand-new’. See 2.3.3.

See 3.1.13 for other methods of strengthening intensification.

3.1.19 Reduplicated anlaut not matching the base

Note: Vowel shortening is discussed in 3.1.20 and omitted here.

The reduplicated anlaut does not always match the anlaut of the base.

In the majority of cases, the mismatch is caused by a change, purely phonetic in nature, which only occurred in the base, or only in the reduplicated anlaut, and did not propagate to the other element. Sometimes also contamination with a foreign pronunciation might need to be taken into consideration. Examples include: Kar.E *ča|yrčebik* ‘very quick(ly)’, *čöpčevirtin* ‘from all around’, *čöpčüvre* ‘all around’, *jymješly* ‘very green’ and *sypslah* ‘completely wet’ (usually along regular forms; see 2.6.3), Ott. *öpuzun* ‘very long’, *topdolu* ‘absolutely full’ (see 2.13.3), Tksh. *epej* &c. ‘quite, fairly’ (see 2.16.3), Karakalpak bases in *ža-*, Uighur unlauded bases, and several special cases (see below for the last three).

Karakalpak bases beginning in *ǰa-* regularly have their anlaut reduplicated as *ǰä-*, and only in a half of the cases also, as an alternative, without modification (e.g. *ǰasyl* ‘green’ → *ǰapǰasyl* ~ *ǰäpǰasyl*, see 2.7.4). In Uighur, reduplications reportedly regularly reflect the pre-umlaut shape, as in *seriq* ‘yellow’ → *sapsერიq*. Two examples are especially instructive: *ješil* ‘green’ → *javjäšel* ~ *japješil*, and *taqir* ‘smooth, bare’ → *tapteqir* ~ *taptaqir*; see 2.19.4.

Three cases are less clear: Tat. *čepči* ‘1. utterly raw; 2. round, complete, true; 3. in veterate, double-dyed, genuine’, Uigh. *tipteč* ‘complete peace’ and Uigh. *tüptekis* ‘very smooth, very even’. In all some type of contamination or other irregular phenomenon might be suspected; see 2.15.3 and 2.19.4, respectively. Lastly, Ott. *aphāzyr* ‘absolutely ready, ...’ arose probably from a combination of poor audibility of *h* and lack of deeper understanding of the mechanism of reduplication, see 2.13.3.

In general, then, it is the stem that is more progressive. A particularly well documented example of this is Ott. *dopdolu* ‘completely full’ < *topdolu* < *toptolu* (see 2.13.3). Uigh. *javjäšel* and *taptaqir* ~ *tapteqir* (above) are also convincing pieces of evidence. The only two exceptions, it seems, where it is clearly the reduplicated anlaut that undergoes a change while the base does not, are Karakalpak bases in *ǰa-* and Ott. *aphāzyr* ‘absolutely ready, ...’.

See also 3.1.10 for cases of reduplicated anlauts being used to intensify entirely different words, and 3.2.6 for conclusions that can be drawn from the examples collected here.

3.1.20 Shortened vowel

In the great majority of cases, the reduplicated vowel is short where the first vowel of the base is long. The traditional interpretation of *shortening*, however, is not necessarily correct, see 3.2.6.

Supporting this observation are 59 examples in nine languages: Dolg. *küök*, Gag. *dōru* (with *-p-* and *-s-*), *māvi*, *sā*, *sūk*, *tülü*, *ülen*, Khak. *čōn*, *hū*, *nā*, *tērpek*, *tōj*, Kirg. *dāna*, *ēn*, *kū*, *ōr*, *sō*, *sūk*, *tūra*, *žōn*, Ott. *hāzyr*, *tāze*, Tksh. *āšikar*, *āyr*, *dānyyk*, *dōru*, *jōun*, *lāživert*, *māvi*, *sā*, *sāde*, *sālam*, *sōuk*, *tāze* (with *-m-* and *-p-*), Trkm. *ājdyň*, *ājsat*, *āžy*, *dīri*, *dōly*, *gīň*, *gōk*, *gūry* (with *-p-* and *-s-*), *jāšyl*, *sāry*, *sūži*, Tuv. *čā*, and Yak. *kien*, *küöχ*, *sikej* (with *-n-* and *-p-*), *sīnes*, *suon*, *tiere*, *tūstaχ*, *ūllayas*, *yarχan*. See the “Structure” subsections in the respective sections in chapter 2.

Five further examples are not entirely clear: Ott. *māvi* ‘blue’, *sāfi* ‘pure’ and *sāry* ‘yellow’, and Trkm. *āk* ‘white’ and *čāl* ‘grey’. In the first three cases, the spelling suggesting a long vowel in the reduplicated anlaut is probably just a matter of orthography; in the third one of a misprint, and in the last one of a deficient notation; see 2.13.3 and 2.17.3, respectively.

Seven words go against the tendency. Six of them are Yakut and Dolgan, and all fall into two groups. One contains two examples where the original diphthong

has been preserved (Yak. *čuoǰcuoyur* ‘very motley’ and *tüöptüökün* ‘big cheater’), and the other one six cases where the originally short vowel has been lengthened (Ott. *tāstamām* ‘absolutely proper, ...’, Yak. *būsütün* ‘absolutely all’, Yak. *suočsoyoŋtoχ* ~ Dolg. *suotčogotok* ‘very lone(ly)’ and Yak. *ūnutary* ‘perversity, ...’). Apparently the only regular phenomenon that can be used to explain the North Siberian cases is the exclamative intonation, whereby the last vowel of the word is lengthened. However, high vowels are diphthongized with it, so a **büösbütün* would have to be expected instead of the attested *būsütün*. See 3.1.12 where an irregular phenomenon is proposed, and also 2.4.4 and 2.21.4 for wider commentary on these forms. The Ottoman form is unclear, see 2.13.3.

There seems to be no correlation between the shortening of the first vowel and the closing consonant used. All the words listed here have their reduplications closed with *m*, *p* and *s*, some even allowing more than one; that other consonants are not represented is most probably due to their general rarity.

The fact that long vowels are generally not reduplicated in full has consequences for the understanding and description of how Turkic reduplication operates, see 3.2.6.

See also 3.1.19 where other cases of reduplicated anlauts not matching the base are discussed.

3.1.21 Spirantized closing consonant

Dialectally, *p* functioning as a closing consonant can be spirantized to *f*, *v* and *w*.

The present work is basically only limited to selected literary languages (see 1.2.1), but four dialectal cases of a spirantized *p* can be mentioned: Az. dial. *gafgara*, *gufguru* and *safsary* ‘completely: black, dry and yellow, respectively’ (Ščerbak 1977: 120), Kar.E *afajsyz* ‘sudden(ly), unawares’ (see 2.6.3), Uigh. dial. *javjäšel* ‘very green’ (Malov 1954), and WYug. *sawsaryy* and *jawja^hsyl* ‘completely: yellow and green, respectively’ (Tenišev 1976b: 70). Note that in each case the same stem is also reduplicated in the literary variant and closed with a stop, which makes this phenomenon different from using double *pp* in the way of a closer (see 3.1.8).

Nothing seems to point to such spirantized reduplications possessing a greater degree of intensification than the standard occlusive variants. Phonetics, too, do not form a pattern or offer a ready explanation as it is only in Kar.E *afajsyz* that the **p* is intervocalic.

Essentially, three explanations present themselves: 1. ancient legacy, 2. independent evolution, and 3. Mongolic influence.

In the light of the data collected here, the first option seems to be the least likely because it would have to rely on three examples in Azeri and one slightly uncertain in Karaim as the only ones that have indisputably preserved what would then be the original spirant closing consonant, against the great wealth and variety of reduplications closed by occlusives.

For Azeri and Karaim then, it is perhaps the second explanation that is the most plausible.

In Mongolic, *C*-type reduplications are very often closed by *w* (see 3.4.3). This opens the possibility for the Uighur and Western Yugur forms to be explained differently than the Azeri and Karaim ones. While no final conclusions can be drawn at present, the option seems to be at least equally likely as that of independent evolution.

3.1.22 Verbs

Verbs can be reduplicated.

Examples are very few. Three cases are certain: Kar.E *apačmak* ‘to break open, to throw open’, Uzb. *qâp̄qârajmâq* ‘to turn completely black’ and also Uzb. *qipqizarmâq* ‘to redden intensively, to turn completely red, to flush strongly’. See 2.6.5 and 2.20.5, respectively. In both languages, the corresponding adjectives, Kar.E *ačyk* ‘open’, and Uzb. *qâra* ‘black’ and *qizil* ‘red’, also have reduplications.

Additionally, K.D. Harrison claims that in Tuvinian reduplication not only can be applied to verbs, but also that it is fully productive. This is a rather startling statement because it goes strictly against the picture of reduplication in the Turkic languages as a whole, and in the South Siberian branch in particular since it is there in fact that it is the least developed. No one seems to confirm Harrison’s opinion, and he himself appears to be only reiterating the same three examples in his works. See 2.18.2.

3.1.23 Vocalic anlaut

With one exception, reduplications of bases beginning with a vowel are closed by *p*. This is apparently the only phonetic rule for the choice of the closing consonant.

This restriction is included in Hatiboğlu’s first rule (1973: 34) and repeated in Müller’s “Regel C” (2004: 150). Of 186 words beginning with a vowel in the present collection, the rule holds for 185. The sole exception is Yak. *utary* ‘across, opposite’ which reduplicates rather extraordinarily to *ūnutary*. Neither the *n* nor the length of the reduplicated vowel are clear. It cannot be excluded that the word is not in fact a reduplication at all; see 2.21.3.

3.1.24 ‘White’

Bases meaning ‘white’ have been pronouncedly more productive than any other. In the North Siberian languages, this refers to *ürünj*, and in all the others to *ak*.

Tkc. *ak* is the only word whose reduplications have been reduplicated (see 3.1.9), the only clear example of a base being prepended to a reduplication (see 3.1.16), and the only base to accumulate, and do so repeatedly, more than two intensifying elements (see 3.1.13).

It is also one of the three reduplications that can have their closing consonant lengthened (3.1.8, sometimes with an additional change in the meaning, see 3.1.2) and one of the seven roots in which reduplication combines with diminutive (3.1.7).

The only method of intensification that can combine with reduplication, and has not been applied to *ak*, is nominal composition (3.1.18) of which, in fact, there is but one example.

Moreover, *ak* has given rise to probably the largest family of related reduplications in the present collection, which comprises thirteen forms in total in Khakas, Oïrot and Shor (3.1.11).

The achievements of NSib. *ürüŋ* appear amateurish in comparison, but it should be mentioned here that it is the only Yakut base whose reduplication can be closed by a double *pp*. This is slightly more telling in a language that appears to employ lengthening for intensification, see 3.1.12. In Dolgan, *ürüŋ* does not seem to be special in any way, but its collection is probably incomplete (see 2.4.1).

3.2 Structure

This section explores some interwoven aspects of the structure of Turkic reduplications. First, it gives an overview of the patterns of use of closing consonants in different languages (3.2.1), and then of the general state of reduplication in them (3.2.2). Next, it tests whether synchronic phonetic rules can accurately describe the choice of closing consonants with specific stems (3.2.3), and this failing, it proceeds to identify stems shared by different languages (3.2.4), and to inspect the use of closing consonants in those stems (3.2.5), and to conclude with remarks about the process of formation of reduplications (3.2.6). The whole closes with a summary of the most important findings and conclusions (3.2.7).

3.2.1 Closer

Overall, twelve different closers are attested (this includes *pp* as a separate one but does not include the spirantized versions; see 3.1.8 and 3.1.21). Their patterns of use differ considerably, see tab. 3.6.

The domination of *p* is undisputable in the majority of languages, but the distribution of alternative closers varies considerably. Two groupings are presented here, both serving as convenient abbreviations in the description of the alleged phonetic relation between the closer and the stem in the following subsections: a more subjective bipartite distinction between the *mprs*- and the *p*-languages, and a more objective one based on entropy.

See 4.1.3 for a summary of the problem of the choice of the closing consonant.

| | Language | č | j | k | m | n | ŋ | p | pp | r | s | š | t | Entropy |
|------------------------|----------------|---|---|---|----|---|---|-----|----|----|----|---|---|---------|
| <i>mprs</i> -languages | Azeri | 0 | 0 | 0 | 20 | 0 | 0 | 30 | 0 | 12 | 6 | 1 | 1 | 0.565 |
| | Eastern Karaim | 0 | 0 | 0 | 12 | 0 | 0 | 22 | 2 | 4 | 9 | 0 | 0 | 0.508 |
| | Gagauz | 0 | 0 | 0 | 6 | 0 | 0 | 27 | 0 | 1 | 9 | 0 | 0 | 0.409 |
| | Ottoman | 0 | 0 | 0 | 13 | 0 | 0 | 24 | 1 | 5 | 14 | 0 | 0 | 0.535 |
| | Turkish | 0 | 0 | 0 | 29 | 0 | 0 | 101 | 0 | 8 | 38 | 1 | 0 | 0.465 |
| | Turkmen | 0 | 0 | 0 | 2 | 0 | 0 | 39 | 0 | 0 | 13 | 0 | 0 | 0.286 |
| | Uzbek | 0 | 0 | 0 | 6 | 0 | 0 | 36 | 1 | 0 | 2 | 0 | 1 | 0.275 |
| | Yakut | 1 | 1 | 1 | 0 | 3 | 1 | 88 | 1 | 1 | 8 | 0 | 1 | 0.295 |
| <i>p</i> -languages | Bashkir | 0 | 0 | 0 | 2 | 0 | 0 | 90 | 1 | 0 | 0 | 0 | 0 | 0.043 |
| | Dolgan | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0.000 |
| | Karakalpak | 0 | 0 | 0 | 0 | 0 | 0 | 59 | 1 | 0 | 0 | 0 | 0 | 0.000 |
| | Kazakh | 0 | 0 | 0 | 0 | 0 | 0 | 107 | 1 | 0 | 0 | 0 | 0 | 0.000 |
| | Khakas | 0 | 0 | 0 | 0 | 0 | 0 | 33 | 0 | 0 | 0 | 0 | 0 | 0.000 |
| | Kirghiz | 0 | 0 | 0 | 0 | 0 | 0 | 91 | 1 | 0 | 0 | 0 | 0 | 0.000 |
| | Kumyk | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 1 | 0 | 1 | 0 | 0 | 0.070 |
| | Oirot | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 1 | 0 | 0 | 0 | 0 | 0.000 |
| | Shor | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0.000 |
| | Tatar | 0 | 0 | 0 | 3 | 0 | 0 | 55 | 1 | 0 | 0 | 0 | 0 | 0.084 |
| | Tuvinian | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0.000 |
| | Uighur | 0 | 0 | 0 | 0 | 0 | 0 | 33 | 1 | 0 | 0 | 0 | 0 | 0.000 |
| | Western Karaim | 0 | 0 | 0 | 3 | 0 | 0 | 20 | 1 | 0 | 0 | 0 | 0 | 0.157 |

Table 3.6: Number of examples for different closing consonants in different languages. *Entropy* stands for ‘normalized Shannon entropy’ (calculated for *p* and *pp* treated as the same closer), see the main text.

mprs- vs. *p*-languages

When discussing the phonetic relation between the closing consonant and the first and second consonant of the stem (3.1.4–3.1.6), it has proven convenient to introduce two abbreviations: “*mprs*-”, and “*p*-languages”. The principle of division is this: a language in which at least three of *m*, *p*, *r* and *s* are attested as a closing consonant, is an *mprs*-language; a language which is not an *mprs*-language, is a *p*-language. The first employ different closing consonants more evenly, the other use *p* almost exclusively (see tab. 3.6, ignoring entropy for the moment).

This division is secondary. It does not implement any general theoretical distinction; rather, it simply captures the difference between two groups which have apparently developed different strategies for dealing with homolocality between the closing consonant and the first consonants of the stem.

The two groups are not homogeneous. Yakut is clearly different than the other *mprs*-languages; Bashkir, Tatar and Western Karaim are visibly more of borderline cases than clear *p*-languages, and so are Uzbek and, in a way, Turkmen in the *mprs* group. It appears that, if anything, this grouping better reflects geography and zones of influence than genealogy, because it hardly coincides with the classical classification.

It is not a trivial outcome, that a geographical-cultural classification is more convenient in the description of the phonetic relation between two elements of reduplications, than a genealogical one. But far-fetched conclusions should not be drawn from this fact. The *mprs*-languages comprise basically the entire Oghuz group, Eastern Karaim (under heavy influence of Ottoman), Uzbek (genetically mixed, primarily Oghuz-Karakhanid), and Yakut (broke contact very early on).

It seems then that the initial stock of reduplications should have contained seeds of both types of evolution; the Oghuz and the Yakuts have developed one, the other languages the other. Except for the isolated Yakut, all were exposed to the influence of the neighbouring languages which reinforced their chosen path or, less often, drew them away towards the opposing group. This is a more moderate and appears to be a more plausible conjecture than convergent evolution from varied starting points. Nevertheless, it must be approached with reserve as yet because the similarities between Yakut and the remaining *mprs*-languages are not particularly striking. See 3.2.7 for additional support for this inference, and 3.4 for a continuation of these considerations.

Entropy

A more objective way to describe how different closing consonants are employed in a language, is entropy. Here, normalized Shannon entropy is always used.⁴³ Map 3.2 shows the geographical distribution; the exact values are given in tab. 3.6.

Apparently, entropy divides the Turkic languages into four groups: 1. zero ($H_n = 0$), 2. mid-low ($0 < H_n \leq 0.2$), 3. mid-high ($0.2 < H_n \leq 0.4$), and 4. high ($H_n > 0.4$); see fig. 3.3. This corresponds quite well to a more intuitive distinction between the *mprs*- and *p*-languages (see above). Therefore, it also corresponds reasonably well to geography, and not very well to genealogy. This fact provides additional support for the supposition made above about the history of evolution of reduplication. Entropy-based grouping will also prove convenient in the discussion of phonetic rules for choosing the closing consonant below (3.2.3).

⁴³ The term *entropy* may refer to a number of related concepts. In the present work, it is employed in its classical information theoretical interpretation, which is a measure of uncertainty in a random variable. That is to say, if a word is chosen at random, and a guess ventured what closing consonant its reduplication will have, then entropy quantifies the uncertainty of this guess. Here, normalized entropy is used, so that 0 denotes a complete lack of uncertainty (the only possible closing consonant is $p(p)$, and the guess must be correct), and 1 denotes maximal uncertainty (different closing consonants are possible, and all are used with equal frequencies, i.e. equally probable).

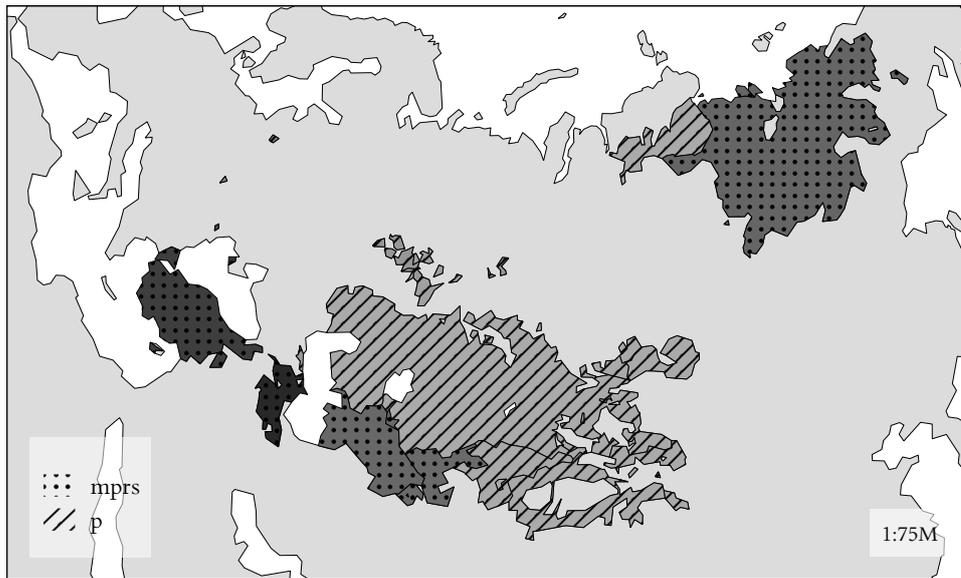


Figure 3.2: Geographical distribution of normalized Shannon entropy. Darker regions correspond to higher values; the exact numbers are in tab. 3.6.

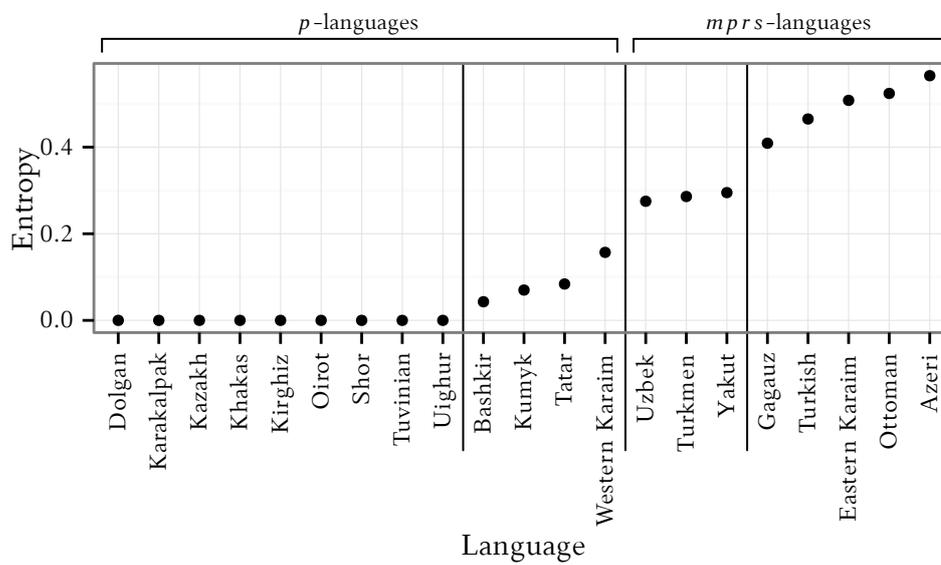


Figure 3.3: Normalized Shannon entropy of closing consonants in different languages; the exact numbers are in tab. 3.6. Vertical lines show the proposed grouping (see the main text).

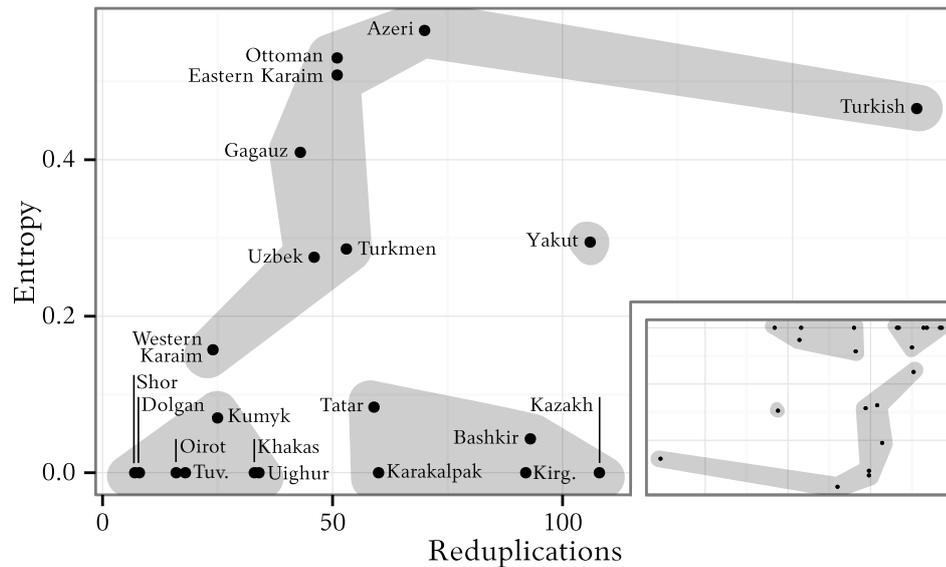


Figure 3.4: Number of reduplications in different languages against normalized Shannon entropy of closing consonants. The miniature plot is a mirror reflection. See the main text, where it is also explained why Turkish is grouped together with the other Oghuz languages rather than as a separate group like Yakut.

Another point of interest in map 3.2 is Turkish. In general, entropy appears to be rising as one advances westwards from the Altai Mountains, but Turkish is an exception. Not much can be concluded based on entropy alone but it should be noted that in this regard, Turkish lies almost precisely in the middle between Azeri and Turkmen, and that the entropy for Ottoman – although these data span a very long period and are surely incomplete (see 2.13) – is higher, and closer to Azeri.

3.2.2 General characteristics

It seems that the defining parameters of the stage of evolution of reduplication in different languages are the number of reduplications and the entropy of closing consonants. They are visualized in fig. 3.4; see also 3.4.3 on how the Mongolic and Tungusic languages relate.

Viewed through the lens of the number of reduplications and entropy of closing consonants, all languages appear to fall into four groups (greyed in fig. 3.4; see also fig. 3.6):

- “Karakhanid” / “South Siberian”, where reduplications are few, and all or almost all are closed by p (Dolgan, Khakas, Kumyk, Oiro, Shor, Tuvinian, and Uighur),

- “Kipchak”, where reduplications are many, and all or almost all are closed by $p(p)$ (Bashkir, Karakalpak, Kazakh, Kirghiz, and Tatar),
- “Oghuz”, where reduplications are usually many, and p is the dominating consonant but m , r , and s are also relatively common (Azeri, Gagauz, Karaim, Ottoman, Turkish, Turkmen, and Uzbek), and
- “Yakut”, where reduplications are many, and p is the dominating consonant, but many other consonants are also used as closers (Yakut).

First, a remark about the location of Turkish should be made. It is a very distant outlier in the “Oghuz” group, due to the unusually high number of reduplications that are attested in it. Next in order, Kazakh and Yakut, have less numerous collections by about a third, and this is the only such big difference across all the languages (see also fig. 3.10). In reality, however, probably just about a half of those reduplications are in common use throughout its territory (see 2.16.4). This would place Turkish far to the left and probably a little higher in the figure, and rather close to Azeri.

Now, the four groups distinguished here to correspond quite well to the genealogical classification of the Turkic languages, and it appears that they can be interpreted as different paths of evolution of reduplication.

If so, fig. 3.4 can also be seen as a diachronic-like scheme.⁴⁴ It begins foreseeably at (0, 0), from whence along the diagonal extends an (imaginary) axis representing the stage of development of reduplication. Only two directions are available: horizontal, where new reduplications are added, and diagonal, where new reduplications but also new closing consonants arise. (A strictly vertical advance is not possible as new closers cannot be added when no reduplications exist.) The “Karakhanid”/“South Siberian” and “Kipchak” groups went essentially the first way. The “Oghuz” and Yakut went diagonally, but at different angles.

Another interesting property of fig. 3.4 is its relation with geography, as can be seen in the miniature plot in the corner. It is a mirror reflection of the main plot; this little legerdemain breaks the arbitrary Cartesian convention and locates the point (0, 0) in the upper right corner, in order to bring the plot closer to the arbitrary Ptolemy-inspired mediaeval convention of orienting maps with north at the top.

Thus, the “Karakhanid” / “South Siberian” group is in the northwest, the “Kipchak” group west from it, the “Oghuz” group southwest from it, and it is only at Yakut, the only Turkic language that advanced north or west from the Altai homeland, that this metaphor must stumble.

The general outline is preserved, but the placement of specific languages is incorrect where the diachronic-like interpretation interferes, e.g. the fact that Kumyk appears to be east from Tatar results from that having almost the same entropy, Kumyk has a considerably smaller number of reduplications attested in it, which locates it at an

⁴⁴ See fig. 3.9 for a similar scheme of semantics.

earlier stage of development, i.e. in this case at almost the same latitude, but at a lower longitude.

What makes this property interesting, however, is why it actually generally holds. A simple plot of the number of reduplications against entropy of closing consonants can only be expected to show similarity to geographic distribution if both the evolution of reduplication, and territorial advance were generally linear, and coincided.

Yakut aside, the Turkic peoples mainly advanced on a broad front straight west. The languages in the vanguard, it appears, must have developed reduplication more vigorously than those in the centre, and did so along two main paths. Both involved adding new reduplications (hence the advance “westwards” in fig. 3.4), but only the “Oghuz” group also added new closing consonants (advance “southwards”).

Also signs of secondary, contact-induced developments can be found in fig. 3.4. Uighur and Uzbek, for example, are far apart despite close genealogical affinity. But both are closer to languages that they are in reality neighbouring with. Uighur does not have closing consonants other than $p(p)$, and this locates it far “north”, and close to the other p -languages, while for Uzbek the opposite is true. Likewise, Eastern and Western Karaim are very far from one another, because one was under heavy Ottoman influence, while the other was surrounded by a non-Turkic element, and did not greatly develop reduplication.

See 3.4 for a more detailed discussion of the history of Turkic reduplication.

3.2.3 Synchronic phonetic rules

A greater part of the effort invested in the study of Turkic reduplications so far has concentrated on formulating synchronic phonetic rules for choosing the closing consonant (see 1.1.2). This attitude is quite frail from the point of view of methodology (see 3.4.1). Nonetheless, let us now inspect the examples from different languages to find what degree of accuracy can be expected from such phonetic tendencies. See 3.2.5 for an attempt at identifying patterns across languages.

Only bases beginning with a consonant are considered here. Those with a vocalic anlaut have their reduplications closed with p in all but one case (see 3.1.23). This is truly a phonetic tendency, a very strong one, and the only one so strong. Yet, the actual phonetic motivation behind it remains mysterious to me.

Languages are discussed in groups based on their entropy. The concept has been introduced in more detail in 3.2.1; the general idea is that the lower the entropy, the more the reduplications are dominated by one closing consonant (that consonant is always p). The Turkic languages fall into four distinct groups, presented here in the order of increasing entropy.

Languages with entropy of 0 (group one) can be put aside immediately because p and pp are the only sounds that they employ to close the reduplicated anlauts. Here, the

rule accounts for all the available examples and it is always true. It is only rather questionable whether it can be considered phonetic in nature.

Group two ($H_n \leq 0.2$) contains four languages: Bashkir and Tatar, Kumyk, and Western Karaim.

In Bashkir (see 2.3.4), there are only two reduplications closed by *m* against 90 closed by *p(p)*. They are *jäšel* ‘green’ and *kük* ‘blue’, but the first can alternately be closed by *p*. One possible formulation of a phonetic tendency is this: all Bashkir reduplications are closed by *p* except for those where the base is two-syllable long and begins with *jäš*, when it can also be closed by *m*, and those where the base is front and has *k* for both C_1 and C_2 , when only *m* is possible. The additional restrictions, such as being two-syllable long or frontness, were forced by the words *jäš* ‘young’ (plus 15 other beginning in *j-*) and *kak* ‘1. naked, bare; 2. very slim, skinny’ (plus 10 other in *k-*), whose reduplications are all only closed by *p*. Such a rule is technically correct but it lacks an actual phonetic motivation, is clearly overfitted and without general value, or in short, absurd. It is simply pointless to attempt to describe Bashkir reduplications in terms of the phonetic shape of the stem.

In Tatar (2.15.4), the situation is very similar. One more word can have a reduplication closed by *m* (*tügäräk* ‘round’) – or, alternately, by *p*. The rule: *m* is limited to reduplications of bases which are two syllable-long and begin with *jäš*, monosyllabic bases with *k* for both C_1 and C_2 and, as an alternative closing consonant, for front labial bases beginning with *t-*. The formulation for *k* must have been changed compared to Bashkir because in Tatar, the front word *käkre* exists and has a reduplication closed by *p*. This fact can serve very well to demonstrate the nonsense and extreme overfitting of such a description.

In Kumyk (2.11.4), there is one reduplication in *s* (*bütün* ‘whole’) against 24 in *p(p)*. The rule might be such: all reduplications are closed by *p*, except for when the base begins with *b-* and has *t* for C_2 . Again, an actual phonetic motivation for the distribution is completely missing.

In Western Karaim (see 2.6.4), there are few examples in general and any formulation of phonetic tendencies could only be fragile at best.

To sum up, it is technically possible to formulate phonetic rules which describe the choice of the closing consonant in the languages in group two, but they are clearly overfitted, not in fact phonetically motivated, and overall quite useless.

Group three ($0.2 < H_n \leq 0.4$) contains three languages: Turkmen, Uzbek and Yakut. Unlike in the previous two groups, here phonetics does seem to have a limited degree of influence on the choice of the closing consonant.

As far as closers other than *p* go, Turkmen (see 2.17.4) clearly prefers *s* over *m*. It is used for all the three bases beginning with *b-*, almost a half of those in *d-*, and

approximately a third of those in *g-* and *t-*. At the same time, *m* is only attested in two examples, *dik* ‘steep’ and *gōk* ‘1. blue; 2. green’.

The case of Uzbek (2.20.4) is more complicated, possibly due to its genealogically mixed history. Formulation of rules is possible but, apart from being lengthy and complex, they will even surpass those offered for Bashkir and Tatar above in lack of actual phonetic justification.

Finally, Yakut (2.21.4) has closing consonants galore. Eight are attested beside *p(p)*, but six of them seem to only have one example each – which hardly suffices for phonetic conclusions. The other two are *n* and *s*. For *n* no pattern emerges but *s* is used mainly for bases beginning with *b-*, and no such base has a reduplication closed by *p-*. Bases beginning with *p-* do not seem to be attested at all. Moreover, the only example closed by *r* is *bosχo* i.a. ‘straight, upright’ – a word which begins with *b-* and has *s* for *C*₂, like no other Yakut stem in *b-*.

Concluding, phonetics is not without importance for the choice of the closing consonant in the languages in group three, but the relation is neither direct nor exception-free. The significant part of the base seems to be effectively limited to its initial consonant, and only once in Yakut does also *C*₂ appear to play a role.

Lastly, the five languages from group four ($H_n > 0.4$): Azeri, Eastern Karaim, Gagauz, Ottoman, and Turkish.

For Azeri (see 2.2.4), as many as 20 and 12 examples are available for *m* and *r*, respectively, but rather than remove uncertainty, this fact actually only increases it. I could identify just one tendency, that reduplications of words beginning with *b* are not closed by *p*, but even this is not without exception (*balaža* ‘small, tiny’, can be closed with both *m* and *p*). Apart from that, different consonants for *C*₁ and *C*₂, different vowels, different number of syllables, all seem to be distributed approximately evenly between *m*, *p*, *r* and *s*.

In Eastern Karaim (2.6.4), the number of examples is lower than in Azeri, but they are very similar and allow for the same, somewhat uncertain conclusion.

The case of Gagauz (2.5.4) is not clear. It seems as if hints of patterns were visible but the number of examples in each category is too low to allow conclusions. In particular, *p* is probably avoided if the base begins with *b-* or *p-* (base on three examples).

As for Ottoman (2.13.4), conclusions are also uncertain because the available data is most probably not representative for the language as a whole (2.13.4). But even despite this, it is clear that there is generally no obvious phonetic motivation behind the choice of different consonants. Similarly to the languages above, *p* is undesirable for bases beginning with *b-*.

Finally, Turkish has already been extensively discussed in the previous literature (see 1.1.2, 2.16.1, and appendix A). Even a relatively complex, visibly overfitted, and hardly phonetically motivated set (Müller 2004) did not account for more than 79% of examples.

To sum up, phonetic tendencies in group four, instead of emerging with greater clarity due to the larger number of examples, are in fact being obscured by it. The role of phonetics of the base appears to be limited to the avoidance of *p* when the base begins with *b-*. Just as the distribution of different closing consonants is here the most even, so apparently is the distribution of phonetic features of the bases across the different closing consonants.

The overall picture is thus very simple and hardly novel: in the *mprs*-languages *p* is avoided for bases beginning with *b-*, while in the *p*-languages it is used indiscriminately.

One slightly more revealing observation is that phonetic tendencies, if any can be really talked about, seem to emerge slightly more clearly in the middle of the field, in group three, than in the more uniform groups one and two, or in the quite evenly distributing group four.

See 3.2.5 for a continuation of these considerations, which remains synchronic but operates on a hypothetical set of stems from the times when the *mprs*-languages were part of one community.

3.2.4 Common stems

Reduplication has existed in the Turkic languages since at least the 11th century (see 2.1). It is to be expected, then, that a proportion of modern examples has been inherited from the earlier stages, and therefore, that the collections of reduplications in languages which belong to one genealogical group will show a degree of similarity.

It is not assumed here that if a stem is common to two languages, then its reduplications in these two languages must necessarily be a common inheritance (see 3.4.2 for an overview of arguments behind this approach). Nevertheless, a synchronic identification of common stems will be able to shed some light on the beginnings and diachrony of reduplication.

Only inflectional stems are compared here, so that e.g. Az. *širin* ‘sweet’ and *širinlik* ‘sweetness’ are considered two separate units. This is to help minimize the number of false positives, even if probably at the expense of false negatives.

Note that the term *inflectional stem* is used etymologically here, which makes it slightly different from the term *base* as used in chapter 2 and elsewhere. Doublets such as Az. *jalgyz* : *jalnyz* ‘lone(ly)’ are considered two bases but one (etymological) stem. These are sporadic cases.

Taking into account the limitations of the sources, and to prevent an excessive reduction of the results, the identification will only rely on languages with the largest dictionaries. The final sets will be checked against the omitted languages, but without consequences.

Where applicable, closing consonants are also taken into consideration to help separate those stems whose reduplications in different languages are more likely to be independent innovations.

Stems are referred to by the literary or dialectal Turkish shape where available, and by some other where not. Meanings, however, are given according to the currently discussed languages, not always Turkish.

The numbers of common reduplicated stems between specific languages are given in tab. 3.7. The analogous collective comparison between the five groups is in tab. 3.10.

Karakhanid

Only two Karakhanid languages are discussed in the present work, Uighur and Uzbek (see 2.19 and 2.20). Their largest sources contain ca. 33 000 and 40 000 entries, respectively, allowing them to be compared without fear of extensive and unrealistic reduction of the number of results.

Seventeen stems are present in the two languages simultaneously: *ačyk* ‘open, clear’, *ak* ‘white’, *baravar* ‘same’, *boš* ‘empty’, *dōru* ‘straight, true’, *gök* ‘blue’, *jaruk* ‘light, bright’, *jeni* ‘new’, *ješil* ‘green’, *karanly* ‘dark’, *kara* ‘black’, *kyzyl* ‘red’, *sary* ‘yellow’, *takyr* ‘smooth, bare’, *tāze* ‘clean’, *tekiz* ‘smooth’, and *tyn* ‘still, silent’.

Their closing consonants are not entirely clear, especially in the case of *tyn*. In Uighur, it seems that only *p* and *pp* are used as closers. In Uzbek, also *m*, *s*, and *t* are attested, and in particular, of the seventeen stems listed here, *boš*, *jašil*, and *kök* have in Uzbek reduplications closed by both *m* and *p*, but *tin* apparently only reduplicates to *timtin*, and not to **tiptin*.

The first three can perhaps be blamed on the Oghuz part of Uzbek’s past, but *tin* is more ambiguous. As far as reduplications go, the word seems to have no relatives in Uzbek. The Uighur side of the family is larger as it contains *tin* i.a. ‘silence, peace’, *tinč* ‘quiet, still peaceful’, *tiniq* ‘clear, transparent’, and possibly also *teč* ‘quiet, peaceful’. Apart from these, related forms also appear in Bashkir (*tyn*, *tynys*, and *tymyk*), Karakalpak (*tynyk*), Kazakh (*tynyk*, and *tynyš*), and Tatar (*tyn*, *tynyč*, and *tymyzyk*), and possibly also Ottoman (**diņsüz*, see 2.13.3). All of them have reduplications closed by *p*. Uzbek, just barely an *mprs*-language, has the only representative of this relatively large family, that has a reduplication not closed by *p*.

Leaving Uzb. *tin* aside, the reduplications of all the sixteen words are closed by *p*. This is not very telling since, apart from Ottoman, closers other than *p* are in general very rare in these languages (see 3.2.1 for details). Nonetheless, they are also quite uncommon in Uzbek, and yet *tin* has a reduplication closed by *m* without any obvious reason.

Thus, sixteen of the seventeen stems can be considered common to Uighur and Uzbek with relative confidence. The unusual closing consonant in Uzb. *timtin* ‘completely still, ...’ requires further study.

Kipchak

The Kipchak group is the largest, and it is also well documented (see the “Sources” subsections in the respective sections in chapter 2). Bashkir, Kirghiz, Karakalpak, and Tatar all have dictionaries with 30 000 or more entries. The largest Kazakh dictionary used here is smaller (ca. 21 000 entries) but since the Kazakh collection of reduplications is still one of the richest in the present work, it will also be included in the comparison. Omitted will only be Kumyk (ca. 13 000 entries), and the three Karaim languages (ca. 17 400 entries).

Only ten stems are present in all Bashkir, Karakalpak, Kazakh, Kirghiz, and Tatar simultaneously: *ačyk* ‘open, clear’, *ak* ‘white’, *jaš* ‘young’, *jenil* ‘light’, *ješil* ‘green’, *jumušak* ‘soft’, *kara* ‘black’, *karanly* ‘dark’, *kyzyl* ‘red’, and *sary* ‘yellow’.

This is surprisingly little. Taken pairwise, the average number of common stems between these languages is as high as 28.9, with the standard deviation of 7.14, and no obvious outliers.

| | Bashkir | Karakalpak | Kazakh | Kirghiz | Tatar |
|------------|---------|------------|--------|---------|-------|
| Bashkir | 89 | | | | |
| Karakalpak | 28 | 59 | | | |
| Kazakh | 29 | 37 | 108 | | |
| Kirghiz | 23 | 37 | 35 | 90 | |
| Tatar | 37 | 22 | 21 | 20 | 56 |

Table 3.8: Number of common reduplicated stems between the Kipchak languages with large sources, taken pairwise. The numbers on the diagonal are effectively the number of stems in the given language. See tab. 3.7 for all the modern languages discussed in the present work.

It can be seen from tab. 3.8, that the number of stems common to any two languages does not correlate with how many reduplications are attested in them. It is not Karakalpak and Kirghiz, the two languages with visibly less numerous collections, that limit the final set.

Overall, 37 stems are shared by exactly two languages, 25 by three, 13 by four, and 10 by all five. The descent is gradual, without any sudden drops. From among stems common to four languages, four are missing from Kirghiz and Kazakh each, three from Tatar, and two from Bashkir.

Clearly, there is not one culprit language that decimates the set of common stems. Rather, a small continuum can be observed, where neighbouring languages share many stems with one another, but less with the geographically more distant ones. This suggests that secondary contacts might have had a greater impact on the formation of reduplications than inheritance. See 3.4.5 for a continuation of this thought.

In all the five languages, *p* visibly dominates as the closing consonant (see 3.2.1), but two reduplications closed by *m* can still be found: Bshk. *jämjäšel* ~ *jäpjäšel* ‘very (light) green’, and Tat. *jämjäšel* ‘very green’. It seems quite possible, that the change of the closing consonant in the reduplications of this stem is an areal feature that these two languages have in common with Western Karaim and maybe Chuvash (see 3.2.5). Unlike the unusual *m* in Uzb. *timtin* (above), this deviation will not be considered so serious as to exclude *ješil* from the set of common stems.

Three languages have been omitted because of the modesty of their sources: Kumyk, and North- and South-Western Karaim. Out of the ten common stems, two are missing from the Kumyk data (*jaš* ‘young’ and *jumušak* ‘soft’), and as many as five from both Karaim languages (*ačyk* ‘open’, *jaš* ‘young’, *jenil* ‘light’, *jumušak* ‘soft’, and *karanly* ‘dark’). It is not possible at the moment to determine the cause of this absence. The sources might be incomplete, Karaim might have lost some of its reduplications being surrounded by a non-Turkic element (merely 24 examples are attested in both North- and South-Western Karaim together), but also the five words could possibly be a common Central Asian innovation, from whence both Karaim and Kumyk are detached to various degrees (note that the two words missing from Kumyk are included in the five missing from Karaim).

Overall, the set of ten stems common to Bashkir, Karakalpak, Kazakh, Kirghiz, and Tatar will be considered sufficiently certain and representative for the entire Kipchak group.

North Siberian

The approach adopted in this subsection fails to produce any results for the North Siberian group, as it only consists of two languages, Dolgan with the sources containing ca. 8900 entries, and Yakut with a dictionary of ca. 40 000.

Just eight reduplications are attested in Dolgan, and six of them have their Yakut equivalents: *gök* ‘blue’, *kara* ‘black’, *kyra* ‘small’, *kyzyl* ‘red’, *tekerlek* ‘round’, and *ürünj* ‘white’. The two Dolgan stems which apparently do not have counterpart reduplications in Yakut are *karaŋa* ‘dark’, and *kytarkaj* ‘red’.

It is not clear what this implies. The Dolgan data might be, and in fact probably are, incomplete, but it is anyone’s guess how many *reduplications* are not attested. Dolgan separated from Yakut not later than in the beginning of the 17th century (Stachowski M. 1996). In theory, the current richness of Yakut reduplications could have arisen through a more or less explosive evolution which occurred after that point, but it is also possible that the evolution was in fact fairly gradual, and only its fruits were abandoned by Dolgans due to the Tungusic substrate (see 2.4).

Overall, no certain conclusions can be drawn from the North Siberian group. It will be included in further considerations on special rights.

Oghuz

Identification of common stems in the Oghuz languages is made more difficult by their use of different closing consonants. The two questions are entangled, and dealt with separately in 3.2.5. Here only the results are presented.

It is not certain how the differences in the closing consonants should be interpreted. A somewhat arbitrary solution will have to be adopted for our purposes. The clear group of stems which are common to the three languages with large sources (Azeri, Turkish, and Turkmen), and have the same closing consonant in all three, contains eight examples: *ajdyn* ‘bright’ (closed by *p*), *ak* ‘white’ (*p*), *bütün* ‘all’ (*s*), *dik* ‘steep’ (*m*), *gök* ‘blue’ (*m*), *kara* ‘black’ (*p*), *kyrmyzy* ‘red’ (*p*), and *sary* ‘yellow’ (*p*).

Further five stems have one closer in Azeri, a different one in Turkmen, and both as alternatives in Turkish: *beter* ‘worse’ (*š, s*), *dolu* ‘full’ (*p, s*), *düz* ‘smooth’ (*m, p*), *ješil* ‘green’ (*m, p*), and *temiz* ‘clean’ (*r, p*). They will be considered common Oghuz stems.

Six more stems are common to the three languages, but the closing consonants in their reduplications are not uniform: *dōru* ‘straight’, *diri* ‘alive’, *duru* ‘clear’, *jeni* ‘new’, *kuru* ‘dry’, and *tāze* ‘fresh’ (see tab. 3.12 for the closing consonants). These will not be considered common Oghuz stems. It might be noted, however, that a half of them (*dōru*, *jeni*, and *tāze*) can be also found among stems common to the Karakhanid languages, see tab. 3.9.

The only Oghuz language discussed in the present work and omitted from this comparison, is Gagauz. From the first group, reduplications of *ajdyn* (albeit *apajdynnyk* is attested), *bütün*, *dik*, and *gök* appear to be missing from it; from the second – of *beter* and *temiz*. This is almost a half of the thirteen stems which are here considered common. All are rather basic and quite commonly reduplicated words. It seems more likely than not, that this absence results merely from the lack of a larger source.

South Siberian

The South Siberian group is represented by four languages in the present work: Khakas, the largest source for which contains ca. 14 000 entries, Oirot (ca. 13 000), Shor (ca. 4000), and Tuvinian (ca. 22 000). In other words, neither is in fact very well suited for a comparison.

Five stems are common to Khakas, Oirot, and Tuvinian: *denk* ‘same’, *kara* ‘black’, *kyzyl* ‘red’, *sary* ‘yellow’, and *uzun* ‘long’.

Only the three colour names are also attested for Shor. The lack of *ak* ‘white’ among the common stems is surprising. It was eliminated by Khakas, which is the only non-North-Siberian language discussed in the present work that does not have a simple reduplication of it. It has the form *appagas*, but its history is not clear, although it is clear that it belongs to a larger family of similar shapes in Khakas, Oirot, and Shor, and is not necessarily a Khakas own innovation.

| Stem | Meaning | Kara kh khanid | Kipchak | Oghuz | South Siberian | Yakut | Count |
|----------------|----------|------------------------------|---------|-------|-------------------|-------|-------|
| <i>ačyk</i> | open | + | + | | | | 2 |
| <i>ajdyn</i> | bright | | | + | | | 1 |
| <i>ak</i> | white | + | + | + | | | 3 |
| <i>baravar</i> | same | + | | | | | 1 |
| <i>beter</i> | worse | | | + | | | 1 |
| <i>boš</i> | empty | + | | | | | 1 |
| <i>bütün</i> | all | | | + | | + | 2 |
| <i>denk</i> | same | | | | + | | 1 |
| <i>dik</i> | steep | | | + | | | 1 |
| <i>dolu</i> | full | | | + | | | 1 |
| <i>dōru</i> | straight | + | | | | | 1 |
| <i>düz</i> | smooth | | | + | | | 1 |
| <i>gök</i> | blue | + | | + | | + | 3 |
| <i>jaruk</i> | bright | + | | | | | 1 |
| <i>jaš</i> | young | | + | | | | 1 |
| <i>jeni</i> | new | + | | | | | 1 |
| <i>jenil</i> | light | | + | | | | 1 |
| <i>ješil</i> | green | + | + | + | | | 3 |
| <i>jumušak</i> | soft | | + | | | | 1 |
| <i>kara</i> | black | + | + | + | + | + | 5 |
| <i>karanly</i> | dark | + | + | | | | 2 |
| <i>kyrmyzy</i> | red | | | + | | | 1 |
| <i>kyzyl</i> | red | + | + | | + | + | 4 |
| <i>sary</i> | yellow | + | + | + | + | | 4 |
| <i>takyr</i> | smooth | + | | | | | 1 |
| <i>tāze</i> | fresh | + | | | | | 1 |
| <i>tekiz</i> | smooth | + | | | | | 1 |
| <i>temiz</i> | clean | | | + | | | 1 |
| <i>uzun</i> | long | | | | + | + | 2 |

Table 3.9: Common stems across genealogical groups. For North Siberian, Yakut has been used in place of the set of common stems due to the scarcity of Dolgan data, see subsection “North Siberian” above.

General

Stems common to the entire Turkic family can be viewed from at least three angles. First, the above considerations will be continued to identify the specific stems that are common to all groups. Then, all stems will be inspected to find how many are shared by different pairs of groups, i.e. how strong the connections between the different genealogical groups are. Finally, the two points of view will be combined to yield a kind of a map of distances between the various languages.

Tab. 3.9 collects and contrasts the sets of common stems that have been identified above. Only one stem appears in all five, *kara* ‘black’. In fact, it can be observed that two subsets are almost identical: that of stems present in three or more groups, and that of names of colours. The only element differentiating them is *kyrmyzy* ‘red’, which is only common to Oghuz, and opposed to *kyzyl* id., common to all the other groups but not to all the Oghuz languages. This induces another observation, that there are visibly less meanings in the table than there are stems. Semantics have clearly played an important role in the evolution of Turkic reduplications; see 3.3.4 for a more detailed discussion.

One conclusion that it is certainly too early to draw from tab. 3.9, is that the entire phenomenon of reduplication was initiated by *kara*. This table is based as much on which reduplications are attested as on which ones are not, and neither of these can actually prove or disprove common descent. Rather, it should be understood to be a general picture and a suggestion where a more detailed research might want to begin.

That said, the reduplications of six colour names, *ak* ‘white’, *gök* ‘blue’, *ješil* ‘green’, *kara* ‘black’, *kyzyl* ‘red’, and *sary* ‘yellow’ can, in all probability, be assumed to have existed at the very earliest stages of reduplication, and considered part of the common inheritance in all those Turkic languages in which they are still used. It is only less clear whether the relation between them is of an areal or genealogical nature. Since, however, this question probably cannot be answered before the entire Altaic debate is settled, it will not be given much attention in the present work.

A different kind of observations can be made by looking at the number of common stems between pairs of genealogical groups. The exact numbers are given in tab. 3.10, and visualized in fig. 3.5.

The most visible property in fig. 3.5, is perhaps the proportion between the total number of stems with attested reduplications, and the number of stems shared with other groups. Common stems constitute visibly the smallest part in the North Siberian.⁴⁵ There are more in the Kipchak, slightly more in the Oghuz, and then much more in the Karakhanid and in the South Siberian. In the last two, in fact, the sum of all sets shared

⁴⁵ Effectively, the data refer to Yakut alone as the Dolgan input to North Siberian reduplications is negligible.

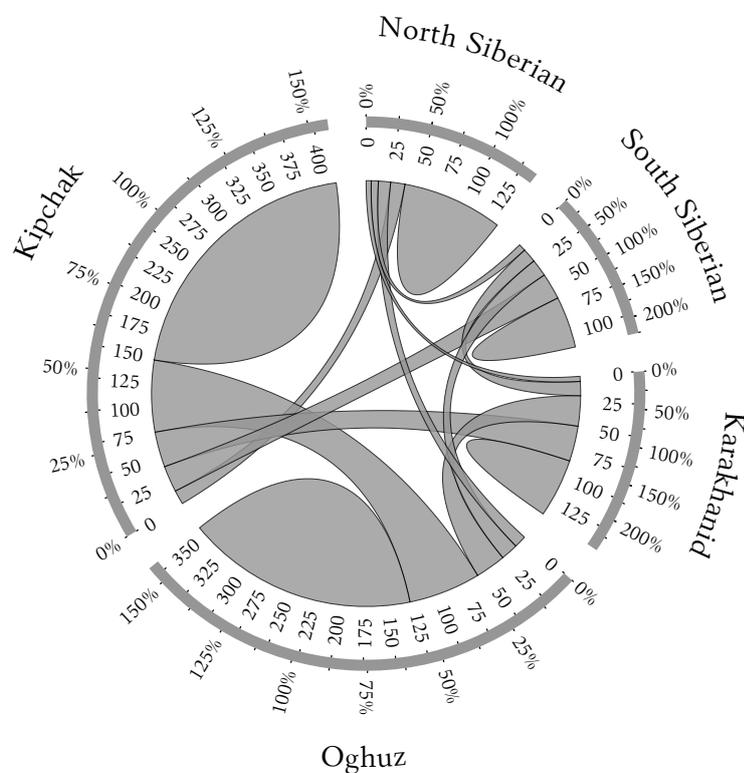


Figure 3.5: Visualization of the number of common reduplicated stems between groups, taken pairwise. See tab. 3.10 for the exact numbers, and fig. 3.8 for an analogous visualization of common semantic units.

The width of the base of a link corresponds to the number of stems common to the two groups it connects. Links limited to one group are the size of the group's entire collection of unique stems (100%). Links are ordered by size within each group. The inner circle of labels denotes the absolute number of stems; the outer circle the percentage.

Note that from this figure, both the absolute and relative numbers of stems can only be read for single links. This is because stems common to more than two groups (and therefore, partially overlapping links) cannot be represented here without harming legibility. Instead, all links are spread and depicted separately, which is why percentages grow to above one hundred.

| | Karakhanid | Kipchak | North Siberian | Oghuz | South Siberian |
|----------------|------------|---------|-------------------|-------|-------------------|
| Karakhanid | 58 | | | | |
| Kipchak | 35 | 264 | | | |
| North Siberian | 5 | 14 | 99 | | |
| Oghuz | 30 | 71 | 12 | 228 | |
| South Siberian | 14 | 26 | 7 | 17 | 51 |

Table 3.10: Number of common reduplicated stems between groups, taken pairwise. The numbers on the diagonal are effectively the number of stems in the given group. Note that this comparison is not limited to stems common to the entire group; all matches between specific languages are included, so that e.g. a stem that has a reduplication in Turkish and Kazakh is counted as an Oghuz-Kipchak match regardless of whether it is also reduplicated in Azeri, Tatar or elsewhere. See fig. 3.5 for a visualization, and tab. 3.16 for an analogous table for meanings.

with other groups amounts two about one and a quarter, and one and a half of the total number of unique stems attested in them. (This is possible because many stems are shared with more than one group, and therefore are counted more than once.)

Further, in the South Siberian, the number of stems shared with the other groups decreases in a relatively gradual manner. In the Karakhanid, a sudden drop can be observed between the link with the Oghuz and that with the South Siberian. In the Kipchak and the Oghuz, similar drops are to be seen after their mutual links, and finally in the North Siberian own stems stand in a marked opposition to all the other ones.

This corresponds well to geography and history. Yakut has been isolated; the South Siberian languages have been closer to the Kipchak group, the Karakhanid to the Kipchak and the Oghuz, and finally the Oghuz and the Kipchak have been close to one another. Again, it seems that reduplications have formed a kind of continuum, where contact with the neighbouring languages has had a greater impact than genealogical affinity. See 3.4.5 for a continuation of this thought.

Moreover, it implies that the South Siberian have been relatively the most conservative, the Karakhanid slightly more innovative, then the Oghuz and the Kipchak, propelling each other, and lastly Yakut that has shaped its rich set of reduplications almost all by itself. Similar conclusions flow from the general characterization of reduplications in specific languages (see 3.2.2).

Finally, focus on specific common stems, and focus on the number of correspondences between different languages can be combined to create a kind of a map of similarities, as in fig. 3.6. (See 3.9 for an analogous visualization of semantics.)

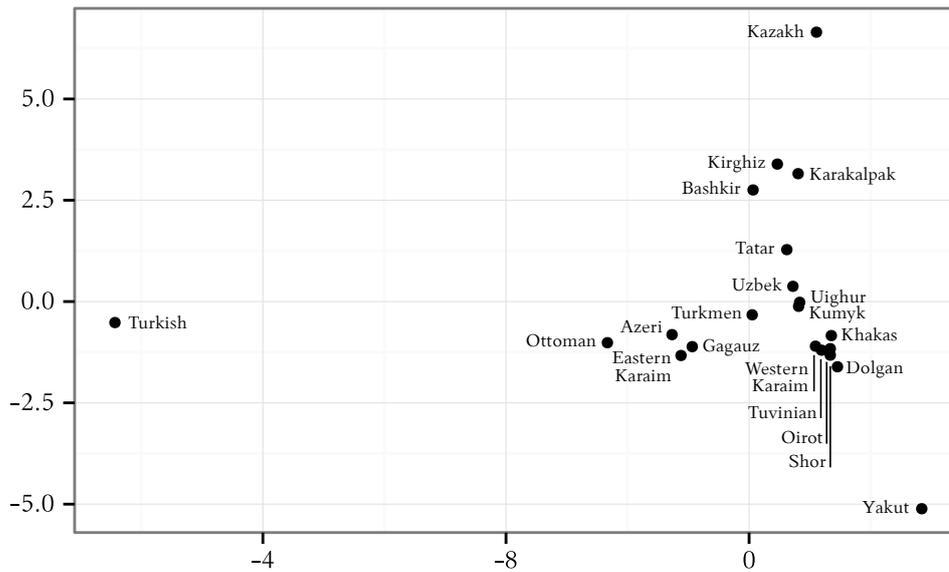


Figure 3.6: Multidimensional scaling⁴⁶ of common reduplicated stems (Euclidean distance; see fn. 72). For each stem (i.e. in each dimension), the distance between two languages is 0 if both have, or both do not have a reduplication of this stem, and 1 if only one of the languages has such a reduplication. Closing consonants are not taken into account, and as a consequence, neither is existence of alternative reduplications of one stem (as e.g. Tksh. *jepjeni* \sim *jesjeni* ‘brand-new’). See fig. 3.9 for an analogous plot of common semantic units.

⁴⁶ In general, data described by two features can be conveniently presented in a two-dimensional plot, as in fig. 3.4, and the distances between them measured easily with a ruler. Addition of a third feature would require the visualization to extend into the third dimension. Beyond this number, graphical representation in print becomes difficult, but theoretical n -dimensional distances between specific items can still be calculated. In essence, multidimensional scaling serves to so arrange items in a y -dimensional space, that the distances between them possibly correspond to the original n -dimensional ones. Superficially, this can be likened to an equidistant projection in geography, where a three-dimensional object (Earth) is so represented in two dimensions (a map), that the distances between specific points on it are preserved.

Here, items are languages, and features that describe them (i.e. that define their location in successive dimensions) are reduplicated stems. The distance between two languages is 1 in a given dimension if one of them reduplicates the given stem, and the other does not. If they both do or both do not reduplicate the stem, they are considered to be located in the same point in this dimension.

In effect then, the data analysed here are simply presence/absence binary data, and therefore the Euclidean distance between two items is equal to $\sqrt{|AB| + |BA|}$. This measure is rather sensitive to the cardinalities of the compared sets, which helps highlight outliers like e.g. Turkish, and through them define the paths of evolution of reduplication. See fn. 72 on other distances.

Perhaps the most interesting conclusions are to be drawn from a comparison with fig. 3.4 which depicted the structural similarity between reduplications in different languages: the size of the collections, and the entropy of closing consonants, while fig. 3.6 visualizes what can be to a certain degree considered genealogical similarities: which languages reduplicate the same stems.

The most visible similarity, it seems, is that in both figures several languages are clustered in one point, from which all the other appear to be departing along three paths. This agreement suggests not only that the simplistic picture in fig. 3.4 is essentially true, but also that, despite all the difficulties involved in their identification, common reduplicated stems do to a certain degree reflect history and genealogy, and thus may be used as an argument while determining the details of the evolution of reduplication in 3.4. Note that this model does not directly correspond to the semantic evolution (see 3.3.4).

There is also a difference between the two figures. Structurally, the Karakhanid languages were divided: Uighur was more like the South Siberian, and Uzbek like the Oghuz. But when specific stems are analysed, the two can be seen to stand much closer to one another, and slightly apart from all the other groups. Somewhat surprisingly, however, Kumyk appears even closer to both than it did in the “structural” figure. It is more probably a result of all three having small collections which barely extend beyond just the basic words, than a sign of actual close genealogical affinity.

From fig. 3.5, it was clear that there is a strong link connecting the Kipchak and the Oghuz languages. Fig. 3.6 seems to identify this link as Turkmen, and to a lesser degree, Uzbek. This is in perfect accordance with geography, suggesting once again the importance of secondary contacts in the evolution of reduplication.

It can be also seen that the second strongest link, the one between Karakhanid and Kipchak, is not due to mainly one language, but that it is both Uighur and Uzbek, that went the same path as the Kipchak, and only did not travel it as far.

Lastly, a note on the position of Turkish. It is a clear outlier because of the large, and unrealistically so (see 2.16.4), size of its collection. But unlike with regard to entropy, here it does not occupy a position intermediate between Azeri and Turkmen. It seems that the Kipchak influence on Turkish consolidated in it the domination of *p* as a closing consonant, but did not strongly affect the choice of specific stems for reduplication.

3.2.5 Common stems and closers

Common stems identified in 3.2.4 above are not necessarily related with one another in the genealogical sense. If, however, an assumption to the contrary were made, the original phonetic motivation behind the choice of the closing consonant could potentially emerge. Let us test this hypothesis.

All the general remarks from 3.2.4 apply. In particular, only inflectional stems are compared, and only those from languages with large sources.

The *mprs*-languages

The *mprs*-group contains eight languages: Azeri, Gagauz, Ottoman, Turkish and Turkmen (Oghuz), Uzbek (Karakhanid/mixed), Eastern Karaim (Kipchak), and Yakut (North Siberian). The genealogical differences manifest themselves in the number of common stems in the case of Yakut, but not any more in the cases of Uzbek and Eastern Karaim; see tab. 3.11.

The general picture is this: there are 340 unique stems, 70 ($\approx 20\%$) of which are attested in at least two languages. Out of the total of 494 reduplications in the *mprs*-languages, these recurring stems are the base of 224 ($\approx 45\%$). The more certain part of them is listed in tab. 3.12.

| | Azeri | Gagauz | Ottoman | Turkish | Turkmen | Uzbek | Eastern Karaim | Yakut |
|----------------|-------|--------|---------|---------|---------|-------|----------------|-------|
| Azeri | 66 | | | | | | | |
| Gagauz | 13 | 41 | | | | | | |
| Ottoman | 19 | 22 | 48 | | | | | |
| Turkish | 32 | 29 | 41 | 159 | | | | |
| Turkmen | 20 | 14 | 17 | 24 | 48 | | | |
| Uzbek | 14 | 10 | 16 | 18 | 16 | 41 | | |
| Eastern Karaim | 17 | 17 | 22 | 32 | 15 | 12 | 42 | |
| Yakut | 3 | 2 | 5 | 11 | 7 | 4 | 5 | 97 |

Table 3.11: Number of common reduplicated stems between the *mprs*-languages, taken pairwise. The numbers on the diagonal are effectively the number of stems in the given language. See tab. 3.7 for all languages discussed in the present work.

Not all *mprs*-languages are equally well suited for comparison. From history, genealogy, and also from tab. 3.11, it is clear that Yakut took its own, isolated path of evolution (which, incidentally, Dolgan did not quite follow; see 2.4). On the other hand, Ottoman data are probably incomplete; they will be essentially omitted here and contrasted with Turkish in more detail in 3.4.6. Further, the largest sources for Gagauz and Eastern Karaim have both less than 20 000 entries, while those for Azeri, Turkish, Turkmen and Uzbek all have 40 000 or more (see the “Sources” subsections in the respective sections in chap. 2). The main comparison will be limited to these four languages in order to avoid excessive reduction of the results, and only the final finds will be checked against Gagauz, Eastern Karaim and Ottoman at the end of this subsection.

But first, a note on Eastern Karaim and Uzbek. The first is not genealogically related to the Oghuz languages in a direct way (although its exact place in the classification is discussed, see e.g. Jankowski 2003) while the latter is essentially a mixture with a substantial Oghuz share. Almost the same number of reduplications is attested for both (43 in Eastern Karaim and 45 in Uzbek), and the data for the latter are in fact probably more complete because the source used was considerably larger. Yet, Eastern Karaim has more stems in common with the Oghuz languages than Uzbek (see tab. 3.11). Most likely, a greater part of the Karaim forms are actually loanwords from Ottoman – the two Western Karaim languages only have twelve reduplications each – but this is impossible to prove without prior detailed and extensive research.

| Stem | Meaning | Azeri | Turkish | Turkmen | Uzbek |
|----------------|-----------|-------------|-------------|-------------|--------------|
| <i>ačyk</i> | open | <i>p</i> | <i>p</i> | – | <i>p</i> |
| <i>ajdyn</i> | bright | <i>p</i> | <i>p</i> | <i>p</i> | – |
| <i>ak</i> | white | <i>p</i> | <i>p</i> | <i>p</i> | <i>p, pp</i> |
| <i>ažy</i> | bitter | – | <i>p</i> | <i>p</i> | – |
| <i>baška</i> | different | <i>m</i> | <i>m</i> | – | – |
| <i>belli</i> | obvious | – | <i>s</i> | <i>s</i> | – |
| <i>beter</i> | worse | <i>š</i> | <i>s, š</i> | <i>s</i> | – |
| <i>boš</i> | empty | <i>m</i> | <i>m</i> | – | <i>m, p</i> |
| <i>boz</i> | grey | <i>m</i> | <i>m</i> | – | – |
| <i>bütün</i> | all | <i>s</i> | <i>s</i> | <i>s</i> | <i>s, t</i> |
| <i>buz</i> | cold | <i>m</i> | <i>m</i> | – | – |
| <i>dik</i> | steep | <i>m</i> | <i>m</i> | <i>m</i> | – |
| <i>diri</i> | alive | <i>m</i> | <i>p</i> | <i>s</i> | – |
| <i>dolu</i> | full | <i>p</i> | <i>p, s</i> | <i>s</i> | <i>p</i> |
| <i>dōru</i> | straight | <i>p, s</i> | <i>s</i> | <i>p, s</i> | <i>p</i> |
| <i>duru</i> | clear | <i>m</i> | <i>p</i> | <i>p</i> | – |
| <i>düz</i> | smooth | <i>m</i> | <i>m, p</i> | <i>p</i> | – |
| <i>gök</i> | blue | <i>m</i> | <i>m</i> | <i>m</i> | <i>m</i> |
| <i>jalnyz</i> | lone | <i>p</i> | <i>p</i> | – | <i>p</i> |
| <i>jassy</i> | flat | <i>m</i> | <i>m, p</i> | – | – |
| <i>jeni</i> | new | <i>p</i> | <i>p, s</i> | <i>p, s</i> | <i>p</i> |
| <i>jenil</i> | light | – | – | <i>p</i> | <i>p</i> |
| <i>ješil</i> | green | <i>m</i> | <i>m, p</i> | <i>p</i> | <i>m, p</i> |
| <i>jumru</i> | round | <i>p</i> | <i>s</i> | – | – |
| <i>jumušak</i> | soft | <i>m</i> | <i>s</i> | – | – |
| <i>kara</i> | black | <i>p</i> | <i>p</i> | <i>p</i> | <i>p</i> |
| <i>karanly</i> | dark | – | – | <i>p</i> | <i>p</i> |
| <i>katy</i> | hard | – | <i>p, s</i> | – | <i>p</i> |

| Stem | Meaning | Azeri | Turkish | Turkmen | Uzbek |
|-----------------|---------|-------------|-------------|-------------|----------|
| <i>kuru</i> | dry | <i>p</i> | <i>p</i> | <i>p, s</i> | <i>p</i> |
| <i>kyrmyzy</i> | red | <i>p</i> | <i>p</i> | <i>p</i> | – |
| <i>kyorak</i> | lithe | <i>s</i> | <i>s</i> | – | – |
| <i>kyzyl</i> | red | – | <i>p</i> | – | <i>p</i> |
| <i>sā</i> | healthy | <i>p</i> | <i>p</i> | – | – |
| <i>sālam</i> | sturdy | <i>p</i> | <i>p</i> | – | – |
| <i>sary</i> | yellow | <i>p</i> | <i>p</i> | <i>p</i> | <i>p</i> |
| <i>sijāh</i> | black | – | <i>m, p</i> | – | <i>m</i> |
| <i>silik</i> | worn | – | <i>p</i> | – | <i>p</i> |
| <i>širin</i> | sweet | <i>p</i> | <i>p</i> | – | – |
| <i>sōuk</i> | cold | – | <i>p</i> | <i>p</i> | – |
| <i>tajjar</i> | ready | – | – | <i>p</i> | <i>p</i> |
| <i>takyr</i> | bare | – | – | <i>p</i> | <i>p</i> |
| <i>tāze</i> | fresh | <i>p, r</i> | <i>m, p</i> | <i>p</i> | <i>p</i> |
| <i>tekiz</i> | smooth | – | – | <i>p</i> | <i>p</i> |
| <i>tekerlek</i> | round | – | <i>s</i> | <i>s</i> | – |
| <i>temiz</i> | clean | <i>r</i> | <i>p, r</i> | <i>p</i> | – |
| <i>turš</i> | sour | <i>m</i> | – | <i>p</i> | – |
| <i>uzun</i> | long | – | <i>p</i> | <i>p</i> | – |

Table 3.12: Common stems in Azeri, Turkish, Turkmen, and Uzbek. The main entry is literary or dialectal Turkish where available, and some other where not. Here, only the main meaning is given; see 3.3.4 for more discussion.

To facilitate identification of patterns, all stems have been assigned to groups describing their distribution across Azeri, Turkish, Turkmen and Uzbek, see tab. 3.13. Always only the largest groups have been used, so that e.g. *ajdyn* ‘bright’ was labelled “P123” rather than “P123, P12, P13, P23”. This is to reduce the number of linguistically irrelevant combinations, and highlight patterns rather than singular exceptions.

It can be seen from tab. 3.13, that the more frequent patterns are simply those that tie successive pairs and triples together. Two points need to be noted.

Firstly, it seems that for any pair of languages, stems which occur in both of them also typically have the same closing consonant in them both (the *P*- and *S*- groups refer to the same languages).

Secondly, the pairs with the highest number of common stems and closing consonants are Azeri-Turkish, Turkish-Turkmen, and Turkmen-Uzbek, which would only be consistent with geography if Turkey and Azerbaijan were swapped. The same observation, that Turkish occupies a place intermediate between Azeri and Turkmen, follows from the entropy of closing consonants, see 3.2.1 above. Particularly interesting in

| Groups | Examples | Stems |
|-------------------|----------|---|
| P1234, S1234 | 8 | <i>ak, bütün, gök, jeni, kara, kuru, sary, tāze</i> |
| P12, S12 | 8 | <i>baška, boz, buz, jassy, kyorak, sā, sālām, širin</i> |
| P23, S23 | 5 | <i>ažy, belli, sōuk, tekerlek, uzun</i> |
| P34, S34 | 5 | <i>jenil, karanly, tajar, takyr, tekiz</i> |
| P123, S23 | 4 | <i>beter, duru, düz, temiz</i> |
| P24, S24 | 4 | <i>katy, kyzyl, sijāh, silik</i> |
| P123, S12, S23 | 3 | <i>beter, düz, temiz</i> |
| P123, S123 | 3 | <i>ajdyn, dik, kyrmyzy</i> |
| P124, S124 | 3 | <i>ačyk, boš, jalnyz</i> |
| P1234, S124 | 2 | <i>dolu, ješil</i> |
| P1234, S23 | 2 | <i>jeni, dolu</i> |
| P1234, S123, S134 | 1 | <i>dōru</i> |
| P1234, S1234, S23 | 1 | <i>jeni</i> |
| P1234, S124, S23 | 1 | <i>dolu</i> |
| P1234, S124, S234 | 1 | <i>ješil</i> |

Table 3.13: Frequent itemsets in tab. 3.12. Names of groups are composed of a letter and a sequence of digits, where *P* stands for ‘present in’, *S* for ‘same closing consonant in’, and *1* for ‘Azeri’, *2* for ‘Turkish’, *3* for ‘Turkmen’, and *4* for ‘Uzbek’, i.e. the same as the columns in tab. 3.12. Redundant subsets have been pruned (e.g. {P123, S134} because it is equivalent to {P1234, S123, S134}).

this context are *beter* ‘worse’, *dolu* ‘full’, *düz* ‘smooth’, *ješil* ‘green’, *temiz* ‘clean’, and partly *jeni* ‘new’, where Turkish has two alternative closing consonants: one as in Azeri (often *m*), and one as in Turkmen (often *s*).

The least frequent combinations are caused by just four words with four different and atypical distributions of closing consonants: *dolu* ‘full’, *dōru* ‘straight’, *jeni* ‘new’, and *ješil* ‘green’. With the exception of *jeni*, they form a rather peculiar pattern: all are present in all four languages, and in all there is one closing consonant that is common to all languages but one. This outlier is twice Turkmen (in *dolu* and *ješil*), and once Turkish (in *dōru*). It seems unlikely that this pattern should be anything more than a coincidence in isolated, more or less random aberrations from the general picture.

Finally, it ought to be noted that the combinations P134 and P14 are missing entirely, and P13 is attested once, but not together with S13. Only six words are not attested in Turkish; five of them (*jenil* ‘light’, *karanly* ‘dark’, *tajar* ‘ready’, *takyr* ‘bare’ and *tekiz* ‘smooth’) are unique to Turkmen and Uzbek, and the last one (*turš* ‘sour’) has always a different closing consonant. It seems that this state might be reflecting more than just the fact that Turkish has the biggest stock of reduplications.

Eleven stems are common to all four, Azeri, Turkish, Turkmen, and Uzbek, and they can be divided into two groups: 1. *dolu* ‘full’, *dōru* ‘straight’, *ješil* ‘green’, and 2. *ak* ‘white’, *bütün* ‘all’, *gök* ‘blue’, *jeni* ‘new’, *kara* ‘black’, *kuru* ‘dry’, *sary* ‘yellow’ and *tāze* ‘fresh’.

The first group was discussed above. The closing consonants are not uniform in it, and therefore too uncertain for phonetic conclusions.

In the second group, there is one closing consonant for every stem that is present in all four languages. In the cases of *gök*, *kara*, and *kuru*, no language has any additional alternative closer. In the cases of *ak* and *bütün*, Uzbek has variants apparently not attested anywhere else. In *jeni*, Turkish and Turkmen share *s* as an addition to the common *p*. In *kuru*, alternative *s* is only found in Turkmen. In *tāze*, Azeri has *p* in addition to the general *p*, and Turkish has *r*.

It seems, then, that all these additional alternative closing consonants are isolated cases which can be safely ignored here. The core set of common stems for the four languages contains eight words (the second group). Reduplications of six are closed by *p*, one by *m* (*gök*), and one by *s* (*bütün*). Given this (relative) uniformity, this set will be assumed to be at least a subset of the common heritage, rather than a mixture of later independent innovations.

As such, it is suitable for an attempt at discovering the alleged original phonetic rules for the choice of the closing consonant. Unfortunately, the results are rather modest because in such a small collection, phonetic particularities rarely recur. The stem *bütün* is the only one that begins with *b-*, has a front high rounded vowel in the first syllable, or *t* for C_2 ; *gök* is not the only one that (historically) begins with *k-*, but all the other ones have their reduplications closed by *p*. As for the vowel, or *k* for C_2 , it is the only one. Let us then broaden the set and include stems which are not common to all four languages.

Tab. 3.12 contains seven words in *b-* in 21 forms and languages in total. The closer *p* only appears once, and it is in Uzbek. Apart from that, *s*, *š* and *t* are used in four words, three of which are more than one syllable long, and *m* in another four, three of which are monosyllables. The tendency is quite clear: in bases beginning with *b-*, *p* is avoided as the closing consonant, and replaced with *s* (typically in polysyllables), *m* (typically in monosyllables), or sporadically some other consonant.

As for *gök*, it is one of seven words originally beginning with *k-*, the only one of them that ever has a reduplication closed by *m*, the only one with an original long vowel or *k* for C_2 , and the only monosyllable. This last feature is consistent with the tendency visible in words beginning with *b-*.

In fact, there are eight monosyllabic words in tab. 3.12 (*ak* ‘white’, *boš* ‘empty’, *boz* ‘grey’, *buz* ‘cold’, *dik* ‘steep’, *düz* ‘smooth’, *gök* ‘blue’, and *sā* ‘healthy’), and six of them have reduplications closed by *m*. The nonconforming two are *ak*, which begins with a vowel, and *sā*, for which there is no ready excuse. The tendency seems to be quite strong, but it is one-way. There are seven more stems, i.e. almost as many, which have

reduplications closed by *m*, and which are not monosyllabic: *baška* ‘different’, *duru* ‘clear’, *jassy* ‘flat’, *ješil* ‘green’, *sijāh* ‘black’, *tāze* ‘fresh’, and *turš* ‘sour’. An observation can be made that the bases of all reduplications closed by *m* begin with an occlusive consonant, *j-*, and only once *s-*, but it will be deceptive: there is actually only one consonant that is attested in anlaut in tab. 3.12, and does *not* have a reduplication closed by *m* (*š-* in just one word, *širin* ‘sweet’).

Overall, stems common to all or some of Azeri, Turkish, Turkmen, and Uzbek, allow for four observations to be made about the phonetic relation between the closing consonant and the base. Their actual phonetic motivation, however, is questionable, as is most clear from tendency number four below. It seems more likely that they either describe the continued use of the original, very simple mechanism (tendency one), or a state created in fact by analogy to a random (?) mutation (i.e. they capture the effect, not the cause; tendency two). Only in one case a phonetic motivation is visible (tendency three), but even then only partially so, because why *m* should be preferred for mono-, and *s* for polysyllables, is not clear at all.

The tendencies, ordered by ranking:

1. Bases beginning with a vowel have their reduplications closed by *p*.
2. Monosyllabic bases typically have their reduplications closed by *m*.
3. Bases beginning with *b-* typically have their reduplications closed by *m* or *s*, or rarely other consonant, but almost never *p*.
4. Closing consonants other than *p* are used almost as often in reduplications of bases that meet the previous restrictions, as in those that do not.

As for stems, it appears that the following set of eight words can be assumed common for the whole of the *mprs*-languages, with satisfactory certainty: *ak* ‘white’, *bütün* ‘all’, *gök* ‘blue’, *jeni* ‘new’, *kara* ‘black’, *kuru* ‘dry’, *sary* ‘yellow’, and *tāze* ‘fresh’. More cautiously, the following three can be added: *dolu* ‘full’, *dōru* ‘straight’, and *ješil* ‘green’.

For pure Oghuz languages, i.e. without Uzbek, this set may be further extended with *ajdyn* ‘bright’, *dik* ‘steep’ and *kyrmyzy* ‘red’, and possibly also *beter* ‘worse’, *diri* ‘alive’, *duru* ‘clear’, *düz* ‘smooth’, and *temiz* ‘clean’, i.e. up to 19 stems in total.

It remains to be seen how these conclusions compare to the languages which were previously put aside for insufficiency of their sources or for complete historical and genealogical detachment, i.e. Gagauz, Eastern Karaim, Ottoman, and Yakut. In tab. 3.14, they are contrasted with what was found above to be the common set of stems and closing consonants.

Overall, the correspondence is quite good. The closer does not match in just one case, Ott. *tāze* ‘fresh’. Since, however, in modern Turkish both *m* and *p* are attested for this word, it will be justified to suspect that this is merely a case of a missing attestation for Ottoman.

| Stem | Common | Gagauz | Eastern Karaim | Ottoman | Yakut |
|--------------|----------------|-------------|----------------|--------------|----------|
| <i>ak</i> | <i>p</i> | <i>p</i> | <i>p, pp</i> | <i>p, pp</i> | – |
| <i>bütün</i> | <i>s</i> | – | <i>s</i> | <i>s</i> | – |
| <i>gök</i> | <i>m</i> | – | <i>m</i> | <i>m</i> | <i>p</i> |
| <i>jeni</i> | <i>p</i> | <i>p</i> | – | – | – |
| <i>kara</i> | <i>p</i> | <i>p</i> | <i>p</i> | <i>p</i> | <i>p</i> |
| <i>kuru</i> | <i>p</i> | <i>p</i> | <i>p</i> | <i>p</i> | – |
| <i>sary</i> | <i>p</i> | <i>p</i> | <i>p</i> | <i>p</i> | – |
| <i>tāze</i> | <i>p</i> | – | – | <i>m</i> | – |
| <i>dolu</i> | <i>p, s</i> | <i>p</i> | <i>p</i> | <i>p</i> | – |
| <i>dōru</i> | <i>p, s</i> | <i>p,s</i> | – | <i>s</i> | – |
| <i>ješil</i> | <i>m, p</i> | <i>m</i> | <i>m</i> | <i>m</i> | – |
| <i>ajdyn</i> | <i>p</i> | – | – | – | – |
| <i>dik</i> | <i>m</i> | – | <i>m</i> | – | – |
| <i>beter</i> | <i>s, š</i> | – | <i>s</i> | – | – |
| <i>diri</i> | <i>m, p, s</i> | <i>p</i> | – | <i>p</i> | – |
| <i>duru</i> | <i>m, p</i> | – | – | <i>p</i> | – |
| <i>düz</i> | <i>m, p</i> | <i>m, p</i> | <i>m, p</i> | <i>m, p</i> | – |
| <i>temiz</i> | <i>p, r</i> | – | <i>p, r</i> | <i>r</i> | – |

Table 3.14: Counterparts of the common stems and closers in languages with incomplete sources, and in Yakut. The stems are divided into four groups: 1. stem and closer common to 1234 (‘Azeri, Turkish, Turkmen, and Uzbek’, see tab. 3.13); 2. stem common to 1234; 3. stem and closer common to 123; 4. stem common to 123.

As for the stems, Yakut is unsurprisingly very different. The other three languages generally attest the common stems, although not without gaps. If the main comparison took all seven into account (Azeri, Turkish, Turkmen, Uzbek + Gagauz, Eastern Karaim, and Ottoman), what can be considered common inheritance of the Oghuz languages would be reduced to nine stems at most: *ak* ‘white’, *diri* ‘alive’, *dolu* ‘full’, *dōru* ‘straight’, *düz* ‘smooth’, *ješil* ‘green’, *kara* ‘black’, *kuru* ‘dry’, and *sary* ‘yellow’. It seems more probable than not, that this would have been an exaggeratedly cautious conclusion.

The *p*-languages

Closing consonants other than *p* do not only occur in the *mprs*-languages. There are nine examples in total in the four languages with mid-low entropy (see 3.2.1),

i.e. in Bashkir, Kumyk, Tatar, and Western Karaim; see tab. 3.15. Interestingly, only *m* and *s* are attested in them, i.e. the second and third most common closers in the *mprs*-languages.

| Stem | Meaning | Oghuz &c. | Bashkir | Kar.NW | Kar.SW | Kumyk | Tatar |
|-----------------|---------|--------------|--------------------|-----------------|-----------------|-----------------|-----------------|
| <i>boř</i> | empty | <i>m</i> | <i>p</i> | <i>m</i> | <i>p</i> | <i>p</i> | <i>p</i> |
| <i>bütün</i> | whole | <i>s</i> | <i>p</i> | – | – | <i>s</i> | – |
| <i>gök</i> | blue | <i>m</i> | <i>m</i> | – | – | – | <i>m</i> |
| <i>jeřil</i> | green | <i>m</i> | <i>m, p</i> | <i>m</i> | <i>m</i> | <i>p</i> | <i>m</i> |
| <i>tekerlek</i> | round | <i>m, s*</i> | <i>p</i> | – | – | – | <i>p, m</i> |

Table 3.15: Closing consonants other than *p* in the *p*-languages. The consonant in the “Oghuz &c.” column is the typical closer in the *mprs*-languages without Yakut. Even where it is attested, *p* is omitted from this column.

* The stem appears five or six times in the *mprs*-languages: Kar.E *tömtögerek*, Tksh. *testekerlek*, Trkm. *testegelek*, *tostogalak*, maybe also **teptegelek* (see 2.17.2), and Yak. *töptögürük*, all meaning ‘completely round’. The geographical distribution of closing consonants is clear: *m* in the Crimea, *s* and possibly *p* in the Greater Middle East, and *p* in the Far East.

It can be seen from tab. 3.15, that where the *p*-languages deviate from *p*, they do so in accord with the *mprs*-languages. What makes the two groups different is really only the proportion of *p* to other closing consonants: low in the *mprs*, and high in the *p*-languages. This suggests that the few non-*p* examples in otherwise purely *p*-languages are in fact just later loanwords.

A closer inspection, however, seems to open another possibility. Except for Tat. *tügäräk* ‘round’, all these stems belong to the core set of the *mprs*-, and the Oghuz languages in particular; *gök* ‘blue’ and *jeřil* ‘green’ are also two of the three earliest attestations of *m* as a closing consonant, and the earliest attestation of *s* is with *tägirmä* ‘round’ (see 2.1.4). Except for Kmk. *büsbütün*, all are closed by *m*, and attested in northern and north-western Kipchak languages.

Kmk. *büsbütün* ‘absolutely all’ is an outlier, both geographically, and as the only example with *s* for the closing consonant. Perhaps it is indeed a later Azeri or Ottoman loanword (see 2.11.3) and unrelated to the other examples, but see similar shapes from the Crimea below.

The unusual geographic distribution of closing consonants in the reduplications of *boř* ‘empty’ in Karaim has already been mentioned in 2.6.3, and the conclusion reached that *m* in this function is probably a common Karaim feature. In *tekerlek* ‘round’, three closing consonants are attested in the *mprs*-languages: *m* in the Crimea, and *p* and *s* elsewhere.

Apart from Kmk. *büsbütün*, all the examples are from Bashkir, Tatar, and Western Karaim, i.e. from a both geographically and genealogically relatively consistent area.

Attempting to dismiss them all as late, independent loanwords might be an overlooking. Let us widen the picture to see if an explanation based on geography or inheritance would be more promising.

Some supplementary data can be found in languages essentially not analysed in the present work (meanings are not given where they are typical): CTat. (AiM) *bomboš*, *büsbütün* ‘absolutely, fully, completely’, *jemješil*, *kömkök*, *tömtögerek* ~ *töstögerek*,⁴⁷ Krm. (Rebi 2004) *bomboš*, *busbutun*, *jemješilik*, Urum (Garkavec’ 2000) *bomboš* ~ *bonboš*, *büsbütün*, *gömgök* ‘completely grey’,⁴⁸ *jemješil*, *tömtögerek*. Reduplications closed by *p* do not seem to be attested for any of those words. Unfortunately, the usefulness of these examples is limited because it is not clear what the general picture of reduplication in these languages is, and how these forms relate to it. It can be only guessed that given the languages’ genealogy, history and geography, they might be similar in this regard to Eastern Karaim.

To this, two Chuvash examples can be added: *jemješal* ‘very green’,⁴⁹ and *kömkövpök* ~ *könkövpök* ‘very blue’ (Ašmarin 1928–50, Krueger 1961: 128). For *pätäm* ‘all, whole’, *pušö* ‘empty’, and *tükerek* ‘round’ no reduplications seem to be attested. It is not clear what can be made of this. According to Krueger 1961: 128, Chuvash reduplications are closed by *p*, “but sometimes -m or even other consonants”. Among fourteen or fifteen examples I know of, ten are closed by *p*, three or four by *m* ~ *n* (*jemješal*, *köm|nkövpök*, *tämtättän* ‘dark’,⁵⁰ and possibly *tämtäkär* ‘downtrodden’⁵¹), and one by *k* (*taktakör* ‘flat’, Krueger 1961: 128).

⁴⁷ Also *tostomalak* ‘very round’.

⁴⁸ Appears to be an Oghuz loanword. The Kipchak shape *kök* is also attested, with more meanings, but apparently without a reduplication.

⁴⁹ Probably a loanword because the regular Chuvash shape is *šeš-* (Egorov 1964), for which no reduplications seem to be attested. Perhaps from Tatar (see Fedotov 1996)?

⁵⁰ Chuv. *tämtättän* ‘completely dark’ (Ramstedt 1952: 250, Räsänen 1949: 239) is not completely clear. It is attested as a separate word with the meaning ‘all black, very black, completely black’, and also in such phrases as *täm karök* ‘a pitch-black capercaillie’, *täm tupolha* ‘a very black filipendula (таволла)’, *täm hura* ‘pitch-black’ (Ašmarin 1928–50). The phrase *täm tākär* ‘downtrodden’ is probably not related, see fn. 51. Quite probably, therefore, Chuv. *täm* is the same stem as Tat. *döm* in *döm karangy* ‘very dark’ (Räsänen 1955: 206), Trkm. *dym* in *dym garangy* ‘very dark’ (TrkmRS), and Uzb. *tim* in *tim qāra* ‘completely black’ (see 3.1.10). Possibly, also Ott. *düm* ~ *tüm* ‘thick, rounded, humped’ (VEWT) might belong here together with Tkc. *tümän* ‘very many &c.’, which is sometimes suspected of semantic interplay with PSlav. **tma* ‘darkness’ (Boryś 2005 s.v. *ćma*, VEWT). Rybatzki 1994: 241 derives Chuv. *täm* < **√tem* > Tkc. *tün* ‘night’ and *temir* ‘iron’, and connects it not only with *tüm kara*, but also with Ott. *düm düz sijah* ‘completely black, pitch-black’. This is perhaps too optimistic. Overall, whether the structural resemblance to reduplications is accidental in these words, or not, must be at present considered debatable. See also 3.1.10 on, at least superficially, similar intensifiers.

⁵¹ Ašmarin 1928–50 lists this phrase inside the entry for *täm* ‘all black, very black, completely black’, but in light of the semantics, it seems more probable that the phonetic similarity is here accidental (see fn. 50). This might be the only Chuvash example that I am aware of, that is a reduplication closed by *m*, and does not correspond to the exceptional forms in the *p*-languages.

The final solution for non-*p*-reduplications in otherwise *p*-languages will require an attitude to be assumed to Chuvash data, which are not clear. If anything, *jemješal* points towards a geographical convergence, *kəmkkvovk* towards genealogical affinity, and the other stems apparently do not reduplicate in Chuvash at all.

Taken together, all these loose pieces of information seem to point to the conclusion that the relation between the deviation from *p* in the five stems in the *mprs*- on one hand, and the westernmost *p*-languages on the other, is deeper than merely a series of relatively modern, independent borrowings. The existence of *büsbütün* ‘absolutely all’ in various languages suggests it might be genealogical in nature, but the general scarcity and diffusion of examples across different languages is perhaps more reminiscent of an areal feature. Nothing certain can be concluded before the many loose ends are attended to but, especially in light of Chuvash data, this is a task beyond the scope of the present work.

See also 3.4.5 on inheritance vs. influence in Turkic reduplications, and see in particular the close link between the Kipchak and Oghuz languages in both Baskakov’s and Tekin’s classifications.

Conclusions

Aware of the limitations of the method, the above subsection tried to husk out a possibly certain rather than numerous set of common reduplicated stems.

In the *mprs*-languages (see 3.2.5), the closing consonants proved to be employed quite uniformly in this common set. However, few phonetic regularities could be discerned. Stems beginning with *b* and monosyllables typically have their reduplications closed by *s* or *m*, but both are used almost as frequently for stems which do not share any particular phonetic feature. Turkish was shown again to occupy an intermediate place between Azeri and Turkmen with regard to the choice of the closing consonant.

The relatively few non-*p*-reduplications in the *p*-languages were found to follow something of a pattern, and to deserve a more minute and more extensive analysis than can be carried out in the present work, because their structure and geographical distribution might suggest the existence of a deeper relationship between them on one hand, and the reduplications in the *mprs*-languages on the other.

As an intermediate result, a set of common stems in the Oghuz languages was identified. In the more conservative version, it contains eight stems (*ak* ‘white’, *bütün* ‘all’, *gök* ‘blue’, *jeni* ‘new’, *kara* ‘black’, *kuru* ‘dry’, *sary* ‘yellow’, and *tāze* ‘fresh’), but in more optimistic variants it can be extended to up to nineteen (first with *dolu* ‘full’, *dōru* ‘straight’, and *ješil* ‘green’, then with *ajdyn* ‘bright’ and *dik* ‘steep’, and finally with *beter* ‘worse’, *diri* ‘alive’, *duru* ‘clear’, *düz* ‘smooth’, and *temiz* ‘clean’). See 3.2.4.

See 4.1.3 for a continuation of considerations on the choice of the closing consonant.

3.2.6 Formation

Most commonly, the formation of Turkic reduplications is described as copying of the first syllable of an adjective, appending of a consonant to it if it was open, or replacing its final consonant with a different one if it was not, and prepending of the whole to the unmodified base. This is an imprecise and biased depiction.

It is imprecise because it does not account for the shortening or any other alteration of the reduplicated vowel, and does not allow for a modification of the base, or for classes other than adjectives. All of these might occur. Admittedly, most are rare phenomena, but incorporating them into the description can in fact result in a more concise formulation (see below).

The customary description is also biased because it implies that the mechanics of the phenomenon is actually known, and this is not the case. I am also not aware of attempts to establish whether it is the case that the mechanics are the same now, when reduplication is essentially no longer productive, as they were when it was still thriving, or when it was only beginning.

For modern Turkish, Sofu 2005 and Sofu/Altan 2009 have shown that “[w]ords beginning with vowels are rule-governed” while “of the words beginning with consonants, frequently used ones seem to be stored individually in the lexicon”, and the infrequent ones are not, and “pose problems in production and [are] more prone to errors” (Sofu/Altan 2009: 72). I believe that this diagnosis can be (provisionally) extended to all the modern Turkic languages which allow more than one closing consonant.

In 3.2.3, it was shown that synchronic phonetic rules cannot accurately explain the use of different closing consonants in modern languages, and in 3.2.5 that correspondences in their distribution are very few even across the most tightly related languages. Combined, these two conclusions suggest that the process of formation of reduplications might have been in fact the same when the phenomenon was still productive, as it is now. Together with multiple closing consonants to choose from, ‘errors’ – i.e., *variation* – in their usage appeared. Bases with a vocalic anlaut are reduplicated uniformly (with just one exception in Yakut, see 3.1.23) and seven bases with a consonantal anlaut have the same closing consonant across Azeri, Turkish, Turkmen, and Uzbek, but the majority appear to be closed somewhat randomly.

Partial corroboration for this diagnosis can be found in those examples where the reduplicated anlaut no longer matches the head of the base (e.g. *žäpžaman* ‘very bad’, see 3.1.19 for a list). In the case of Karakalpak bases in *žä-*, it is the reduplicated anlaut that underwent a change that the head did not undergo, but in almost all the other ones, it is the base that evolved phonetically, and the reduplicated anlaut that remained petrified. It must have been, it seems, perceived as a fixed intensifier linked with a specific word, rather than with this word’s current phonetic shape, that is to say one that is “stored individually in the lexicon”.

As for the earliest period of reduplication, nothing really can be said with certainty. It is not clear how the phenomenon began, and it cannot be determined at present, whether it started with one closing consonant, or more. See 3.4.4 for propositions. Nonetheless, it seems more likely than not that it began from very few, perhaps just one, examples, based on which new ones were formed by analogy. Being emphatic constructions, it might be expected that new reduplications were sometimes slightly modified for greater phonaesthetic effect. This is a very similar picture to the one that emerges for later periods.

Summing up, Sofu and Altan's discovery does not fully explain the mechanics of reduplication, but it does indirectly suggest that, apart from bases with a vocalic anlaut, the two driving forces of the phenomenon are memory and analogy. Naturally, the memorized examples must have been coined at some point as well, and it can be guessed that in the majority of cases, analogy played a significant role in the process.

However, a linguistic description must be more precise than one based on analogy alone ("method for creating forms similar to ..."), which means that it must depart from what appears to be the psychologically most probable explanation. The following formulation could be proposed:

C-type reduplication is the doubling of the initial mora of a word and inserting a single or double consonant in between.⁵² Secondary phonetic modification can be applied independently to any part of the resulting form,

but it should be borne in mind that although simpler, it is not, in fact, less biased than the traditional description mentioned at the beginning of this subsection.

In 1.1.1, the term *reduplication* was defined in terms of the form it produces ("method yielding a form composed of ..."). Of course, this was just evasion, but perhaps such machination is necessary, if the description that is probably the closest to the truth happens to be unsatisfactorily imprecise.

A remark must be made about the theoretical implications. Regardless of whether the description refers to syllables or to morae, the base can be divided into two parts: one that is repeated (the "head"), and one that is not (the "tail"). The C-type is a partial reduplication, therefore neither of these parts can be null (\emptyset). Logically, the two must be also disjoint because no non-null part of a word can be claimed to be simultaneously repeated and not repeated. The border between them falls by definition between the first and the second mora. Now, in the case of bases with a long vowel in the first syllable, this means precisely in the middle of that vowel.

⁵² Here, a syllable is considered to have as many morae, as many of the following features are present in it: any vowel, a long vowel or a diphthong, consonantal auslaut. For example, Tksh. *gök* 'blue' is two morae long, Trkm. *gök* and Yak. *küöç* id. are both three morae long. Thus, the initial mora is always a short vowel with or without a consonant before it.

At least two solutions can be proposed: 1. to interpret long vowels as sequences of two short vowels (Trkm. /gö.ök/ → /gö.m.gö.ök/), or 2. to assume vowel quality and length are separable features (in the IPA notation: /gö:k/, where the head is the first mora, i.e. /gö/, and the tail the remaining two, i.e. /:k/).⁵³ This issue lies beyond the scope of the present work, and will not be discussed further.

3.2.7 Conclusions

This section attempted to investigate some of the main aspects of the structure of Turkic reduplications, and prepare ground for a more diachronic analysis in 3.4.

Several observations were made more than once, independently, and based on different kinds of data. In particular, the conclusion that the evolution of reduplication was influenced by secondary geographical proximity as much or more as by genealogical inheritance, was arrived at in 3.2.1, 3.2.2, 3.2.4, and 3.2.5 (see 3.4.5 for a continuation of these considerations). Linked with it is the observation that the number of reduplicated stems common to two languages tends to be relatively high, but it drops very rapidly as new languages are added to the comparison (3.2.4 and 3.2.5). The Karakhanid and South Siberian languages were found to be the most conservative in the area of reduplication, followed by Oghuz and Kipchak which probably mutually propelled the development of reduplication in each other, and finally by Yakut which, it appears, arrived at a comparable level of advancement all on its own (3.2.2 and 3.2.4). The general tendency (poorly attested languages excluded) for the stage of development of reduplication to be proportional to the distance from the homeland in the Altai Mountains was shown to not hold for Turkish, which instead occupies a place intermediate between Turkmen and Azeri (3.2.1 and 3.2.5, but see also the “General” subsection in 3.2.4). Finally, the tendency for stems beginning with *b-* to not have their reduplications closed with *p-* in the *mprs*-languages was confirmed in 3.2.3 and 3.2.5.

In addition, in 3.2.1 it was suggested that the seeds of multiple closing consonants might have been already present in the initial stock of Turkic reduplications. In 3.2.2, it was shown that there were essentially four paths that the evolution of reduplication took in different languages, and that the peoples in the vanguard of the march west from the Altai Mountains developed this method with greater energy than those closer to the centre. It was also suggested in the same subsection, that the collection of Turkish reduplications is unrealistically numerous. Further, in 3.2.4 it was shown that semantics must have been an important factor in the evolution of reduplication, that all stems which are common to many languages in different genealogical groups,

⁵³ Such dissection might seem bizarre, but it is not unique; see e.g. Pol. /vźeli/ ‘they took’, where *vz-* is a word-formative prefix, *-eli* is a personal suffix, and what is left, the palatalization of *z*, is the root. It was originally *j* (see *j.qć* ‘take-INF’), but note that the sequence [zj] is perfectly valid, e.g. *zjem* ‘I will eat’ : *źem* ‘lands-GEN’.

are names of colours – including the only stem common to all groups, *kara* ‘black’, and that common reduplicated stems can, to a limited degree, be used to draw conclusions about the history of reduplication. Lastly, in 3.2.5 it was observed that among the stems common to the *mprs*-languages, *p* is avoided as the closing consonant for stems beginning with *b*-, and typically replaced by *m* in mono-, and *s* in polysyllables, but both *m* and *s* also frequently occur with other, seemingly random stems, further undermining the conjecture about the phonetic nature of this distinction. It was also suggested that the reduplications in the westernmost Kipchak languages, and maybe also Chuvash, might be related to Oghuz at a deeper level than a series of late borrowings would be. Finally, in 3.2.6 it was proposed that Turkic reduplication be described in terms of morae rather than syllables.

3.3 Semantics

This section gives an outline of the semantics of Turkic reduplications. Unlike stems, meanings can rarely serve as a base for genealogical conclusions, and so extensive comparisons as in 3.2 are generally not made here. First, the grouping underlying all further reasoning is explained (3.3.1), and then the overview proceeds from the most general observations (3.3.2), through broad semantic ranges (3.3.3), down to the more detailed units (3.3.4). The whole closes with a summary and conclusions (3.3.5).

3.3.1 Grouping

The meanings of reduplications that are given in the present work were mostly taken from general purpose dictionaries of various sizes (several thousand to more than seventy thousand entries), and written in various languages (most often Russian, but also Belorussian, English, French, German, Italian, Latin, Polish, Spanish, Turkish, Ukrainian, and others). There is a great variability in the extensiveness of definitions, and in the choice of specific synonyms in translation (e.g. ‘беленький’ : ‘белый-пребелый’ : ‘completely white’ : ‘Schneeweiss’). In this collection, unification of the depth and breadth of definitions is only achievable towards the general.

As it stands, a more detailed collection of equal scope cannot be assembled without prior extensive field research. This is beyond the scope of the present work, and beyond its primary goal which is to give a general overview of Turkic reduplications as a whole, and of the main patterns in their evolution.

Because of this, all reduplications have been assigned to semantic groups at two levels, the more general *ranges* (3.3.3) and the more detailed *units* (3.3.4), and all the reasoning in the current section will be based on them rather than on the exact definitions provided by the sources. The exact assignment is detailed in the index on p. 273.

3.3.2 General

The primary meaning of Turkic reduplication is intensification of adjectives and adverbs.

Given the generally weak distinction between parts of speech within nomina, it is not surprising that also bases with a stronger or even strong nominal character occur quite frequently, but their reduplications are almost always adjectival. There are only three clear exceptions: Kirg. *ynak* ‘1. clean; 2. close friend’ → *ypynak* ‘very close friend’, and two words in Azeri where reduplication apparently has the meaning of plural, *bäzäk* ‘decoration’, and *söküntü* ‘chip, splinter’; see 2.10.5 and 2.2.5, respectively.

Other parts of speech, pronouns and verbs, are very rare. Also in these cases, reduplication serves primarily to intensify (see 3.1.17 and 3.1.22).

Semantic shifts during reduplication are not particularly common, mostly trivial, and the majority are independent, singular cases. If any pattern can be observed, it is that typically the first, and often only the first, meaning is reduplicated. This tends to be concrete, but figurative meanings are sometimes added to the reduplication, which were apparently missing from the base, as in Dolg. *kytarkaj* ‘red’ → ‘pretty as a picture’, Kkpl. *kojyw* ‘thick, dense’ → i.a. ‘Okay! Good!’, Kmk. *tegiš* ‘smooth, even’ → i.a. ‘dog-poor’, and Oir. *tünej* ‘same, similar’ → i.a. ‘all right, okay’ (see the “Semantics” subsections in the respective sections in chapter 2).

In particular, reduplications of *ak* ‘white’ function in many languages with such meanings as ‘good’, ‘nice’, &c., which are often not explicitly present in the semantics of the base. A similar evolution, interestingly, can be observed in Tuv. *kara* ‘black’. These figurative reduplications of *ak* are usually closed with a double *pp*, and stand in opposition to alternative forms closed with a single *p* which only intensify the original, concrete meaning. See 3.1.2 for examples of different closing consonants carrying different meanings.

As far as meanings can be considered positive or negative, the former seem to outnumber the latter. This is also reflected in the fact that the privative suffix *-syz* is apparently only attested in three bases: Tkc. *ansyz* ‘sudden(ly), unawares’, Tksh. *išsiz* ‘unemployed, jobless’, and Tksh. *yssyz* ‘desert(ed), isolated’, while its inverses, *-ly* and others, are fairly common, e.g. Az. *bäzäkli* ‘decorated, fancy, chic’, Gag. *kirli* ‘dirty’, *tülü* ‘hairy, feathery’, Khak. *tadylyg* ‘sweet’, Kirg. *tyrmaktaj* ‘with nails, with claws’, Tksh. *renkli* ‘coloured, colourful’, Yak. *tüstāχ* ‘salted’&c.

Incidentally, a similar observation can be made about Polish diminutives. This analogy might be flawed by that they very often have an additional either hypocoristic or contemptive side to them (e.g. *śoneczko* i.a. ‘[my] sunshine of a person’, or *cwaniaczek* ‘small-time crook’), but a general, vague similarity in the choice of meanings for intensification can nevertheless be seen. Most probably, it should be explained by simply common sense which suggests that only selected meanings have the capability to be meaningfully intensified.

3.3.3 Ranges

The units of the more general level of grouping will be here called *ranges*. Perhaps the term *domain* would also be applicable as what they attempt to capture is essentially “a coherent knowledge structure [...] against which other conceptual units such as concepts are characterised” (Evans 2007 s.v. *domain*), but the term *range* will be preferred because the specific choice of levels of organization differs from that most often found in cognitive linguistics. This is because adjectives expressing the middle values on a scale can only rarely be intensified in a meaningful way (‘fiercely mild’), and many traditionally discerned domains, e.g. TEMPERATURE, would be reduced to just two polar antonyms. Instead, wider and more vague ranges were used, e.g. comprising the whole of experiences coming from one sense, such as *touch* which includes TEMPERATURE, but also ROUGHNESS, WETNESS, &c. The small caps notation, however, will be employed. Polysemantic words were assigned to multiple ranges.

A completely objective, impartial and unbiased grouping does not seem to be attainable, and the one employed here is certainly not the only one possible. For this reason, its role will be limited to providing a general overview of the semantics of Turkic reduplication, and conclusions flowing from it will only serve as auxiliary for the ones drawn from the analysis of semantic units in 3.3.4 below.

Overall, the reduplications collected here fell into 14 ranges so chosen as to minimize the unavoidable overlap and mixture of levels of generalization:

- physical:
 - external: COLOUR (e.g. ‘black’, ‘motley’, ‘redhead’), SIZE (e.g. ‘big’, ‘low’, ‘shallow’), and SHAPE (e.g. ‘bulge’, ‘sloping’, ‘squat’),
 - internal: NATURE (e.g. ‘elastic’, ‘liquid’, ‘sturdy’), and
 - other: STATE (e.g. ‘ripe’, ‘still’, ‘young’),
- psychic: CHARACTER (e.g. ‘agile’, ‘angry’, ‘friendly’), PERCEPTION (emotional, qualitative, e.g. ‘awkward’, ‘obvious’, ‘unpleasant’), and QUALITY (subjective evaluation, e.g. ‘bad’, ‘master’, ‘wrong’),
- senses: APPEARANCE (e.g. ‘blurry’, ‘clean’, ‘shiny’), TASTE (including SMELL; e.g. ‘juicy’, ‘sweet’, ‘unpalatable’), and TOUCH (e.g. ‘cold’, ‘dry’, ‘rough’), and
- other: LOCATION (in time and space, e.g. ‘around’, ‘noon’, ‘moment’), QUANTITY (e.g. ‘all’, ‘few’, ‘sparse’), and OTHER (e.g. ‘deletion’, ‘itch’, ‘together’).

As can be seen from the enumeration, the semantic scope of Turkic reduplications reaches far beyond colour names, which is what it is most commonly associated with. Almost all the most important aspects of everyday life are included in one language or other. I could only find one range that seems to be essentially omitted, namely HEARING. Only such meanings as ‘silence’, ‘silent’, and ‘quiet’ are attested, but these

tend to be entangled with ‘calm(ness), peace(ful(ness))’. Meanings more readily associated with HEARING, ‘roar’, ‘squeek’, ‘whisper’, &c., apparently do not undergo reduplication, even though their intensification is certainly conceivable. Possibly, Tuv. *sapsajtyk* ‘1. imitation of a gopher’s scream, ...’ is an exception – if it is indeed a reduplication (see 2.18.4).

There are considerable differences in the cardinality between the attested ranges. The exact numbers, however, are purposefully not given at the moment because they would be misleading for two reasons. One is the vague and arbitrary nature of the grouping, and the other is that each range has a different, and not quite definable, natural limit of semantic units that can belong to it, and be meaningfully intensified. No range is composed of fewer than eight units; the least numerous are (in the alphabetical order) TASTE, TOUCH and QUANTITY, and the most numerous SHAPE, STATE, and PERCEPTION. (If necessary, the exact assignments are detailed in the index on p. 273.)

Also the number of bases representing the given range in different languages varies significantly, see fig. 3.7. Again, the exact numbers will not be given, and no conclusions should be drawn from the details, but the general picture is quite clear.

For all ranges, the number of representatives grows together with the total number of reduplicated bases in the given language, and for all ranges except COLOUR, there are very few or just no representatives at all in the languages with the lowest total number of reduplicated bases. Taking into account the vagueness and arbitrariness of ranges, together with their natural limitations, the growth rate is approximately similar for most.

This suggests that COLOUR is a privileged range, which is consistent with the results of the comparison of common stems (see 3.2.4). It seems that all languages inherited the basic stock of reduplications of colour names and not much more, and developed it without any particular semantic or structural preference, in approximately the same directions.

3.3.4 Units

The units of the more detailed level of grouping will be called simply *semantic units* (see the index on p. 273). To better reflect the Turkic specificity, the division was based not only on semantics but also on etymology. The highly fragmented and diversified dictionary definitions (see 3.3.1) were first grouped by meaning, and then a series of rearrangements has been made, joining into one unit those meanings that are typically carried by historically one stem across various Turkic languages (e.g. Tkc. *ačyk* ‘1. bright; 2. open;’), and separating those that seem very close from the English perspective, but are typically carried by historically different stems across the Turkic languages (e.g. Tksh. *berrak*, *duru* ‘limpid, clear, transparent’ : *ačyk*, *ajdyn(lyk)* ‘bright, clear, light, open’ : *āšīkar*, *belli* ‘obvious, certain, clear, explicit, well-known’). Turkic polysemantic words were assigned to more than one unit. A particular effort was made

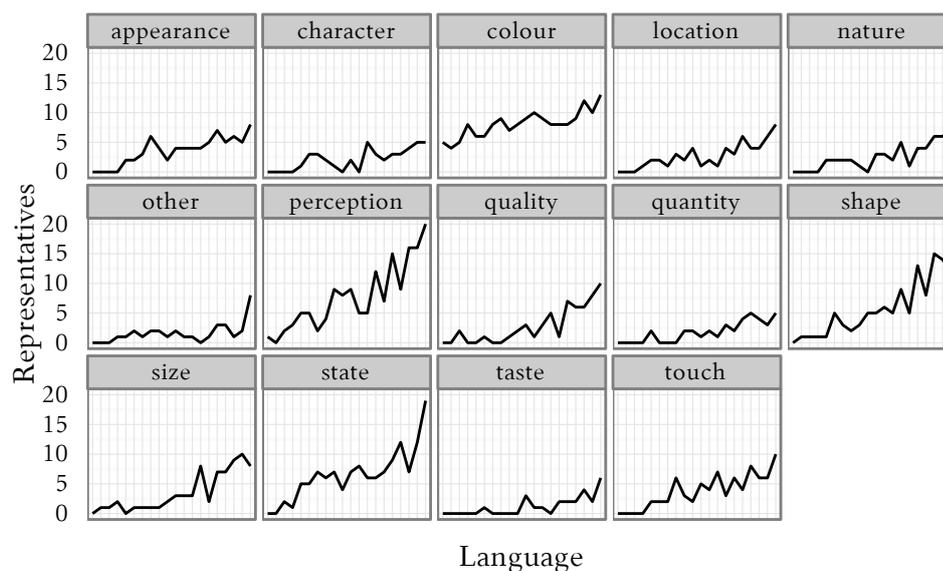


Figure 3.7: Number of representatives of semantic ranges in different languages. Languages are not labeled, but are ordered in each plot by the total number of reduplicated bases, same as in fig. 3.10.

to find the golden mean of the level of detail, but a significant proportion of arbitrariness and subjectivity could not be avoided in the final result.

Overall, the present collection fell into 214 semantic units. However, the oldest attestations will be ignored here because their dialectal affiliation is unclear. In particular, three bases not attested in later sources were the only representatives of their respective units: *alčak* ‘friendly, mild’, *kötgi* ‘protruding’, and *tutčy* ‘1. contiguous, osculant; 2. permanent, continuous’ (see 2.1.2).

The great majority of units are represented by very few examples. In fact, 93 ($\approx 44\%$) units are only attested in one language, and further 35 ($\approx 16.6\%$) in just two, more often than not, unrelated languages. The top of the list is appropriated by basic colour names: ‘black’, ‘blue’, ‘green,’ ‘red’, ‘white’, and ‘yellow’, all attested in sixteen or more languages out of twenty-one. The only non-colour unit with so many attestations is ‘round’ (sixteen languages). Following them are some basic adjectives that do not belong to a single semantic range: ‘accurate’, ‘all’, ‘bright’, ‘clean’, ‘direct’, ‘dry’, ‘empty’, ‘even, smooth’, ‘full’, ‘long’, ‘naked’, ‘new’, and ‘obvious’. All the remaining units are represented in ten or fewer languages.

Several colour names can also be found among the less numerous represented units: ‘brown’, ‘grey, pale’, ‘orange’, ‘pink’ and ‘purple’, together with bases only indirectly referring to colour: ‘bright’, ‘dark’, ‘motley’ and ‘redhead’ (all attested in no more than nine languages).

Common units

Let us now inspect how many semantic units are shared between the five genealogical groups the Turkic languages are divided into, and whether their distribution in the specific languages forms a pattern of some kind.

The number of common units are given in tab. 3.16, and visualized in fig. 3.8. It can be seen that the width of links decreases gradually in each group; a larger drop can only be observed between the North Siberian – Oghuz, and the North Siberian – South Siberian links, but it seems to have only been caused by that the South Siberian languages have a significantly less numerous collection than the North Siberian. Most importantly, semantic units shared with other groups constitute a very high percentage of the total number of units attested in each group, which appears to imply that each group’s own contribution to its collection was relatively small.

This is in stark contrast to the analogous visualization of common stems in fig. 3.5, where stems unique to a given group typically dominated the picture. There, I interpreted this situation as a sign that the inherited stock must have been relatively small, and the large Kipchak, Oghuz and North Siberian collections arose independently from one another.

Coming back to semantic units, the picture that they create is to be expected only if each group has inherited a considerable portion of its units from an earlier stage of development, or if all groups expanded their respective collections as if according to a common plan. Clearly, the first possibility cannot be the case here, and therefore the second explanation must be true. It is in accord with what could be deduced based on semantic ranges in 3.3.3.

| | Karakhanid | Kipchak | North Siberian | Oghuz | South Siberian |
|----------------|------------|---------|-------------------|-------|-------------------|
| Karakhanid | 42 | | | | |
| Kipchak | 39 | 138 | | | |
| North Siberian | 18 | 54 | 79 | | |
| Oghuz | 37 | 85 | 49 | 135 | |
| South Siberian | 23 | 37 | 20 | 30 | 40 |

Table 3.16: Number of common semantic units in genealogical groups, taken pairwise. The numbers on the diagonal are effectively the number of semantic units in the given group. Note that this comparison is not limited to units common to the entire group; all matches between specific languages are included, so that e.g. a unit that is present in Turkish and Kazakh is counted as an Oghuz-Kipchak match regardless of whether it is also there in Azeri, Tatar or elsewhere. See fig. 3.8 for a visualization, and 3.10 for an analogous table for stems.

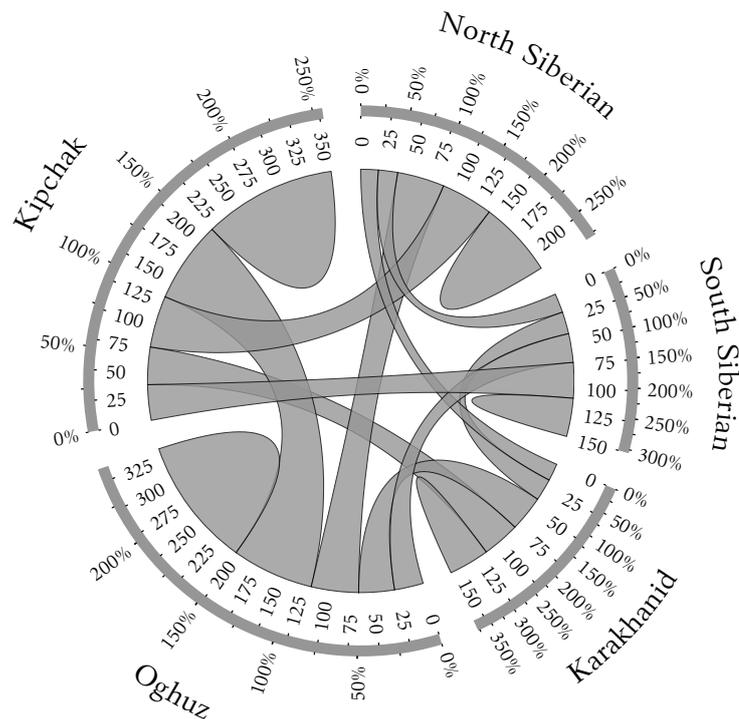


Figure 3.8: Visualization of the number of common semantic units between genealogical units, taken pairwise. See tab. 3.16 for the exact numbers, and fig. 3.5 for an analogous visualization of common stems, and an instruction on how to read it.

A similar, albeit not identical, picture emerges from a comparison of specific units shared by specific languages, as illustrated in fig. 3.9. Yakut is an outlier in it, but all the remaining languages are located approximately along a single line. This suggests, again, that as far as semantics is concerned, reduplication evolved according to a quite similar plan in all languages, but it also shows that one language deviated from it, and that this language is unsurprisingly Yakut. This observation was lost earlier because fig. 3.8 only gave a general picture of the *number* of common units between *groups*, whereas fig. 3.9 offers a more detailed view of specific units in specific languages.

It can be also seen in this figure that the spread of languages from right to left corresponds to the total number of reduplications in them (see fig. 3.10), which is understandable. Reduplication began with colour names, and then new semantic units were added. The process stopped at various points in various languages, and those that obtained a similar set in the result, are closer to one another in fig. 3.9. In this

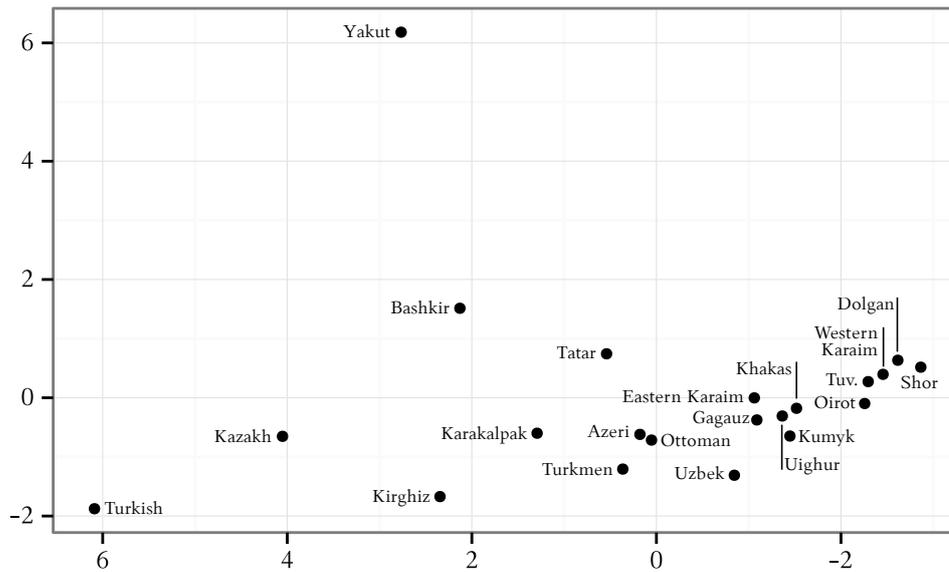


Figure 3.9: Multidimensional scaling of common semantic units (Euclidean distance; see fn. 72). See fig. 3.6 for an analogous plot of common stems, and an instruction on how to read it. For each unit (i.e. in each dimension), the distance between two languages is 0 if both have, or both do not have a reduplication with this meaning, and 1 if only one of the languages has such a reduplication. Note that similarly to fig. 3.4, the horizontal axis has been reversed here to highlight the likeness to the geographical distribution (both plots and maps are typically oriented according to arbitrary conventions, see 3.2.2).

sense, this figure might be also regarded as a diachronic scheme of semantic evolution of reduplication, not unlike fig. 3.4 and 3.6 could be for stems.

Overall, the evolution of reduplication appears to have taken two paths when viewed through the lens of semantics, which is two less than when viewed through the lens of common stems (see 3.2.4). This is only seemingly contradictory. There must have been four paths, but each of them has independently involved similar, perhaps the most frequent of the intensifiable semantic units, and these were expressed by different stems in each group. The paths were separate but semantically almost parallel – with, of course, the exception of Yakut which has deviated quite far from its kin in both regards.

This observation has serious repercussions for the reconstruction of the detailed history of Turkic reduplication; see 3.4.2.

Relation to bases

The numbers given above might suggest that the semantic units discerned here are in fact tied to bases in a nearly one to one relationship, but it is not quite so. Fig. 3.9 shows that the independence of semantics has been preserved sufficiently to allow purely semantic conclusions. Nevertheless, a close correlation with the number of bases is observed, and illustrated in fig. 3.10 and 3.11.

Although far-reaching conclusions cannot as yet be drawn from this fact, it might be interesting to note that this relation can be quite well approximated using linear regression.⁵⁴ Let us abbreviate the number of bases to n_b , and the number of semantic units to n_u . The formula is: $n_b = 1.413 \cdot n_u - 3.924$, which results in $R^2 = 0.9826$, i.e. a very good fit. However, the fact that it is linear implies that both n_b and n_u can in theory continue to grow at a constant pace until one of the resources is exhausted in the language, and their mutual relation will always remain the same. In other words, it suggests that this relation might not be so much a property of Turkic reduplications, but actually of the Turkic languages themselves – or, of how semantic units have been established here.

It seems impossible to determine which is the case without an investigation reaching beyond the scope of the present work. Nevertheless, a mention can be made that, unsurprisingly, non-linear regression can also be used to approximate our relation. The formula is: $n_b = 0.8839 \cdot n_u^{1.0982} + 0.6671$, which results in $R^2 = 0.9838$, i.e. a marginally better fit. This slight improvement might be a sign of overfitting, but from fig. 3.11 it can be seen, that the difference lies primarily in the languages with the lowest, and the highest n_u . Extrapolated, this implies that after a certain point, the growth of n_u should begin to decelerate in relation to the growth rate of n_b , i.e. that after that point new bases can still be reduplicated, but they will mostly belong to the already represented semantic units. If this is the case, then apparently the Turkic languages have not yet reached that point.

The above considerations operate on the silent assumption that the semantic evolution of reduplication proceeded according the same plan in all the languages, and therefore that they can all be regarded as successive snapshots of just one history. This appears to be true in general, as can be inferred from fig. 3.9, but Yakut is clearly an exception. In fig. 3.11, however, it is only slightly more of an outlier than Azeri or Turkmen. It is not clear to me what should be made of this fact.

Overall, the relation might be interesting. Research that would reach out to other languages, however, is necessary to determine whether it is a singular coincidence, or a more general tendency. It should also help assign a linguistic meaning to the coefficients, because without it the value of this observation is negligible.

⁵⁴ Regression is a technique of estimating relationship between variables. Various indices are employed to evaluate the goodness of fit of the resulting model to the observed data. Here, only the R^2 coefficient of determination will be used, which essentially represents the proportion of variability in the dependent variable that the model accounts for.

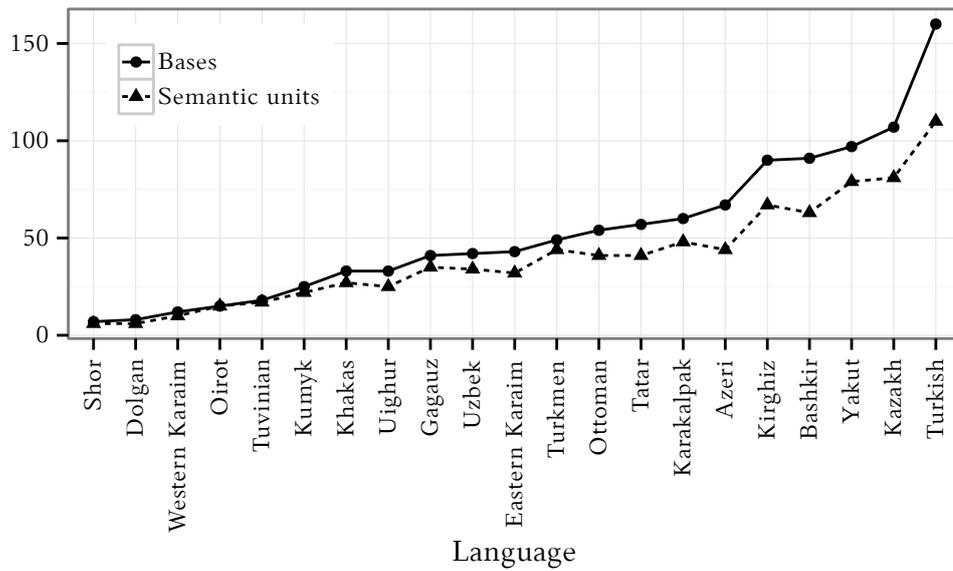


Figure 3.10: Number of reduplicated bases and semantic units that they represent, in different languages.

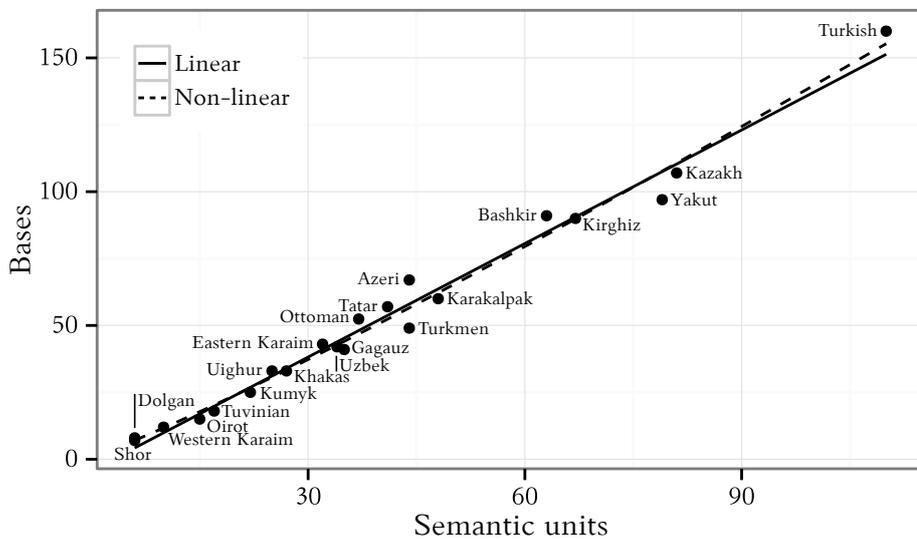


Figure 3.11: Number of reduplicated bases against the number of semantic units that they represent, in different languages. The parameters of the two regression lines are given in the main text.

Relation to the closing consonant

Maḥmūd al-Kāšyarī mentions eight or nine reduplications. The unclear one is *čimjig* ‘very raw’ (see 2.1.3). In most cases, his definitions do not suggest a connection between the “particle”, as he interprets the reduplicated anlauts, and the meaning of the stem. On one or two occasions, the wording may be seen as slightly misleading (e.g. “When the Oγuz wish to exaggerate the description of anything round they say: TAS·TAKIR·MA’ **täs tągirmä**.” (Dankoff/Kelly 1982: 261)), but in two cases the existence of such link is declared quite clearly: “**üp** A particle of emphasis for colors.”, and “**čim** An exaggerative particle of dampness or rawness.” (Dankoff/Kelly 1982: 87, 267).

If al-Kāšyarī’s definitions are adequate, then what is interpreted here as emancipated reduplicated anlauts (e.g. Kirg. *kyp* in *kyp žyłaŋaç* ‘completely naked’ < *kyp kyzyl žyłaŋaç* lit. ‘red-naked’), might be in fact ancient intensifiers with a strictly limited scope. A number of such formations is listed in 3.1.10. However, semantic links between them are very few. Effectively the only one that could perhaps be interpreted in a way similar to al-Kāšyarī’s *čim*, is Uzb. *žim* ‘quietly, calmly; silently, tacitly’, but the etymology of this form is not clear. It might have once been an independent word, but it is also possible that it was extracted from the reduplication *čymčyrt* or similar; see 2.19.3.

Full reduplicated anlauts are not, in fact, likely to have ever been strongly connected to semantics on a regular basis, but before abandoning this idea entirely, let us see whether a connection can be observed with just the closing consonant.

Three or four of the reduplications mentioned by al-Kāšyarī are closed with consonants other than *p*. Apart from *čimjig* ‘very raw’, it is *kömkök* ‘very blue’, *sümsüčig* ‘very sweet’, and *tästägirmä* ‘very round’. In 3.2.5, a set of stems common to the Oghuz languages was established; of them, the following have their reduplications not closed by *p*: *bütün* ‘all, whole’ (*s*) and *gök* ‘blue’ (*m*), and, with a lower probability, *dik* ‘steep’ (*m*), *diri* ‘alive’ (*m, s*), *duru* ‘clear’ (*m*), *dolu* ‘full’ (*s*), *dōru* ‘direct, straight’ (*s*), *düz* ‘smooth’ (*m*), and *ješil* ‘green’ (*m*).

Very little, it seems, can be deduced from this collection. Three words with reduplications closed by *m* form a pattern: *jig* ‘raw’ – *ješil* ‘green’ – *kök* ‘blue’ (note that the two colours were often not distinguished in the Turkic languages). The latter two, surprisingly, also have reduplications closed by *m* in Bashkir and Tatar, where closing consonants other than *p* are almost never used. Another triple can be found in *dolu* ‘full’, *tägirmä* ‘round’, and *bütün* ‘all, whole’, all with reduplications closed by *s*. However, five words with reduplications closed by *m* (*dik*, *diri*, *duru*, *düz*, and *süčig*), and two with reduplications closed by *s* (*diri* and *dōru*) are thus left with no company at all.

An inspection of the stock of modern languages brings no conclusions, either. Semantically, words with reduplications closed by consonants other than *p* are equally

diverse in the Oghuz languages, in Yakut, and in the few Kipchak languages where such forms are attested.

Overall, what could be considered a hint in al-Kāšyārī's dictionary proves to be a cul-de-sac. There does not seem to exist any systematic correlation between the meaning and the phonetic shape of the intensifying elements.

3.3.5 Conclusions

This section attempted to give an overview of the semantics of Turkic reduplications, but generally abstained from drawing definite conclusions on genealogy because meanings do not, in and of themselves, provide sound support for such considerations.

To minimize the inequalities and irregularities in the original definitions, which have been extracted from a number of dictionaries in various languages, the meanings have been grouped into more general *ranges* and more detailed *units* (3.3.1). Following some more or less loose general remarks (3.3.2), semantic ranges were presented showing that the semantic diversity of reduplications grows together with the total number of reduplicated bases but without preference for any particular range, and that COLOUR is the only range that is well represented in all languages (3.3.3). Next, semantic units were presented (3.3.4), and it was concluded that all languages except Yakut developed reduplication as if following essentially the same semantic plan. This constitutes a serious impediment to the study of the history of reduplication by undermining the feasibility of the historical-comparative method. Also, a correlation was observed between the number of semantic stems in a language, and the number of reduplicated bases in it, but despite what might appear to be a hint in al-Kāšyārī's dictionary, no link was found between the phonetic shape of the reduplication and its meaning.

3.4 History

This section attempts to draw the general outline of the history of reduplication in the Turkic languages. It begins with methodological considerations (3.4.1 and 3.4.2), and later proceeds to sketch the evolution of reduplication from its possibly Altaic beginnings (3.4.3 and 3.4.4) through a period of development which, it appears, was more due to influence than inheritance (3.4.5), up to the changes in the only language with a considerable written history, which is Turkish (3.4.6). The whole is concluded and summarized in 3.4.7.

3.4.1 Synchronicity

A greater part of the study of Turkic reduplications so far has concentrated on establishing synchronic, phonetic rules for choosing the closing consonant in Turkish

(see 1.1.2). This attitude is methodologically fragile because it hinges on two unprovable assumptions: 1. that reduplications are systematic, and 2. that they were created observing some tangible set of rules.

It is known that there are doublets in Turkish, two reduplications of the same stem, each closed with a different consonant (e.g. *ješil* ‘green’ → *jemješil* and *jepješil*, see 2.16.4), as well as etymological doublets where each variant continues a different path of phonetic evolution (e.g. *jašyl* : *ješil* ‘green’). The material collected in the present work shows that this situation is no way a specificity of Turkish (see 3.1.1).

Inconsistencies are also apparent between languages. Very different sets of stems can be reduplicated, and the reduplications of etymologically one stem can be closed with different consonants (see 3.2.4 and 3.2.5).

Thus, the first assumption does not hold. As for the second one, it is debatable whether it can be disproved; so far, it has not been proven. However, since the primary meaning of reduplication is intensification, and there is no particular reason to believe that reduplication is somehow special, it seems more likely that it was applied in the same way that intensification would typically be expected to, which is to say irregularly. For these two reasons – failure of the first assumption and probable incorrectness of the second one – the search for synchronic *rules* must be considered futile. *Tendencies* are the best that can be hoped for.

In light of the presently observed diversity, it seems rather probable that these tendencies varied between periods and places. The synchronic stock in any given language is most likely a mixture of forms, some of them inherited and others borrowed or calqued, which resulted from different sets of tendencies. Not one set is to be sought but many, even to explain the state in just one language.

Therefore, the correct approach is the historical one.

Internal reconstruction, however, seems impossible. To separate groups of reduplications in one language which follow a given set of phonetic or other tendencies, is a trivial and useless exercise. It is trivial because, not being limited by external data, one may create up to as many groups as there are reduplications and claim a different set of tendencies for each. Evaluation of the result cannot be based but on impression and personal taste, which is what also makes the exercise useless.

Thus, the methodological choice is narrowed to effectively one method, the historical-comparative one. Unfortunately, its implementation proves rather difficult in practice. See 3.4.2 below.

3.4.2 The historical-comparative method

At the current stage of research, the classical historical-comparative method cannot be safely applied to Turkic reduplications. This is for a number of reasons.

Perhaps the most important of them is our insufficient knowledge of the etymology and history of specific words, and of the detailed phonetic history of specific languages. This raises two problems.

Firstly, identification of bases reducible to a common root is severely impeded, and even more difficult it is to determine which ones can be brought down to a common stem. For this task to be completed properly, a comprehensive comparative dictionary is necessary, but this does not exist, and compiling one considerably exceeds the scope of the present work.

Secondly, forms borrowed between the Turkic languages are too often impossible to tell from inherited ones. In the case of reduplication this problem is made worse by the very real possibility of calquing.

Further, many reduplications are derivatives, e.g. Az. *joꝛjoꝟsul* ‘poor, destitute’ : *joꝛjoꝟsullug* ‘poverty, destitution’. The latter might be a derivative of the former, but it might also be a reduplication of *joꝟsullug*. This is probably not possible to determine.

The same applies to related words in different languages, e.g. Tksh. *geniš* ‘wide’ : Kirg. *keŋ* id. : Trkm. *gīŋ* id. : Yak. *kieŋ* id. Here, conjectures are made possible by the existence of Kirg. *keŋiš* and Trkm. *gīŋiš* id. which do not seem to have reduplications, and Kirg. *kenen* i.a. ‘spacious’ which does have one. However, conclusions drawn from what appears to be a lack of attestations, are at best problematic. This difficulty is particularly visible with the Turkic, where the politically less prominent languages sometimes even lack middle-sized dictionaries.

Next, in the case of a widely applicable but not very frequent derivational suffix, the existence of analogous forms in different languages may sometimes suffice to posit a common origin because it might not be very likely that of all words, these languages should all choose this particular one independently. Reduplication is perhaps sufficiently rare for such reasoning but the number of bases it can be used with is significantly lower. The existence in different languages of reduplications of etymologically the same base does not in any way prove a common descent.

Moreover, word formation is in itself not a regular process. The fact that reduplications carry in them an additional emotional load only intensifies this, as is attested by various non-standard formations; see in particular 3.1.13.

This helps understand the results of H.-G. Müller’s questionnaire, where only 84.2% of the 125 interviewed native speakers knew the actual literary reduplications in Turkish (see appendix A.2). Being emotional formations, their temporal and spatial reach may often be limited. Our knowledge of reduplications cannot be full without dialectal data, but unfortunately, comprehensive dialectal dictionaries are a scarcity in the Turkic world.

Furthermore, the exact relations between specific languages are not thoroughly understood. The traditional grouping into Karakhanid, Kipchak, North Siberian, Oghuz, and South Siberian is usually interpreted as genealogical in nature, but this is a simplified

view. It is not certain that e.g. a single Proto-Oghuz *language* can really be postulated. This is a serious impediment to the reconstruction of the history of reduplication.

Finally, a confrontation of the results of the structural and semantic analyses conducted here (see 3.2.7 and 3.3.5) brings another difficulty to light. It appears that the great majority of languages expanded their stock of reduplications as if following almost the same semantic plan (see 3.3.4). This means that quite different sets of stems were reduplicated in different genealogical groups, but inside one such group, this means that a quite similar set of reduplicated stems must be expected in most languages in it, regardless of whether they had inherited their respective stocks or developed them independently. A detailed reconstruction of the genealogy of reduplication seems to not be possible at the present stage of research.

But all is not lost. Some of these limitations can in fact be used to our advantage. The semantic limitation of reduplication suggests that the phenomenon should have begun with the most basic words, and these rarely are derivatives. In fact, this is perfectly consistent with the results of the comparisons of stems made in 3.2.4. Thus, the chicken and egg problems of the *gorjoχsullug* kind are effectively eliminated. Gaps in lexicography are also becoming less vexing as the common origin of these basic words is typically well established. Complete understanding of the mutual relations between the Turkic languages in the remote past is not a *sine qua non* here, too, as it is of secondary importance whether a reduplication which is today shared by two languages was formed in a single proto-language, or simultaneously in both proto-languages tied in a geographical league. In fact, it is not even clear that the two cases are universally distinguishable.

The groups resulting from such comparisons are small, possibly smaller than the actual proto-stock of reduplications, but they are more certain. A full genealogical tree cannot be reconstructed at present, but certain conclusions on the history of reduplication can still be drawn.

3.4.3 Altaic background

Reduplications are present in both the Mongolic and the Tungusic, but to the best of my knowledge, extensive, collective descriptions are missing. The readily available data are rather unsystematic, which prevents a detailed comparison. Nevertheless, the overall similarity seems sufficient to assume that the phenomenon dates back to the Altaic community. Below, only very brief overviews will be sketched. The information on Mongolic has been extracted from Benzing 1985: 143, Bese 1960, Hujiltu 1998, Khabtagaeva 2001: 85f, 96, Kotvič 1929: 97, MoLangs: 12, 135, 269, 332, 352, Poppe 1951: 45, and Svantesson et al. 2005: 58f, and that on Tungusic from Hujiltu 1998, Li/Whaley 2000, and Tsumagari 1997 and 2009. Unlike with the Turkic material, examples will be given in the original transcription of the source, but mostly omitting morpheme-delimitation symbols as they could be misleading in the changed context.

C-type reduplication is a Common Mongolic method (Janhunen 2003: 12). The formation is essentially the same as in Turkic, i.e. the initial mora is duplicated and prepended to the base with a closing consonant in between. Typically, the closing consonant is *b*, rarer *p* or *w*,⁵⁵ but occasionally other consonants are also possible,⁵⁶ or indeed no consonant at all in Monguor and Santa.⁵⁷ Sometimes, an emphatic lengthening of the reduplicated vowel can be observed,⁵⁸ and according to Mönggöngerel [after Svantesson et al. 2008: 59], the reduplicated vowel is not shortened in Naiman (*xiip xiit^hən* ‘very cold’). In some cases, the transparency of the formation has been lost, e.g. **sira* ‘yellow’ → **sibirakan* has yielded in Bonan the pair *shera* : *shewrexang*, while **xubxulaxan* » Bon.dial. *howlang* ‘very red’. Reduplication can be combined with the comparative degree, e.g. *Xlx. æb ær.xɔn* ‘very near’. I understand that the phenomenon is generally no longer productive in the Mongolic languages; Svantesson et al. 2005: 58 estimate the number of examples in Khaklha at about twenty or thirty.

But Mongolic reduplications do not end on the C-type. Bese 1960 and Khabtagaeva 2001 quote examples in *d*,⁵⁹ *do*,⁶⁰ *ra*, *rê*, *ro*, *ri*,⁶¹ and *la*, *lê*,⁶² and also the entire initial syllable can be reduplicated without any modification or insertion at all.⁶³ In addition,

⁵⁵ **CMo.** **xubxulaxan* ‘reddish, quite red’, *sibirakan* ‘quite yellow’; **Bon.** *fuwfulang* ‘very red’, *shewshera* ‘quite yellow’; **Dag.** *xabxar* ‘coal-black’; **Khor.** [хур хуу^hэн]; **Klmk.** *hav-havšun* ‘very quick’, *көб көркүкү* ‘very beautiful’, *ob-oln* ‘very many’, *зөб зөвдөр* ‘very well’; **San.** *hupulaghang* ‘very red’; **SYug.** *abaruun* ‘very clean’; **Xlx.** *dub dugui* ‘dead silent’, *нэw нэгэw* ‘very blue’, *pow poorəŋxi* ‘very round’, *xab xara* ‘completely black’, *xub xufa*: ‘very grey’. See also *Teb Tenggeri* in the *Secret History*, and a comment on it in Erdal 1991: fn. 82.

⁵⁶ **Dag.** *cimcigaang* ‘snow-white’, **Klmk.** *bim* ~ *biŋ* ~ *bis bitü* ‘completely closed’, *bas batu* ‘very firmly’, **Mgh.** *ufulān* ‘very red’, and see fn. 59 for *d*.

⁵⁷ **Mng.** *mu mula*: ‘very small’, *tçi tçigan* ‘very white’, **San.** *qa qaluŋ* ‘very hot’, *džo dzolian* ‘very soft’. From a synchronic typological perspective, this is in fact a separate type of reduplication. It is not clear, however, whether these forms arose from a simplification of the C-type, or whether it is perhaps the C-type that emerged from a modification of this Ø-type. See 3.4.4 on the origins of reduplication.

⁵⁸ **Dag.** *xubxulaang* ~ *xuubxulaang* ‘deep red’, maybe also **Klmk.** *cāb cāxḥ* ‘very good’ (Kotvič 1929: 97).

⁵⁹ **Bur.** *bad balaj* ‘pitch dark’, *bod boro* ‘completely grey’, *mad malān* ‘completely bald’, *mad mayā* ‘very bandy of legs’). It is not clear to me why both Bese 1960 and Khabtagaeva 2001 treat this closer as an alternative version of *do* below rather than of the standard *b*.

⁶⁰ **Bur.** *godo godogor* ‘strictly upright’, *šodo šodogor* ‘very thin’.

⁶¹ **Bur.** *ara arbagar* ‘very shaggy’, *boro borxigor* ‘очень невзрачный (старый дом)’, *tere tesxegeer* ‘very fat’. Interestingly, this closer is apparently only used with words with the -GAr suffix. Also, although non-C-type reduplications are essentially not discussed in the present work, it should be mentioned that a considerable number of similar forms can also be found in Yakut, e.g. *beri beriñex* ‘bountiful’, *čuru čulbugur* ‘very spiky’, *doru dostoj* ‘completely in vain’ (Pekarskij 1907–30, Korkina/Ubrjatova et al. 1982).

⁶² **Bur.** *bal bambaachaj* ‘very densely haired’, *bûlê bûlchêgêr* ‘very lofty’, *dala dalbagar* ‘sehr breitrandig vom Hut’. See also **Tuv.** *kap-la kara* in 2.18.4.

⁶³ **Bur.** *ar arbagar* ‘stark, gesträubt’, *têr têrchêgêr* ‘very thick’, **Ord.** *on onbôn* ‘completely different’.

at least three types of full reduplication are used to form plural-like and collective derivatives, and also for intensification.⁶⁴

The overall picture is not only interesting in itself, but also because of how it harmonizes with the Turkic. In 3.2.2, it was mentioned that two parameters, the number of reduplications and the entropy of closing consonants, suffice to distinguish the different types of collections observed in the Turkic languages. One more, however, will be necessary to do justice to the Mongolic, and preferably, it should be the diversification of types. The South Siberian languages have typically very few reduplications, and all are of the *C*-type, and closed by *p*. Yakut, on the other hand, has large collections of several types, and as many as ten closing consonants. The Mongolic languages, it seems, have relatively small collections, but of many types, and closed by several consonants. Typologically, this locates them precisely between the two Turkic groups, which is just where they are on the map, and thus adds another string to the bow of the dialect continuum interpretation of the Altaic (see 3.4.5).

The Tungusic languages could be expected to be similar to the Turkic and Mongolic with regard to reduplications, but they prove to be anything but. Although several examples are attested for Kilen, Orochen, Sibó, and Solon,⁶⁵ grammars typically do not mention reduplications at all.⁶⁶ An indirect explanation of this incompatibility is provided by Tsumagari 1997 and Li/Whaley 2000 who believe that the phenomenon has “obviously originated in Mongolian (and further in Turkic)” (Tsumagari 1997: 180), and, I understand, never truly flourished being a foreign element in Tungusic. Tsumagari 1997 does not explain the reasons behind his conviction, but Li/Whaley 2000 offer an entire series of arguments, which can be summarized as follows: **1a.** reduplication is found in “an extremely suspicious geographical distribution” in Kilen, Orochen, Sibó, and Solon, which “are associated by the fact that they have been in contact with one another and in contact with the same non-Tungusic languages, especially Mandarin Chinese, Khalkha Mongolian, and Dagur”, while at the same time,

⁶⁴ Plural/collective: MMo. *balaqat balaqat* ‘one city after the other’, Xlx. *arix-marix* ‘all kinds of alcoholic beverages’; intensification: Bur. *meliger-müliger* ‘very smooth’, Klmk. *salwr-sulwr* ‘in Unordnung’.

⁶⁵ **Kilen** *tob tondo* ‘very straight’, *ub ujan* ‘quite watery’; **Orochen** in **b**: *kab kara* ‘very black’, *kob kəŋərɪn* ‘very dark’, *ʃɪb ʃɪŋərɪn* ‘very yellow’, **other**: *bag bagdarɪn* ‘snow-white’; **Sibó** in **b**: *tab tarxun* ‘very fat’, *xab xalxun* ‘very hot’, in **v**: *gov-gocɣun* ‘extremely spicy’, *gov golmin* ‘very long’, *nav narxun* ‘very thin’, and **other**: *faq farxun* ‘very dark’; **Solon** in **b**: *ab aya* ‘very good’, *xob xonnorin* ‘coal-black’, and **other**: *nem nemikkün* ‘very thin’, see also below. Apart from the above, several more examples of a rather different build are attested in Solon: *siŋarin* ‘yellow’ → *siŋa siŋaljaxun* ‘very yellow’, *ularin* ‘red’ → *ula ulariljaxun* ‘quite red’.

⁶⁶ **General**: Benzing 1956, Skorik et al. 1968; **Even**: Malchukov 1995; **Evenki**: Konstantinova 1964, Nedjalkov 1997, Vasilevič 1940; **Kili**: Sunik 1958; **Manchu**: Avrorin 2000, Gorelova 2002; **Nanai**: Avrorin 1959–61, Ko/Yurn 2011 (the Naykhin dialect); **Negidal**: Cincius 1982; **Orok**: Petrova 1967; **Udeghe**: Girfanova 2002; **Ulch**: Petrova 1936, Sunik 1985. Surprisingly, even Sem 1976 (**Kilen**) does not seem to make note of reduplication.

1b. they represent the Southern, Central and Northern branches of the Tungusic, and thus “do not form anything close to a genetic unity” (Li/Whaley 2000: 363); further on, reduplication is structurally incompatible with the Tungusic languages because Orochen – which Li and Whaley consider to be representative – is **2.** exclusively suffixing, **3.** has “a strong propensity for syllables to be CV, CV:, or CVS (where S is a sonorant)”, and **4.** does not have the sequences [gb] or [bj] / [pj] other than in *bagbagdarin* and *fibfinarin* (Li/Whaley 2000: 363); and lastly, **5.** Orochen has fewer closing consonants than “one finds, say, in Turkish”, and since “it is generally agreed upon that the choice of consonant found in C2 must be lexically determined”, and “irregularities in a morphological system are generally taken to represent historical residue rather than innovations, we should expect greater irregularity to more closely approximate the original system” (Li/Whaley 2000: 364).

This reasoning is not immediately convincing. As for the first two arguments, it seems that Li and Whaley see their power in the conjunction of **1a.** (geography) and **1b.** (genealogy), rather than in either of them alone. As for 1a., also Tsumagari (1997: 181 and elsewhere) mentions that reduplication is limited to the Tungusic languages spoken in China. However, proximity to the Mongolic is in no way unique to the four languages that have reduplication, as can be seen in map 3.12. Should reduplication be indeed an areal feature, at least Evenki and Manchu would have to be expected to bear traits of it as well and, in fact, Sibó would have to be expected not to, since it is spoken about 3000 km away from the other three.⁶⁷ The conjunction of 1a. and 1b., therefore, does not quite hold. And on its own, 1b. (genealogical diversity) points in fact to inheritance rather than influence, and especially so, that Li and Whaley’s assertion (p. 370, fn. 1) that reduplication does not exist in Korean is in fact inaccurate, see e.g. *kol-kolu* ‘evenly divided’, *t’ek-t’ekul* ‘rolling’, *tu-tuŋsil* ‘floating’ (Kim 2003, 2009). Argument number **2.** (suffixing) is also not in any way characteristic of Tungusic; the Turkic are not any more prefixing than the Tungusic, and yet the nativeness of their reduplications seems never to be challenged. As for **3.** and **4.** (syllables and sequences), I do not feel competent to comment on them, and will limit myself to rejecting one auxiliary subargument for 3. In order to emphasize the phonological difference between the Tungusic, where reduplications are scarce, and Turkic, where they are abundant, Li and Whaley adduce an observation after Dobrovolsky 1987, that unlike the Tungusic ones, “Turkish monosyllables are always CVC” (p. 363). This is obviously nonsense, see e.g. Tksh. *o* ‘he/she/it’, *ak* ‘white’, *da* ‘also’. This says nothing of the Tungusic but, again, undermines the uniqueness of the unfavourable conditions that they supposedly offer for reduplication. Finally, Li and Whaley themselves acknowledge that argument number **5.** (diversity of closing

⁶⁷ Tsumagari 1997: 180 adduces Ma. *tob tondo* ‘honest, fair’, but immediately explains it away as an accidental combination of otherwise independent *tob* ‘straight’ + *tondo* ‘straight, fair’. He also notes that lack of reduplication in Written Manchu cannot be simply extrapolated onto Spoken Manchu, as “this kind of emphatic expression might be so colloquial that it is not attested in Written Manchu” (p. 180). Without actual proof, however, this remark cannot be treated as an argument.

consonants) is not a strong one. The present work does not make it possible to determine whether the many closing consonants found in certain Turkic languages are secondary, but it does suggest that at least a part of them are, which is against Li and Whaley's reasoning. In any case, their observation cannot be used as a serious argument, especially if based on just a rather shallow comparison of modern Orochen and modern Turkish alone, and at the same time supposed to produce a judgement about the Altaic stage.

Overall, the Tungusic collections of reduplications appear to be vestigial at best, which correlates well with the large gap between the Dolgan and Yakut collections (see 2.4). Whether the few Tungusic examples have been inherited or borrowed cannot be considered finally settled; out of Li and Whaley's six arguments in favour of influence, at best two remain valid under closer scrutiny, and not even their conjunction can be seen as decisive.

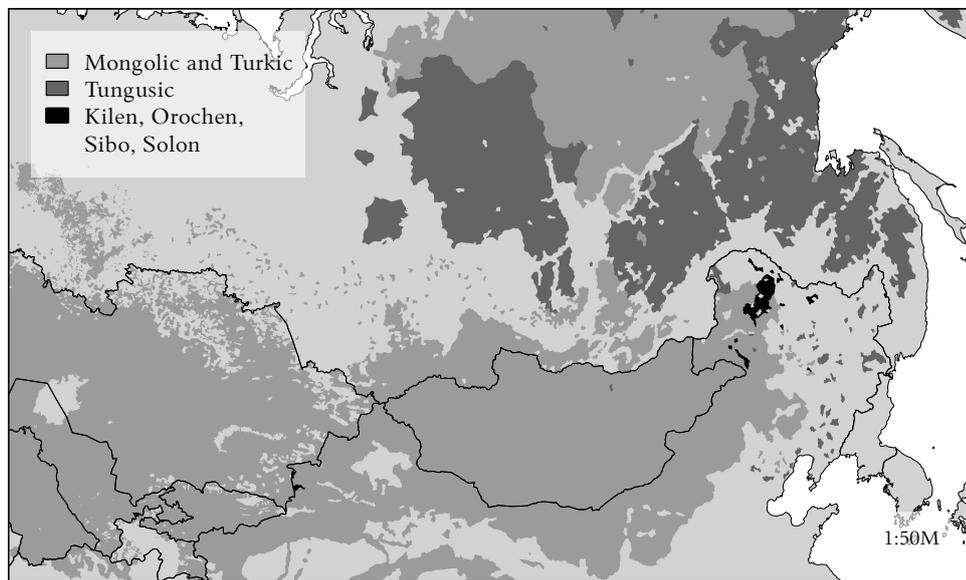


Figure 3.12: Geographical distribution of reduplications in the Tungusic languages. Kilen, Orochen, Sibó, and Solon are the four Tungusic languages with attested reduplications. All locations are after *Атлас народов мира* (Bрук/Apenčenko 1964), except for Kilen and Solon which are after the map in Doerfer/Weiers 1985; the large territories of Even and Evenki are also confirmed by Sablin/Savelyeva 2011.

Based on the brief summary here, it is of course not possible to determine whether reduplications arose when the Altaic peoples still formed one community, as for this a detailed comparison of all the specific collections in all the branches is necessary. Nevertheless, the picture seems to be sufficiently clear to at least provisionally allow for considerations about the origins of the phenomenon at the Turco-Mongolic level, see 3.4.4.

3.4.4 Origins

The Turkic languages are in general quite decisively suffixing, and reduplication is effectively the only process in them that occurs to the front of the stem (if one does not count nominal compounds such as *bašbakan* ‘prime minister’, see Deny 1938 for more examples). This invites suppositions about the very origins of the phenomenon. At present, however, very little can be said with certitude. More likely than not, reduplication did not start after the disintegration of the Altaic community (see 3.4.3), the first closing consonant was *p*, and the first reduplicated stems were colour names (3.2.4). It seems that only two scenarios have been proposed so far. They are presented below, briefly discussed, and supplemented with a third possibility and some additional musings.

Perhaps the most obvious supposition would be that partial reduplication evolved from contracted full reduplication, i.e. **kara kara* lit. ‘black black’ → **karɔkara* → **karkara* → **kapkara*.⁶⁸ Whether one believes that this change would have happened to one or to more phrases, the development of the phenomenon past its initial phase would still be more likely attributed to analogy, i.e. by assuming that **kapkara*, or whichever the first reduplicated stem or stems was, had been misinterpreted, and thus gave impulse for the creation of new formations like it. This scenario assumes that full reduplications were rather common, sufficiently so to undergo this kind of contraction. This is possible but unprovable. Also the dissimilation can be seen as a weak link, and it should be noted that this scheme cannot account for the shortening of the reduplicated vowel (see 3.1.20).⁶⁹

Another scenario derives reduplications from formations of the **kara mu kara* type, where **mu* is “the emphatic enclitic” (Menges 1959: 467, 1995: 116).⁷⁰ Interestingly, Menges’s wording suggests that this transmutation occurred separately for each reduplication.⁷¹ Müller 2004: 316–319 analyses this proposition quite thoroughly, and is right to eventually refute it.

⁶⁸ In particular, Kim 2009 supports this option. His, however, is an account that concentrates primarily on the theoretical aspect of the issue.

⁶⁹ Perhaps, the only way to incorporate shortening into it would be to assume that all the long vowels in the initial syllables of the earliest reduplicated stems are secondary and arose through compensatory lengthening, so that **sagryg sagryg* lit. ‘yellow yellow’ → **sagrɔsagryg* → **sapsagryg* → **sapsāryg*. It seems that this postulate only makes the scenario even less plausible.

⁷⁰ According to Imart 1981: §1110, a similar proposition has been put forward by M.L. Bazin. Unfortunately, the exact source is not specified. See 2.10.1.

⁷¹ The adnominal nouns have an augmentative, [...] by which the word is originally doubled, [...] and connected by the emphatic enclitic *-mu/-mü* or *-ma/mä* to the first noun. The first noun is generally reduced by the effect of accentuation [...] In many cases, *-mu* undergoes occlusive assimilation so that *qap-qara* [...], *sap-saryγ* [...] arose from **qa-m’-qara* < *qara-mu-qara*, **sa-m’-saryγ* < *saryγ-mu-saryγ* etc. (Menges 1995: 116)

A more plausible version of Menges’s idea can be imagined, where the phonetic change happens once, e.g. for **kara mu kara* > **karamkara* > **kamkara* > *kapkara*, and based on this one form other reduplications are shaped through analogy, and differentiation of the closing consonants only occurs much later. This frees the scenario from the inconvenient obligation to explain why “occlusive assimilation” should happen in one stems and not in others, without a clear phonetic reason. Still, a chain of transformations needs to be assumed, which contains one particularly non-trivial link (**karamkara* > **kamkara*), and has in addition no way to explain the shortening of the reduplicated vowel. The overall plausibility of this scheme seems quite low.

If this scenario can be saved in any way, it is perhaps only by limiting it to reduplications closed by the double *pp*. Roos 2000: 79 suggests this solution for WYug. *appaq* ‘very white’. However, it seems more probable that the Yugur word is related to the general Tkc. *appak* id., than an independent innovation, and that is more likely to have resulted from emphatic lengthening than anything else (see 3.1.8).

One more scheme can be proposed. In 3.3.4, it was mentioned that al-Kāšyarī understood reduplicated anlauts to be independent particles with limited semantics or connectivity, and in 1.1.2 we saw that this interpretation has survived well into the 20th century. One particular combination of a particle and an adjective is special because of its phonetics. OTkc. *köp* i.a. ‘many, much, very’ + *kök* i.a. ‘blue’ (Clauson 1972: 686, 708, DTS) is equally or more likely to have existed than **kara kara* or **kara mu kara*, and for this scheme to hold, it need not have been particularly common. Given its particular sounding, it may be assumed that it was mis- or reinterpreted, and new analogous forms based on it. Such scenario does not involve any phonetic changes at all, and accounts for the shortening of the reduplicated vowel, which is unlike the propositions above.

Also unlike the propositions above, it is explicitly limited to one stem, from which the entire phenomenon would have to have evolved. Unfortunately, this stem is not *kara*, which is the only one common to all the Turkic languages with large sources, and not even *kyzyl* or *sary* which are the second most common ones (see tab. 3.9). In genealogical classifications, *kök* does not appear near the root of the tree (3.4.5). However, of the languages compared there, it is only missing from Karakalpak and Kazakh, and only partially from from Khakas (**köpkök* does not seem to be attested, but *köppeges* is, and was only excluded from the comparison because its exact etymology is not clear, see 2.9.3). The picture would be quite different if one were to assume that this particular reduplication had gone out of use in the Kazakh-Karakalpak territory at some later period, or that it has been simply overlooked by lexicographers. Neither of these seems to be exceptionally plausible.

Note that this scenario effectively assigns the creation of reduplication to the Turkic languages, and excludes the Mongolic. As far as Tkc. **kök* is well known to correspond to Mo. *köke* (see e.g. Stachowski K. 2012a (W. Kotwicz’s material): 295), *köp* appears

to only have existed in the former (see TMEN III: 616, No. 1654, where Mongolic is never mentioned, and Doerfer 1982: 105, where Tkc. *köp* ‘excessive, much’ is considered to have originally been equal to *köp-* ‘to swell, to froth over’, and Mo. *köge-sün* ‘froth’ is mentioned without drawing conclusions, but it can be seen that they could only refer to a period prior to that of **köp*). Interestingly, the passing of the first seed of reduplication from the Turkic to the Mongolic, and maybe also to the Tungusic, will be interpreted differently depending on whether the two or three groups were separate at the time, or formed a close community or a continuum. In the first case, it would have to be declared a borrowing, and in the second, common inheritance within the wave model. The scheme is thus independent of which perspective on the Altaic group is adopted.

Be that as it may, one must also mention the possibility that the relation between *köp* and *köpkök* is in fact the reverse. In 3.1.10, several examples of particles were listed, that might have arisen from severed and emancipated reduplicated anlauts. At least in the case of Kirg. *kyp žyłaŋaç* ‘stark naked’ < *kypkyzyl žyłaŋaç*, this is a rather likely explanation. In theory, the same could be true of *köp*.

One common problem of the three schemes above is accent. In general, the Turkic languages have dynamic stress on the final syllable; reduplications with their initial accent are one of the few exceptions. As for the Mongolic languages, the opinions are divided on the place of accentuation (Svantesson et al. 2005: 94f). The issue has not been finally resolved for the Proto-Altaic stage. N. Poppe proposed a coexistence of an initial dynamic stress and a hypothetical mobile pitch accent (VGAS: 146), but his idea is unprovable for the time being. However, a steadfast reconstruction of Proto-Altaic stress is perhaps not necessary for our purpose. It is only natural that words and morphemes which are meant to be emphasized will be pronounced in a slightly overacted way. In a language with fixed accent, such as Polish, this can lead to a shift of stress to the first syllable (e.g. *abso'lutnie* ‘absolutely’ → '*absolutnie*). Regardless of where the Proto-Altaic accent fell in non-emphatic pronunciation – if indeed it was just one place throughout the entire territory the Proto-Altaic community occupied – it well might have shifted towards the first syllable due to emphasis alone. Naturally, this supposition cannot be proved, but I believe it sufficiently plausible as a working hypothesis to allow the research not to stall over the issue.

A remark should be also made about the different closing consonants. It seems to be universally agreed upon that *p* is the primary one, and all the other ones appear to be implicitly assumed to derive from it. This might be true for *f*, *v*, *w*, and perhaps *m*, but will be more difficult to demonstrate for *r*, *s*, and the rarer *š*, *t*, &c. Müller’s 2004 “Kontrast-These” cannot account for the general Turkic differentiation.

Perhaps, then, reduplication began with more than one closing consonant, and the uniformity we observe today in the majority of Turkic languages is a result of later working of analogy? This would be in line with the use of *m*, *s*, and other consonants as closers

in the Mongolic and Tungusic languages (see 3.4.3). If true, this supposition could undermine all the three scenarios discussed here, but as a matter of fact, it could also lead to the conclusion that perhaps even all three are partly true as they combined to create several forms which were then reinterpreted to give rise to the phenomenon of reduplication. One particularly intriguing argument for early diversification of reduplications is the existence of CVC-type reduplications in Dolgan and Yakut (e.g. *to.bus.toloru* ‘absolutely full’) and, it appears, nowhere else save for one single example in Gagauz, *je.piz.jeni* ‘brand-new’. The question probably cannot be answered without a comprehensive analysis of the Mongolic collections, and that is yet to be carried out. See also 3.4.6.

With little data to harness the imagination, scenarios can be multiplied much like Russell’s teapots. The scheme proposed here (**köp kōk*) has some advantages over the two put forward before (**kara kara*, **kara mu kara*), but it is not free of weaknesses, either, and for the time being all three remain merely unprovable ideas. Directly linked to this issue is the question of whether the differentiation of closing consonants is a primary or secondary phenomenon, and this also cannot be settled, it seems, at the present stage of research.

3.4.5 Inheritance vs. influence

It was already mentioned on several occasions that secondary areal influence might have played a greater role in the evolution of reduplication than genealogical inheritance (see 3.2.7). It is not entirely clear how this supposition might be tested. Identification of common reduplicated stems shared by various languages does not in itself prove their common descent in the genealogical sense (see 3.4.2). However, given the absence of continuous, long-standing and exhaustive historical records, it appears to be the best method at our disposal. Fortunately, clarity of the overall picture seems sufficient to compensate for the shortcomings of the method.

Many classifications have been proposed for the Turkic languages. Here, only two will be used, Baskakov 1960 and Tekin 1990, because it will quickly become evident that the eventual conclusions regarding the evolution of reduplication must be essentially the same regardless of which particular one would have been chosen.

Below, reduplicated stems will be fitted onto genealogical trees. The fitting will be from the leaves to the root, i.e. first e.g. the stems common to Azeri, Turkish and Turkmen will be identified, and assigned to the Oghuz level, then the same will be done for the Kipchak languages, and the intersection of the two resulting collections will be considered to be the stock at the Kipchak-Oghuz level. In order to avoid excessive reduction of the results, only the languages with relatively large sources will be taken into account, just as was done and explained in 3.2.4. Nodes which only contain idiolects not covered in the present work, will be omitted. If they were included, the final picture would either remain the same, or contain even less examples in the intermediate nodes.

Baskakov 1960

The classification composed by N.A. Baskakov is quite different, but also more detailed and more explicitly genealogical than most of the others. Its full complexity, however, will not need to be exploited here; the fragment relevant for us is presented in fig. 3.13.

Perhaps the most striking observation to be made, is that the great majority of groups (nodes) have very few reduplications. When only pairs of languages are compared, as in tab. 3.7, the intersections tend to be much more numerous. This phenomenon was already made note of apropos the Kipchak languages in 3.2.4, where the number of common stems decreased very rapidly with the addition of new languages to the comparison, but no single culprit for the ultimately very small collection could be identified. Apparently, the number of common stems in each group here correlates

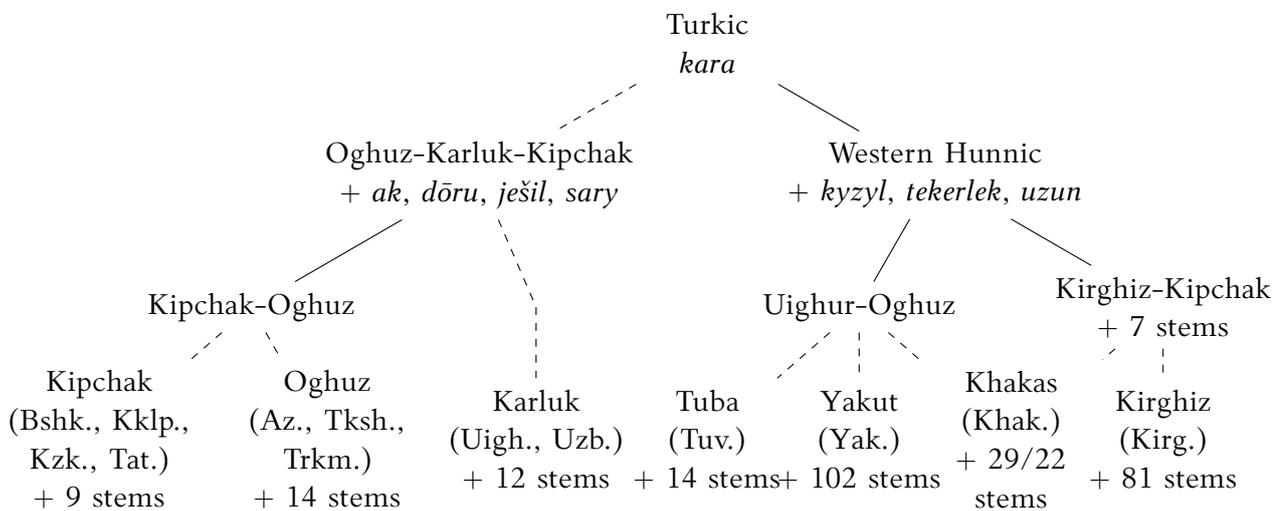


Figure 3.13: Reduplications against a fragment of N.A. Baskakov's 1960 classification of the Turkic languages. Dashed lines denote removed intermediate nodes (e.g. Kipchak-Polovets, Oghuz-Seljuk). Completely omitted branches are not marked (e.g. Chuvash, Khazar). In the leaves, only the languages used in the comparison are listed. The division into chronological epochs has not been preserved. The plus symbol always refers to the immediately preceding node; nodes without comment simply inherit the state from their predecessors. The Khakas group is marked as descending from two groups simultaneously in Baskakov 1960; here, 29 is with regard to the Uighur-Oghuz group, and 22 to the Kirghiz-Kipchak group. The Oghuz group is based on a set of 19 stems, i.e. including the six with not uniform closing consonants (see 3.2.4). This is to maximize the number of matches with the other groups.

not only with the size of the collections in the languages that eventually descend from it, but also with the number of languages compared. This is most visible in the Kipchak and Oghuz groups, which generally boast sizeable stocks, but only have a handful of stems in common.

At this point, the accuracy of identification of stems in the present work might be questioned. The procedure was quite conservative, and in case of doubt false negatives were preferred to false positives. Note also that what is taken into account are stems rather than roots, and this further reduces the possible number of matches, even if, it appears, not by much. Nonetheless, the Kipchak-Oghuz group has merely five reduplications, while the seven languages with large sources that descend from it, have an average of about 88.5.

Overall, it would seem that the languages in the Kipchak group inherited, on average, merely about 6.7% of their reduplication, those in the Oghuz group 6.5%, and those in the Karluk group 12.7%. The remaining groups only contain one language each, and it is clear from fig. 3.13 that the numbers are higher for the South Siberian languages, but even lower for Yakut. Naturally, these calculations must be taken with a grain of salt; see below.

One more observations needs to be made, namely that the few stems common to higher genealogical units are not all colour names. This is somewhat against the supposition that could be made based on our previous considerations. One possible explanation is that the process of semantic diversification had started rather early on. Another is that these few stems are not shared by several languages because each has inherited them, but rather because they all have developed their respective collections of reduplications following a quite similar semantic plan (see 3.3.5). Presently, there does not seem to exist a way to determine what the real cause of this situation is.

Tekin 1990

The classification composed by T. Tekin is closer to the line started by A.N. Samojlovič in 1922, and later continued by the majority of propositions. Like them, it is not explicitly genealogical in nature. The principle of division is phonetics, which suggests that genealogy is indeed what is meant, but intermediate proto-languages are not discussed. The fragment relevant for us is presented in fig. 3.14.

Despite the many differences between Baskakov's and Tekin's classifications, the effects of fitting of reduplications to them are rather similar. Again, the collections at intermediate nodes are strikingly humble and entirely out of proportion with the collections of the languages ultimately descending from them. The numbers in the leaves are quite alike, which is of course due to the fact that the groups themselves aggregate almost the same languages in both models. Thus, the same extends to the calculations of the inherited-to-“influenced” ratio.

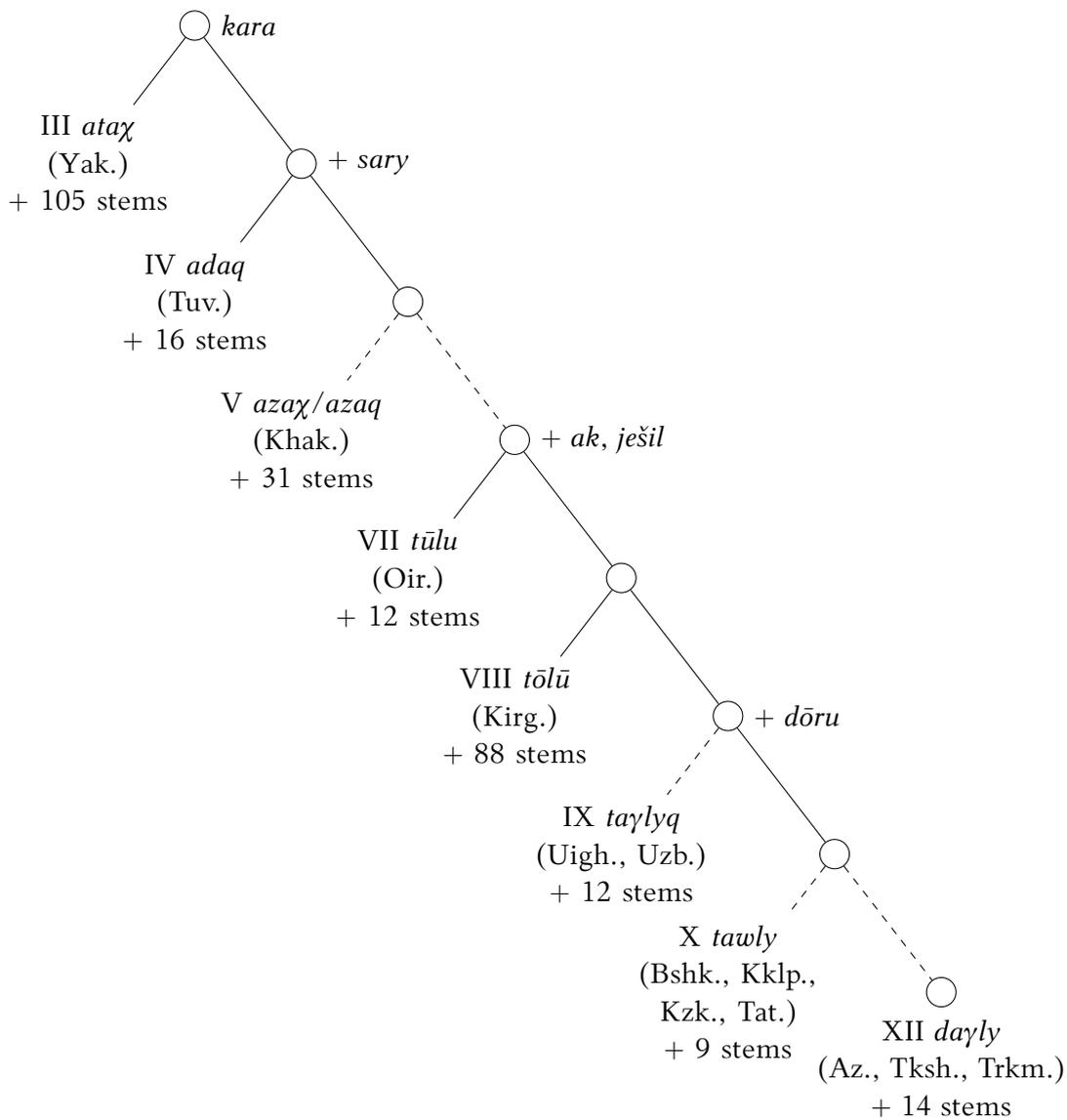


Figure 3.14: Reduplications against a fragment of T. Tekin's 1990 classification of the Turkic languages. Dashed lines denote removed intermediate nodes (e.g. the *taylyq* group (several South Siberian dialects), or the *tayly* group (Salar)). Completely omitted branches are not marked (the *r-* and *l-* group (literary Chuvash), and the *hadaq* group (Khalaj)). In the leaves, only the languages used in the comparison are listed. The plus symbol always refers to the immediately preceding node; nodes without comment simply inherit the state from their predecessors. The Oghuz group is based on a set of 19 stems, i.e. including the six with not uniform closing consonants (see 3.2.4). This is to maximize the number of matches with the other groups.

Perhaps the only more significant difference between how reduplications fit into the two classifications, is semantics. The first stem that is not a colour name appears at the same stage in both, Oghuz-Karluk-Kipchak in Baskakov's terms, but in Tekin's model it is much further down. Possibly, however, no conclusions can be made based on this occurrence. The genealogical nature of Tekin's classification is alleged by me but not, in fact, professed by the author. It is also not at all clear to what degree the commonness of a reduplicated stem implies inheritance. Finally, the very feasibility of the genealogical model to the Turkic languages is disputable (see below). For the moment, this observation will be dropped.

In 3.2.4, reduplicated stems in the five traditionally distinguished genealogical groups were compared, and the conclusion drawn that the number of stems common to more than a handful of languages is quite low indeed. It was shown that even rich collections shrink very rapidly as new languages are added to the comparison. But aggregating multiple languages into a single unit is precisely what genealogical classifications do. It appears that reduplications simply cannot fit into the clear-cut, hierarchical picture painted by genealogical classifications, and the two experiments above attest to this. Before making conclusions, however, let us conduct another one, this time not taking genealogy into account at all.

Clustering

Baskakov's classification is essentially based on all available kinds of material, Tekin's only on phonetics, and neither makes particular use of reduplications. Fig. 3.15 shows one possible classification that is only based on reduplicated stems, and takes neither phonetics, nor morphology into account, nor any other part of prior linguistic research that is not directly related to reduplications.

The resultig scenario is quite interesting. Foreseeably, the collections at intermediate nodes hold many more reduplications than was the case in the two classifications above. It might be also noticed that stems not related to colours appear quite early here, but again, the significance of this fact is dubious. The most striking feature of this classification, however, is perhaps how well it correlates with genealogy. But it also corresponds to geography, and in light of what is known about the history of the Turkic peoples, it seems impossible at the moment to differentiate between the two.

Finally, a warning must be made not to treat this scheme too literally. Finer and more robust methods could be used to minimize the effect of imperfect sources (see in particular Nerbonne/Kleiweg/Manni/Heeringa 2008), but underlying any eventual classification is the shaky assumption that existence of reduplications of a stem in two languages implies a genealogical relation between them (see 3.4.2).

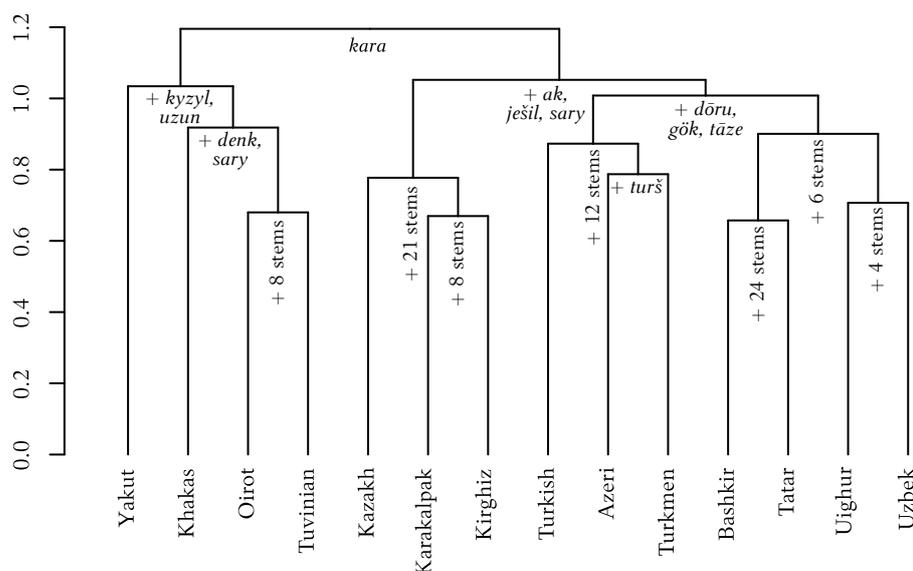


Figure 3.15: Hierarchical clustering of Turkic reduplicated stems (Jaccard distance, Ward method).⁷² Only the languages with large sources have been compared (see the main text). The plus symbol always refers to the immediately preceding node; nodes without comment simply inherit the state from their predecessors.

⁷² Hierarchical clustering, like multidimensional scaling, is derived from a distance matrix. Here, however, Jaccard distance is used ($1 - \frac{|A \cap B|}{|A \cup B|}$) instead of Euclidean distance ($\sqrt{|AB| + |BA|}$) employed in fig. 3.6 and 3.9. Let us note that the former ranges between 0 and 1, and represents the relative dissimilarity between two sets, while the latter has no upper limit and yields the absolute distance between them. As a result, Jaccard distance is not particularly sensitive to the cardinalities of the analysed sets, but Euclidean distance is very much so. The point of MDS as used before, was to expose the different paths that evolution of reduplication has taken in various languages, and Euclidean distance served this purpose well because it highlights the outliers. Used in hierarchical clustering, however, it returns what is effectively the ordering of languages by the number of reduplications. On the other hand, Jaccard distance always compares sets in their own weight class, so to say, and therefore yields a clustering based on actual relative similarity, and not just on cardinality. But it needs to be noted that there are actually tens of indices with similar properties. Here, Jaccard's was chosen for its simplicity, which I believe to be an important quality when analyzing data whose accuracy hinges entirely on imperfect sources.

Conclusions

The material collected in the present work clearly does not fit very well into the traditional genealogical classifications of the Turkic languages. The match can be improved if genealogy is slightly bent, but the resulting scheme still does not do justice to the richness of connections between many languages, as seen in tab. 3.7.

Instead, reduplications apparently might constitute one of the many arguments that are necessary in order to shift the understanding of the evolution of the Turkic languages from the genealogical model to one that is based on dialect continua; see Gadžieva 1975 for a collection of many more. The idea that the Altaic languages are related by contact rather than inheritance was first put forward by W. Kotwicz (see Stachowski M. 2001: fn. 10). Later, the same idea was applied to newer stages of development inside the Turkic family, perhaps most explicitly and recently in Schönig 1997–98. Also the oldest Turkic written sources do not paint the picture of a single language, and from later history we know that at least in Central Asia in a broad sense, the Turkic peoples remained in contact that was sufficiently close to spawn and uphold Chaghatai. Genealogical inheritance cannot be realistically hoped to account for the dominant part of the characters of specific languages, and reduplications only confirm this. A more detailed quantitative estimation of this relation, nevertheless, cannot be performed based only on common reduplicated stems; what else needs to be investigated, and how, remains not quite clear for now.

See 3.4.3 for an overview of how the Mongolic and Tungusic languages fit into this interpretation, and 3.4.7 for a summary and continuation of these considerations. See also 4.1.3, where it is shown that also the distribution of closing consonants might be viewed as resulting more from secondary areal influence than from genealogical inheritance.

3.4.6 Evolution

The great majority of the Turkic languages lack long-standing, continuous historical records that could be definitely attributed to their direct genealogical ancestors. The possibilities of tracking the evolution of reduplications are thus severely limited.

The present work collects two sets of historical data, Ottoman (see 2.13), and the generally oldest attestations (2.1). Few conclusions can be drawn based on the first, and from the second hardly anything can be deduced at all, because most examples cannot be clearly attributed to an earlier stage of any specific modern language. (This is also the main reason why Chaghatai material has been altogether excluded from this work, see 1.2.1.) It would be also quite futile to hypothesize about history based on groups of those modern languages which are known to have developed from a single proto-language, as effectively the only such pair are Dolgan with Yakut, and their collections have almost nothing in common (see 3.2.4).

Let us attempt to gather what little can be inferred about the evolution of Turkic reduplications from a comparison of modern and historical material.

Turkish – Ottoman

The Ottoman data presented in this work are almost certainly incomplete (see 2.13.4). From gaps in Ottoman, therefore, nothing can be deduced. There are, however, reduplications attested in it that are apparently no more in modern Turkish, and also those where the closing consonant has changed.

The first group contains nine examples (see 2.13.2 for details): *apary* ‘very clean’ (attested in the 16th c.), *aphāzyr* ‘absolutely ready’ (17–18th c.), *büsbüyük* ‘very big’ (17th c.), *dosdolajınže* ‘all around’ (17th c.), *epeji* (19th c., but conserved as an independent word *epeji* &c. ‘quite, fairly’, see 3.1.15), *kapkarşy* ‘completely opposite’ (14th c.), *kypkyryk* ‘completely broken’ (18th c.), *sāmsāfi* ‘absolutely pure’ (18th c.), and *sypsyrlak* ‘completely smooth’ (15th c.).

Only one of them can be found in the Turkish dialects, *bösböj(j)ük* (DS; see 3.1.12 on emphatic lengthening), and except for the now-dialectal *syrlak*, all the bases remain in everyday use in the literary language. The nine reduplications do not share any particular structural or semantic features, and there generally seems to be no obvious reason why they would have ceased to be used.

In twelve examples, the closing consonant has changed between Ottoman and modern Turkish: *ak* ‘white’ ($p, pp \rightarrow p$), *čürük* ‘rotten’ ($p \rightarrow m$), *čyplak* ‘naked’ ($m, r \rightarrow r, s$), *dolu* ‘full’ ($p \rightarrow p, s$), *jassy* ‘flat’ ($p \rightarrow m, p$), *jaš* ‘wet’ ($m \rightarrow m, p$), *ješil* ‘green’ ($m \rightarrow m, p$), *katy* ‘hard, solid’, ($s \rightarrow p, s$), *sijāh* ‘black’ ($m \rightarrow m, p$), *tatly* ‘sweet, tasty’ ($m \rightarrow p$), *tāze* ‘fresh’ ($m \rightarrow m, p$), *temiz* ‘clean’ ($r \rightarrow p, r$).

A tendency to standardize can be seen relatively clearly here. In six cases, p was added as an alternative closing consonant, in one the old closing consonant was replaced with p , and in one the irregular pp was discarded leaving only p from the earlier pair of alternatives. Only four words had their reduplications changed not towards the dominating p , but it must be noted that one of them (*čyplak*) has p for C_2 , which is a strong indication against p as the closer (see 3.1.6), and two of them had already had p before, and only new closing consonants were made available for them, while the original p has been preserved as an alternative. Thus, the only word that has clearly gone away from standardization is Ott. *čüpčürük* ‘completely rotten’ (18th c.) > Tksh. *čümčürük* id.

In light of the incompleteness of Ottoman data, entropy of its closing consonants (3.2.1) can only be considered in approximation. Still, it is found to be higher than that of modern Turkish (0.535 : 0.465), which is in line with the standardization seen above.

Overall, the only radical difference between Ottoman and modern Turkish, is the number of attested examples. But this, as we saw in 2.16.4 and 2.13.4, has been largely inflated by insufficient sources on one side, and somewhat zealous research on the other. In the area of closing consonants, a limited amount of standardization can be seen.

One more observation needs to be made about the relation between Ottoman and modern Turkish. According to the so-called Piotrovskij-Altmann law, change in language can be

modelled with an equation which in our particular case would assume the shape $p(t) = \frac{c}{1+e^{-b(t-A)}}$, where p is the number of forms, t is time, and A , b , and C are coefficients (this is a slightly modified version that was proposed in Stachowski K. 2013 as a linguistically more meaningful formulation). The plot of this function is a sigmoid (has the shape of a horizontally stretched letter s), which is entirely consistent with linguistic intuition.

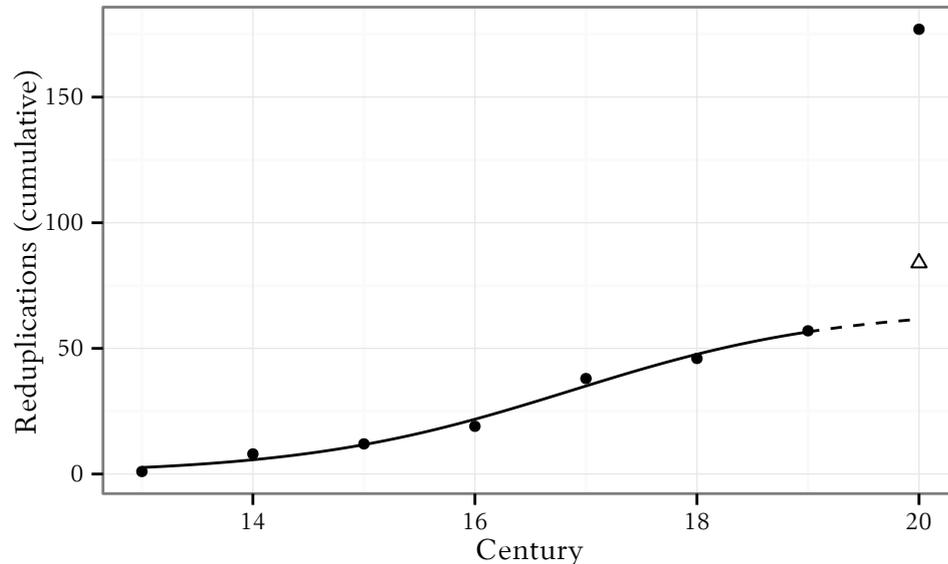


Figure 3.16: Cumulative sum of the number of new reduplications per century in Ottoman and Turkish. The coefficients for the fitted line (see the main text) are as follows: $A = 16.8561$, $b = 0.8285$, $C = 66.1431$. Beyond the 19th century, the line is dashed because it was only fitted for the Ottoman data. The hollow triangle marks the (very) approximate number of reduplications in common use across the Turkish territory, which is 84 (see 2.16.4). The dot at the 20th c., marks the 177 reduplications attested for modern Turkish.

The formula accounts for the Ottoman data quite well ($R^2 = 0.9897$; see fn. 54 for the meaning of this index), but it would be futile to attempt to fit it to both the Ottoman and Turkish data simultaneously (to connect all the dots in fig. 3.16 while maintaining the type of shape of the black line). It seems that the only way in which this could be done, would predict nothing short of an explosion in the 21st and perhaps the following centuries.

Apparently, only three basic explanations of this situation are possible: 1. the Piotrovskij-Altman law is incorrect; 2. the Ottoman data are severely incomplete; 3. the Turkish data are inflated. Or, some combination of the three. In light of the vast supporting material collected for the Piotrovskij-Altman law, the first option will be disregarded. The second possibility has already been supposed in 2.13.4. Fig. 3.16 visualizes the scale of omissions that would have to be assumed if the gap were to be explained by

the second option alone. This seems rather unlikely. The third possibility has also been supposed before, in 2.16.4. It was estimated that perhaps only about 84 reduplications are in actual widespread use across the entire Turkish territory, and this number is represented by a triangle in fig. 3.16.

Thus, it appears that both the second explanation (insufficient Ottoman) and the third one (excessive Turkish) are true. The latter might be responsible for a larger part of the gap, but the former is also definitely not insignificant.

The oldest attestations

It was mentioned in 2.1 that the oldest attested *C*-type reduplications belong to a number of just generally characterized dialects, but are representative of none. In sporadic cases, even such general attribution is missing. The majority of stems can be found together with their reduplications in modern languages, but it is certainly not possible to draw conclusions about those modern reduplications which do not seem to be attested in the oldest sources. In this situation, hardly anything may be deduced apart from *termini ante quos* of two phenomena.

All attestations are post-runic. It must be remembered, however, that being probably colloquial formations, reduplications were likely deemed unsuitable for the typically ceremonious inscriptions. In the 11th century, most notably in al-Kāšyārī's dictionary, more than ten reduplications appear. They are attributed to various Turkic tribes, not limited to colour names, and can be closed with three different closing consonants which show a genealogical and geographical differentiation (*m* in the Oghuz versus *p* elsewhere; *s* in *tästägirmä* 'very round' is puzzling for al-Kāšyārī). Combining these two pieces of information, it might be supposed that the 11th century is no more than the *terminus ante quem* of both, Turkic reduplication, and the diversification of closing consonants, and the phenomena had in reality begun earlier.

Just how much earlier, however, cannot be settled at the current stage of research. Reduplication is present in the Mongolic languages, several examples can be found in the Tungusic (see 3.4.3), and both show a certain degree of differentiation of closing consonants. A comprehensive analysis, however, is necessary to see whether the similarities go any deeper. A hypothetical scenario is presented in 3.4.7.

3.4.7 Conclusions

The goal of this section was to reconstruct what little can be deduced with acceptable certitude about the evolution of reduplication. It was shown that the correct approach is the historical-comparative one (see 3.4.1), but its application runs into serious obstacles (3.4.2). Later the Altaic background was briefly presented (3.4.3), suggesting that reduplication is most highly developed in the Turkic languages, limited but certainly present in the Mongolic, and vestigial in the Tungusic. This is consistent with the most plausible

scenario that led to the emergence of reduplication (3.4.4), and with the fact that its later expansion was mainly due to areal influence rather than genealogical inheritance (3.4.5), which likely was also the case at the Altaic stage. Finally, the oldest attestations were used (3.4.6) to show that the evolution from Ottoman to Turkish was headed towards unification, and that the 11th century is the terminus ante quem for both reduplication itself, and the diversification of closing consonants. It is not clear whether the diversification was an inherent part of the process of emergence of reduplication, or a later development. The uniformity in the Central Asiatic Turkic might have been caused by the working of analogy, but it might also be that the Oghuz and Mongolic/Tungusic diversity is a coincidence caused by the drive to strengthen the emphatic quality of reduplications. A comprehensive, pan-Altaic study is probably necessary to answer this question.

Perhaps the most important of these is the shift of weight from inheritance to influence. The areal model not only accounts much better than the genealogical one for the number of common reduplicated stems between the various Turkic languages, but it can also better explain the diversification of closing consonants. Moreover, it quite well accommodates the Mongolic languages, and, when extrapolated onto the Altaic stage, complements the genealogical understanding to produce a believable explanation for the Tungusic languages.

Especially the fragments not directly related to Turkic are purely hypothetical at the present stage of research, but the provisional scenario can be summarized as such: reduplication begun in the times of Altaic community as an areal feature. Its centre was located in the Turkic territory, and the phenomenon spread onto the Mongolic languages which passed it further to the Tungusic. The beginnings were most likely very modest. It is possible that different closing consonants were used. Later development was caused, again, primarily by areal influence. The vanguard of the Turkic migration to the west (mainly the Oghuz) added both new stems, and new closing consonants. In Central Asia (the Karakhanid, Kipchak, and South Siberian), new stems were also added but all reduplications were closed by *p*, and if any other had been inherited, they have been ironed out in all languages except for the few remaining examples in westernmost Karaim, Kumyk, Bashkir and Tatar, and the half Oghuz Uzbek. Note that the Kipchak collections are typically much larger than the Karakhanid and South Siberian ones. The Mongolic collections, too, are apparently relatively small. They show a degree of diversification of closing consonants, but more characteristically, they also fall into several types other than the *C*-type, which is the only one discussed in the present work. This is quite like Yakut, except that the Yakut collections of different types of reduplications are rather sizeable. This is also unlike Dolgan, the Turkic-Tungusic mix, whose collection is very small, but typologically diversified, perhaps comparably to the little that there is of reduplication in the Tungusic. With their small and hardly diverse collections, the Tungusic languages do not fit very well between the Mongolic and Yakut. In their case, perhaps, the strength of areal influence decreased after the disintegration of the Altaic community, at least as far as reduplication is concerned.

Chapter 4

Summary

The present book can be viewed as a patchwork of topics relating more or less directly to Turkic reduplications. Many are interconnected and interdependent, which renders it impossible to organize the presentation in a linear way. The thematic division adopted here is only one of the possible groupings, and not necessarily optimal for all tasks.

To alleviate this inconvenience, the current chapter first summarizes the whole following a different thematic division (4.1), and then very briefly recapitulates what I consider to be the most important conclusions (4.2). Some thoughts are expressed more clearly here than in the previous chapters, where they were lost between auxiliary observations.

4.1 Summary

This subsection summarizes the entire book following a different thematic organization than was adopted for the main flow of the discussion. Several topics which had to be fragmented before are merged here. First, the present state of Turkic *C*-type reduplications is sketched (4.1.1), followed by what can be deduced about their past (4.1.2). Then, the process of formation of reduplications is discussed, including the question of the choice of the closing consonant (4.1.3), and finally special cases are shortly presented together with the recurring peculiarities (4.1.4).

4.1.1 Present

The present work gathers about 1200 *C*-type reduplications from twenty modern Turkic languages. Despite my efforts, these collections are surely not complete, but it can be quite confidently assumed that they cover a substantial majority of the examples in actual widespread use (1.2.1). Reduplication is essentially unproductive in all the languages discussed here, albeit there is a claim to the contrary about Tuvinian (2.18.2), and some Turkish examples can be found in digital corpora even though they are

apparently missing from the traditional sources (2.16.1). Sporadic, possibly ad hoc and one-time, new reduplications might be also appearing in other languages.

Within the Turkic family, the defining parameters are the size of the collection and the diversification of closing consonants (3.2.2). Within the Altaic, typological differentiation would also need to be taken into account (3.4.3). The collections vary significantly, and genealogical and geographical patterns can be observed. A synopsis is given in tab. 4.1.

The typological classification that emerges from tab. 4.1 is similar but not identical to the traditional genealogical one. For convenience, the slightly more detailed summaries below are ordered according to the latter.

| Group | Number of reduplications | Diversification of closing consonants | Proportion of non- <i>C</i> -type |
|------------------------------|--------------------------|---|-----------------------------------|
| Karakhanid / South Siberian* | low | none: <i>p</i> | low |
| Kipchak | high | low: <i>p</i> , also <i>m</i> , <i>r</i> , <i>s</i> | low |
| Oghuz | high | high: <i>m</i> , <i>p</i> , <i>r</i> , <i>s</i> | low |
| Yakut | high | low: <i>p</i> , also nine others [†] | high |

Table 4.1: General diversification of reduplications across the Turkic languages. Singular exceptions (e.g. the high proportion of non-*C*-type reduplications in Dolgan) have been omitted for clarity.

* Also includes Dolgan.

† The total of ten different closing consonants results in a relatively high entropy (see fig. 3.3), but reduplications closed by consonants other than *p* constitute only 16% of the Yakut collection, and hence their diversification has been classified as “low” here, as opposed to the Oghuz group where only four consonants are used, but in a much more even distribution.

Karakhanid

Two Karakhanid languages are discussed: Uighur and Uzbek. Sizewise, their collections are both in the second quartile, but they differ with regard to the use of various closing consonants: in Uighur, *p* and *pp* are the only ones attested in the literary language, while in Uzbek also *m*, *s*, and *t* can be found, even if in only 19.57% of the examples. This yields the entropies of closing consonants of 0, and 0.275, respectively, classifying Uighur as a *p*-language, and Uzbek as an *mprs*-language (3.2.1).

Peculiarities are relatively rare. In Uighur, the vowel in the reduplicated anlaut has typically the pre-umlaut shape, while the one in the head undergoes umlauting normally, i.e. *seriq* ‘yellow’ → *sapseriq* ‘very yellow’. In Uzbek, two reduplicated verbs

exist, *qipqizarmâq* ‘to turn very red; to blush strongly’, and *qâpqârajmâq* ‘to turn completely black’. Apart from these, the great majority of examples are standard. Special cases are more numerous in Uighur, and just three in Uzbek.

Sixteen or seventeen reduplicated stems are common to both Uighur and Uzbek (3.2.4), i.e. less than a third of the union of their collections. This is unusually many, but one must not forget that only two languages are being compared here, and the number of common stems tends to drop very quickly as new languages are added to the comparison.

Kipchak

Essentially, seven Kipchak languages are discussed: Bashkir, Kirghiz, Karakalpak, Kazakh, Kumyk, Tatar, and Karaim which is usually treated as two languages here: Eastern Karaim (= Crimean), and Western Karaim. In opposition to the Karakhanid ones, Kipchak collections are quite uniform with respect to closing consonants, but rather varied in terms of their size.

As regards the number of examples, Western Karaim and Kumyk fall in the first, and barely in the second quantiles, respectively. All the other languages are in the third and the fourth. With 108 reduplications, the Kazakh collection in particular, is the second largest in the present work.

With reference to the entropy of closing consonants, on the other hand, all but Eastern Karaim are *p*-languages (3.2.1). The exceptional state of that language is no doubt due to the influence of Ottoman and Turkish. The eastern of the Kipchak languages discussed here (Karakalpak, Kazakh, and Kirghiz) use *p* and *pp* exclusively; the western ones (Bashkir, Karaim, Kumyk, and Tatar) have examples closed by *m* (all except Kumyk), *s* (Eastern Karaim and Kumyk), or *r* (Eastern Karaim). These exceptional reduplications are very few (3.2.5).

Peculiarities are not particularly common. In Bashkir, Karaim, and Kumyk, there are the singular reduplications closed by consonants other than *p*. In Eastern Karaim, moreover, a reduplicated verb can be found (*apačmak* ‘to break open, to throw open’). In Kirghiz, multiple intensifications are relatively frequent (see 3.1.13), and one pronoun is reduplicated (*kačan* ‘when?’ → *kapkačan* ‘(very) long ago’). In Karakalpak, bases beginning with *ža-* reduplicate always to *žä-*, and in a half of the cases, also to the standard *ža-*. Apart from the above, special cases are neither very numerous nor very rare.

The Karaim and Kumyk sources are relatively small, and therefore possibly incomplete. The collections of the remaining five languages were found to have merely ten stems in common (3.2.4), which is more than 4% of the union of their collections, and surprisingly few. Taken pairwise, however, the similarities are much stronger, never less than twenty shared stems (tab. 3.8), but the number drops very rapidly with every new language added to the comparison.

North Siberian

Both Dolgan and Yakut are discussed, and their collections prove to be almost polar opposites. For Dolgan, merely eight examples appear to be attested (the second smallest collection), while Yakut has 106 (the third largest). All Dolgan examples are closed by *p*, whereas in Yakut ten different closing consonants are attested, which is almost twice as many as in the second richest language (Azeri with six). Both languages, however, are extraordinarily diversified from the typological point of view, and *C*-type reduplications constitute only about a half of their entire collections (not discussed in full here). The great gap between the two languages might be due to the Tungusic substrate of Dolgan, and lack thereof on Yakut (3.4.3).

In the light of the above, peculiarities are surprisingly few. The Yakut word *utary* ‘across, opposite’ should be mentioned, whose reduplication *ūnutary* is not only closed by a consonant other than *p* (3.1.23), but also has its reduplicated vowel lengthened (3.1.12). Also the shared Dolgan and Yakut *sogotok* ~ *soyotoχ* ‘alone, lone(ly)’ might be of interest as its reduplications display an unusual diversification of forms. Other than these, special cases are quite rare given the very high total number of reduplications in Yakut.

Six reduplicated stems are common to Dolgan and Yakut (3.2.4), which accounts for 75% of the Dolgan collection, and only 5.7% of the Yakut one.

Oghuz

Four modern Oghuz languages are discussed (Azeri, Gagauz, Turkish, Turkmen), and one historic, Ottoman. Apart from the relatively modest Gagauz collection (second quartile), all are in the third and fourth quartiles. In particular, the Turkish collection is not only the largest one, but by two thirds more numerous than the second biggest collection (Kazakh). In reality, however, the use of a half or more of it is limited in various ways, and not quite representative for the language as a whole (2.16.4 and 3.4.6).

As for the entropy of closing consonants, the Oghuz are the *mprs*-languages proper (fig. 3.3). Apart from Eastern Karaim which has always been under heavy influence of Ottoman and Turkish, the only non-Oghuz language where closing consonants other than *p* are fairly numerous, is Yakut in which, however, their number and distribution is quite unlike in the Oghuz (compare 2.21.4 with e.g. 2.16.4).

Peculiarities seem few, especially in light of how numerous the collections are. Two Azeri words might be mentioned whose reduplications have apparently the meaning of plural rather than intensified singular (*bärbäzäk* ‘decorations, ...’ and *sörsöküntü* ‘chips, splinters’). Also in Azeri, reduplications of derivatives are unusually common (e.g. *japjalnyz* : *japjalnyzža* ‘completely: alone, lone(ly)’). In the remaining languages, special cases are rare, with only the exception of Ottoman where, however, orthography and fidelity of the sources might have played a crucial role in rendering certain attestations unclear.

Perhaps as many as nineteen stems are common to the three languages with large sources (all the modern ones but Gagauz), but it is not clear how the differences in the closing consonants used to close their reduplications should be interpreted. The same closer has been used in eight. (3.2.4.) This constitutes a rather modest base for the investigation of common patterns in the distribution of closing consonants. Several tendencies can be established, but they are far from explaining the majority of all the Oghuz reduplications (3.2.3 and 3.2.5.)

The nineteen stems are only less than 9% of the union of the Azeri, Turkish and Turkmen collections. Note, however, that without Turkish, the union shrinks from 216 to 94 reduplications.

Apart from the conspicuous difference in size between the Ottoman and the Turkish collections, not much can be deduced from their comparison. If anything, the evolution seems to have taken a tiny step towards standardization. (3.4.6.)

South Siberian

Four South Siberian languages are discussed: Khakas, Oirot, Shor, and Tuvinian. Their collections are uniform and modest. Khakas, with almost twice as many examples as the second Tuvinian, is in the second quartile; the remaining three are all in the first. Shor, in particular, has the smallest collection in the present work (six reduplications). Closing consonants other than *p(p)* are not attested.

Systematic peculiarities effectively do not occur. In Tuvinian, reduplication is advertised by K.D. Harrison to be fully productive, and to operate on different parts of speech, including verbs, but his claim is yet to be verified (2.18.2). Special cases are surprisingly numerous, but almost all belong to two families of derivatives of *ak* 'white' and *kök* 'blue', with seventeen or eighteen forms across Khakas, Oirot, and Shor (2.12.3).

Excluding the underrepresented Shor, only five stems are common to the South Siberian languages, i.e. a tenth of the union of their collections, which is relatively many.

4.1.2 Past

At the present stage of research, not much can be said with certainty about the history of Turkic reduplications.

The very beginnings of the phenomenon are particularly unclear. Possibly, they are as ancient as the times of the Altaic speech community, if the method is present and, it seems, doing well in the Mongolic languages. It is also present in the Tungusic languages, but on a much smaller scale. A detailed and extensive comparison is necessary for dependable conclusions to be drawn, but provisionally, it might be assumed that Mongolic and Turkic reduplications are continuations of a common heritage rather than independent innovations. (3.4.3.)

How reduplication started, is also unknown. More and less plausible scenarios can be devised but without actual data to rely on, they remain hypothetical. A new, and perhaps more likely, scheme is proposed in the present work, that derives the phenomenon from a misinterpretation of **köp kōk* lit. ‘very blue’, but it is not without weaknesses, too. It is possible that there were in reality several different formations which accidentally just sounded similar, that combined to create the impression of being a separate method of intensification. (3.4.4.)

The earliest certain attestations come from Maḥmūd al-Kāšyarī. It can be deduced from his account, that reduplication was already an established method of intensification in the 11th century, which displayed a geographical and/or genealogical differentiation (*p* being the perhaps universal closer, and *m* limited to the Oghuz languages), and had already had exceptions (*tästāgirmä* ‘absolutely round’). (2.1.1.)

Later attestations do not reveal much, mostly because ascribing them to the predecessors of specific modern languages is more often than not uncertain. (For the same reason Chaghatai material is entirely omitted in the present work.) Ottoman data are an exception, but a comparison with the modern Turkish collection only unearthed a weak tendency to standardize in just several examples. (3.4.6.)

The details of evolution of reduplication are also shrouded from investigation by the relative similarity of the Turkic languages, which makes it difficult or often impossible to tell internal borrowings from calques from innovations. Also, the existence of a stem and its reduplications in several languages cannot be in itself regarded as a proof of common descent. Combined with our insufficient knowledge about the exact relations between specific languages in the past, and several smaller obstacles, this renders the classical historical-comparative method barely applicable. (3.4.2.)

These difficulties recede into the background, however, as it is found that only a very low proportion of reduplicated stems is common to more than just a handful of languages, and their number plunges with every new language added to the comparison. In fact, *kara* ‘black’ seems to be the only base common to all. This situation is not at all consistent with the relatively large sets of stems common to various pairs of neighbouring languages. One conclusion that can be drawn from this incompatibility, is that Turkic reduplications were not so much inherited by every language independently, as jointly developed in local clusters of dialects and languages.

Reduplication can be thus seen as another argument in favour of viewing the Turkic family as a “continuum of dialect continua” rather than as a genealogical tree of distinct, separate branches. (3.4.5.) Although more hypothetically, the same interpretation can in fact be used to explain the modern distribution of closing consonants (4.1.3 below).

If common stems are anything to go by, which they not necessarily are (see above), reduplication seems to have evolved along four separate paths in the Turkic languages: Oghuz, Kipchak, Yakut, and Karakhanid-South-Siberian-Dolgan. This is in accord

ance with the typological diversification of reduplications (tab. 4.1), and also quite similar to the traditional genealogical division of the family. (3.2.2 and 3.2.4.)

Semantically, however, only two main paths can be observed: Yakut, and the rest of the family. This suggests that reduplication encompassed almost everywhere a set of similar, perhaps the most frequent, intensifiable adjectives and adverbs which, simply, were expressed with different stems in various genealogical groups. (3.3.4.)

To sum up, reduplication began probably during the times of the Altaic speech community, and from this time on, it evolved in all the Turkic languages under the influence of their neighbours rather than as a continuously expanded inheritance. For this reason, collections in the modern languages display similarities primarily in geographical clusters, which only happen to coincide with the genealogical division. Semantically, nonetheless, only Yakut departs significantly. The process of diversification of closing consonants began before the 11th century, maybe as early as the very origins of the phenomenon.

4.1.3 Formation

At first glance, the formation of Turkic *C*-type reduplications might seem to be well understood except for the apparently quite complex problem of the choice of the closing consonant. In fact, the traditional account of the process appears to be fallacious, and the mechanism of choosing the closer actually relatively simple.

Mechanism

According to the established and most common account, reduplication proceeds by doubling the first syllable, ejecting its final consonant if there were any, shortening its vowel if it was long, appending a closing consonant to it, and prepending the whole to the original word. However, the same final shape can be obtained by different procedures, e.g. by only doubling the initial mora of a word and inserting a single (or double) consonant in between.

The exchange of syllables to morae is a particularly convenient descriptive manoeuvre as it automatically entails what is typically seen as a separate process of shortening of the reduplicated vowel, and results in effect in a more concise formulation.

Nonetheless, it is probably impossible to determine which description better captures the actual mental processes that have created reduplications. In fact, one may not even confidently assume that the mechanics is the same now as it was in the past, and that it is the same for all speakers – or indeed, that it is at all conscious, as these descriptions imply. A psycholinguistic study conducted on Turkish native speakers suggests that the choice of the closing consonant is only driven by a rule with words beginning with a vowel (when the closer is almost always *p*, with just one exception in Yakut, 3.1.20), while with those beginning with a consonant it is either memorized as

a part of the vocabulary (in the more common reduplications), or unclear to the speakers themselves (in the less common ones). It might be suspected then, that the actual mechanics of the formation of reduplications is, simply, analogy. (3.2.6.)

Also, it should be noted that secondary phonetic modifications might be applied independently to any part of the resulting reduplication. These are not frequent cases, but in Karakalpak and Uighur, they appear to be regular. (3.1.19.)

Note also that the descriptions above only refer to what is called “C-type reduplications” in this work, i.e. reduplications with a single or doubled consonant for a closer. Other types of closers, however, are also attested (e.g. Dolg. *čä.bit.čälkã* ‘snow-white’ (DW), Kklp. *tu.ppa.tuwra* ‘completely straight, very direct’ (KklpRS), or Yak. *bö.rü.böyö* ‘very: strong, sturdy, powerful’ (Pekarskij 1907–30), &c.), and also other types of partial and full reduplications (e.g. Tksh. *ders mers* ‘classes and what not’, *onlar monlar* ‘they and whoever else’, *kara (my) kara* ‘jet-black’, &c.). (1.1.1.)

Choice of the closing consonant

Perhaps the most investigated issue connected with the formation of Turkic C-type reduplications, is the question of the choice of the closing consonant. Previous research has primarily focused on establishing synchronic and phonetic rules for Turkish, and achieved some success, but failed to ultimately resolve the puzzle. Various sets of rules account for about 70–80% of the examples. The more precise ones, however, seem so intricate that a suspicion might arise about whether linguistically naïve speakers really could have been following them while coining reduplications, or in other words, whether these attempts are not in fact inventing a mechanism that never was. (1.1.2.) In other languages, the situation is either very clear (the closing consonant is always *p*), or equally ill-suited to a purely synchronic and phonetic investigation (3.2.3).

In the present work, a comparative and diachronic attitude has been adopted instead. The Oghuz group and Yakut, are those languages where closing consonants other than *p* are not just singular exceptions. The latter has been discarded for lack of close relatives with which to conduct a comparison.

As for the Oghuz group, I identified a set of nineteen common reduplicated stems in them, and compared which closing consonants have been used in their reduplications. The variance proved to be relatively low, but despite that not many phonetic regularities could be established. Stems beginning with *b* and monosyllables were found to typically have their reduplications closed by *s* or *m*, but both these closers proved to be used almost as frequently with reduplications of stems that do not share any particular phonetic feature. Thus, both the validity and the phonetic motivation of these findings is questionable. (3.2.5.)

A comparison of all reduplications in all the modern languages discussed here, yielded results that are more credible but neither very novel nor very accurate. Bases with a vocalic anlaut have their reduplications closed by *p* in 185 out of 186 cases, and this is the only

actual phonetic rule for the choice of the closing consonant (3.1.23). For bases beginning with a consonant, the following set of tendencies has been established (3.1.4 and 3.1.5):

- Regarding the initial consonant of the base (C_1):
 1. p -languages use p as the closing consonant indiscriminately for all bases.
 2. $mprs$ -languages tend to replace p with s , m and rarely other consonants for bases beginning with b -.
 3. $mprs$ -languages use s and, less often, r as the closing consonant for bases beginning with d and t .
 4. $mprs$ -languages almost never use s for bases beginning with s -.
- Regarding the second consonant of the base (C_2):
 1. p -languages use p as the closing consonant indiscriminately for all bases.
 2. $mprs$ -languages do not avoid homolocality of the closing consonant with C_2 .
 3. $mprs$ -languages tend to avoid using C_2 as the closing consonant.
- Tendencies referring to C_2 are weaker than those referring to C_1 .

From these, the primary driving force behind the use of different closing consonants appears to have been the will to avoid repetitions with both C_1 and C_2 (as in Kmk. *papparahat* ‘absolutely calm’ or in Uzb. *bütbütün* ‘quite complete’). This conclusion is consistent with the findings of Hatiboğlu, Demircan, and Müller (see 1.1.2). Unlike theirs, however, the current analysis does not suggest that this tendency was so strong as to impel speakers actively to maximize the phonetic difference between the closing consonant and the first consonants of the base.

Rather, it appears that once the tendency to avoid repetitions was satisfied, the choice between the remaining possibilities was fairly random. It is not a rare occurrence for one base to have different reduplications with various closing consonants (3.1.1), and in some cases even semantic differentiation is to be seen between the alternatives (3.1.2). Nonetheless, a weak geographical pattern can be observed. In Azeri, m is clearly more common than s , and not very much more common than r . In Turkmen, m is exceptional, s is the main non- p closer, and r is missing entirely. In Turkish and Gagauz, m is the second most common, s the first, and r is very rare. (Tab. 3.6.)

The above observations combine to create a picture fairly similar to that drawn by common reduplicated stems, namely one where the evolution of our phenomenon is primarily driven by areal influence (3.4.5). As it is not inheritance that has most actively shaped the modern collections, it seems that it is also not only phonetics that has decided about the closing consonants. The basic phonetic condition being met, various alternative reduplications might have locally arisen, and be forgotten or remembered with or without their sisters, and which it was, would be decided upon more by analogy than by phonaesthetics.

The differences in the distribution of closing consonants in various languages are better accounted for by this scenario than by phonetics and phonaesthetics alone, as these do not in fact vary so dramatically across the Turkic languages as to inspire such diversification in the choice of closers. The reason, then, why synchronic phonetic rules remain elusive, is simply because they do not quite exist, and the variation observed in the modern literary languages is the product of interactions between various dialects, each of which has its own unique collection of reduplications, and its own specific conditions for analogy. Apparently, their intersection is only the tendency to avoid repetitions, and then, to pass the decision process over to analogy (which will also ascertain that accidental similarity to independent words is avoided, &c.).

4.1.4 Exceptions

Almost all of the observations above referred to what might be considered standard cases, i.e. reduplications built according to the same model, with their bases attested as independent words, with the usual closing consonants, and no extraordinary phonetic or semantic changes. But the sources contain also a considerable number of atypical examples. A part of them are singular cases not included in the main stream of considerations due to various problems in their interpretation (“special cases”), others are generally clearer, not isolated, and analysed together with the standard cases, but still irregular in one way or another (“peculiarities”).

Special cases

Altogether, more than eighty words have been classified as special cases, discussed in the “Special cases” subsections in chapter 2, and essentially not considered any further.

By far the most frequent reason for such treatment was that the alleged base did not seem to be attested as an independent word. Quite often, the whole was found to be a loanword, and either originally a reduplication (e.g. Kar.E *simsijah* ‘completely black’ < Ott. *simsijāh* id.), or just accidentally similar in shape (e.g. Tksh. *sersem* i.a. ‘stunned, bewildered’ < MPers. *sarsām* سرسام lit. ‘head inflammation’). Also not rarely, the word proved difficult to dissect, and I could not even be sure that it is indeed a reduplication (e.g. Uigh. *žimžit* ‘sudden silence’ (see 3.1.15 for possible cognates), or Kar.E *komkos* and *tentek*, both ‘very stupid’). In a few cases, the exact phonetic shape appeared to be missing from the sources, either because it was probably just a slovenly pronunciation (e.g. Kar.E **slah*, while *sylak* ‘wet’ is attested), or because the simplified version had already ousted the original one (e.g. Kar.E **jumarlak*, while *jumalak* ‘round’ still exists), or because a secondary simplification had been applied to the reduplication but not to the base itself (e.g. Yak. *soččoyotoχ* &c. ‘lone(ly)’ ← *soyotoχ* id.), or finally because the attestation was in all likelihood erroneous (e.g. Tksh. **sirin* ‘cute’ instead of *širin* ‘sweet, ...’). Lastly, in one or two cases, exclusion from the

main considerations has been only applied for consistency, and to avoid creating vague and subjective precedents (e.g. Az. **büzäklük*, while *büzäk* ‘decoration, decorative’ is attested together with *büzäkli*, *büzäklilik*, *büzäksizlik*, &c.).

The second most frequent reason for exclusion was lack of structural transparency. In the majority of cases, this entailed the inability to identify the base, and as a consequence, to find an attestation for it as an independent word (e.g. Oir. *appāš* ‘snow-white’ or *kökpögöš* ‘very blue’). Less often, my doubts were raised by the reduplicated anlaut (e.g. Kirg. *kyp žyłaŋač* ‘stark-naked’, Oir. *berbek* ‘fat’, Uzb. *pakpakana* ‘very short of a person’).

Finally, several words are exceptional among exceptions. Tksh. *žysžybyldak* ‘stark-naked’, for example, was excluded as a dialectal shape (see 1.2.1 on the choice of sources), Tuv. *šipšimēn* for semantic doubts, Uigh. ⟨āpāq⟩ as a probable loanword, and Uigh. *kim kiček* ‘garment, clothes’ and *žuyžemi* ‘stock and block’ as more likely to be nominal compositions than reduplications.

Peculiarities

Various kinds of peculiarities are discussed in twenty-four subsections in 3.1.

Most often, the anomaly lies in the structure. The base might be a different class than a simple adjective or adverb (3.1.14, 3.1.17, 3.1.18, 3.1.22), a secondary phonetic or other modification might have been applied to a part of the reduplication (the closer: 3.1.8, 3.1.21, the reduplicated anlaut: 3.1.20, any: 3.1.12, 3.1.19), or the whole might be going beyond the standard build of C-type reduplications (3.1.9, 3.1.16). Frequently, the crux is the closing consonant: its shape (3.1.8, 3.1.21), the rules behind its choice (3.1.4, 3.1.5, 3.1.6, 3.1.23), and the results of their ambiguity (3.1.1, 3.1.2).

Etymology is also relatively often the point of interest. Reduplications might belong to families spanning across several languages (3.1.11, see also the unusually productive bases meaning ‘white’ in 3.1.24), might be formed from now-obsolete bases (3.1.15), or despite appearances, they might not in fact be reduplications at all (3.1.3). There is also a suspicion that reduplicated anlauts might be severed and promoted to independent or semi-independent intensifiers (3.1.10).

Finally, several subsections are devoted to multiple intensification, which can be realized in different ways. A collective discussion is given in 3.1.13.

4.2 Conclusions

Below is gathered what I consider to be the most important findings in the material collected in the present work, and in its analysis. The thematic organization is similar, but not identical, to the one used in 4.1.

- Research
 - The quest for all-embracing, synchronic, phonetic rules for the choice of the closing consonant is futile (3.2.3, 3.4.1, 3.4.2).
 - The correct approach is the historical-comparative one, but it is in practice quite limited (3.4.2).
 - Here, well above 566 000 dictionary entries have been checked to collect more than 1200 *C*-type reduplications from twenty modern, and some historical languages (1.2.1).
- Structure
 - Reduplication can be more conveniently described as operating on morae than on syllables (3.2.6).
 - The actual mechanism behind reduplication is probably analogy now, and unknown at the very beginnings of the phenomenon (3.2.6, 3.4.4).
 - Outside of the Oghuz languages, Eastern Karaim, and Yakut, *p* is almost the only closing consonant in use (3.2.1).
 - The distribution of closing consonants is only partly due to phonetics and inheritance, and partly due to areal influence (4.1.3). Semantics does not seem to play a role in the process (3.3.4).
 - A definitive set of phonetic rules for the choice of the closing consonant probably does not exist (4.1.3).
- Semantics
 - Reduplications are not in any way limited to colour names, and are more or less equally diversified in all languages (3.3.3).
 - The semantic scope is quite similar in all languages except Yakut (3.3.4).
 - There is no correlation between semantics and the closing consonant (3.3.4).
- Geography
 - The territorial scope and diversification of reduplications is more due to areal influence than genealogical inheritance (3.4.5).
 - Structural and material similarities correlate with the genealogical classification, except that the Karakhanid, the South Siberian, and Dolgan seem to all fall into one group (3.2.2).
 - Roughly, reduplication is the more developed, the larger the geographical distance from the Altai homeland (3.2.2).
 - The languages with the largest collection are scattered across entire Asia: Turkish, Kazakh with Bashkir and Kirghiz, Yakut (3.3.4).
 - Closing consonants are diversified in the peripheries (Oghuz, Eastern Karaim, Yakut, and singular examples in Bashkir, Kumyk, Tatar, and Western Karaim), and effectively always *p* in the centre (the remaining Kipchak, Karakhanid, and South Siberian, also Dolgan). (3.2.1).
 - Semantics are approximately equally diversified in all languages (3.3.3).

- Evolution
 - Reduplication probably began during the Altaic speech community (3.4.3), maybe with more than one closing consonant (3.4.4).
 - It was already developed and diversified in the 11th century (3.4.6).
 - The evolution of reduplication is more due to areal influence than genealogical inheritance (3.4.5).
 - The Ottoman and Turkish collections are quite similar, except in size. A weak tendency towards standardization to *p* is noticeable (3.4.6).
 - Structural and material evolution occurred in four directions, whereas semantics evolved according to approximately the same plan in all languages but Yakut (3.2.4, 3.3.4).

Appendix A

Müller 2004

This section has been moved to an appendix to improve the consistency and readability of the main part of the work. In it, Müller's 2004 rules for choosing the closing consonant in Turkish reduplications are examined in more detail (A.1), and his test on 125 Turkish students is analysed (A.2), and simulated with the use of artificial neural networks (A.3) to better understand Müller's results (A.4).

A.1 Rules

Müller 2004: 149f gives a set of seven rules governing the formation of reduplications in Turkish. Rules B through F refer to the choice of the closing consonant and will be discussed below; rules A and G are only adduced for completeness. Footnotes with references to other parts of Müller 2004 have been left out.

Regel A

Ein Adjektiv mit erster Silbe $(C_1)V_1(C_{2/3})$ wird intensiviert durch Voranstellung der Silbe $(C_1)V_1\mathbf{X}$ (hier I[ntensiv-]A[djektiv-]R[eduplikation]-Silbe oder IAR-Morphem genannt [...]), wobei \mathbf{X} einer der Überleitungslaute \mathbf{m} , \mathbf{p} , \mathbf{r} oder \mathbf{s} ist, d.h. es fällt die Koda ($C_{2/3}$), so vorhanden, aus, und es wird dem Adjektiv als IAR-Silbe [bei $C_1 = \emptyset$] Nukleus $V_1 + \mathbf{m}$, \mathbf{p} , \mathbf{r} oder \mathbf{s} bzw. [bei $C_1 \neq \emptyset$] Anlaut $C_1 +$ Nukleus $V_1 + \mathbf{m}$, \mathbf{p} , \mathbf{r} oder \mathbf{s}] vorangestellt. („IAR-Strukturregel“)

Regel B

Normaler Überleitungs laut ist der stimmlose labiale Plosiv \mathbf{p} . Er wird immer eingesetzt, wenn nicht bestimmte Gründe, die in den folgenden Regeln angegeben sind, dagegen sprechen. Spricht etwas (mangelnder Kontrast) dagegen, wird zwischen \mathbf{m} , \mathbf{r} und \mathbf{s} als Ersatz-Überleitungs-lauten gewählt. („p-Grundregel“)

Regel C

Beginnt das Wort mit einem Vokal, wird **p** und nur **p** als Überleitungslaut verwendet. („**Vokal-p-Regel**“)

Regel D

p ist als Überleitungslaut dann nicht (mehr) möglich, wenn das Adjektiv selber mit einem labialen Laut (*b, p, m*) beginnt. Als Ersatz-Überleitungs-laute kommen **m, r** und **s** und nur diese in Betracht (s. obige Konsonan-ten-Tabelle, Reihenfolge wie in der IAR-Matrix). („**p-Alternativregel**“)

Regel E

m (und kein anderer Laut) kommt als Überleitungslaut bei Wörtern dann in Betracht, wenn diese mit *b, p, d, t, s, c, ç* oder *y* beginnen und als weitere Konsonanten den Laut *k* (bei einsilbigen Wörtern in jedem Falle) oder Zischlaute, aber nicht *m* enthalten. Für Wörter, die mit *g/k* beginnen, gibt es nur die Zwei Fälle [13] *gök* und [143] *kör*. Aber *gök* ist einsilbig [...] und auslautend auf *-k* und passt somit auch wieder in den Rahmen. Und *kör* nimmt wie die meisten einsilbigen Adjektive ebenfalls **m**.

s kommt als Überleitungslaut dann in Frage, wenn die Adjektive mit *b, p, d, t, g, k, c, m* oder *y* beginnen und keine Zischlaute (von den Affrikaten *c* in *koca(man)* und *ç* in *topaç* und von *ş* in *yumuşak* abgesehen) als Bin- nenlaute vorkommen und sie nicht einsilbig auf *k* enden. („**m-/s-Regel**“)

Regel F

Für die 8 Adjektive *perişan, temiz, top, çabuk, çiplak, sebil, sefil* und *sıklam* (also ohne *sem*) wird als Überleitungslaut **r** verwendet. Für diese sind an- dere Überleitungslaute mangels Kontrastmöglichkeit wenigstens teilweise ausgeschlossen. („**r-Regel**“)

Regel G

Die (mit einem einsilbigen Morphem gebildeten) Intensiv-Adjektive tra- gen ihren Hauptton auf der ersten Silbe. („**IAR-Betonungsregel**“)

Some comments on these rules were given in 2.16.1. Let us briefly recapitulate here the points dealing with the choice of the closing consonant:

- The rules are synchronic and phonetic. Hence they can never achieve absolute accuracy, as is demonstrated by the pair *jepjeni* : *jesjeni* ‘brand-new’ and 16 more like it (see 2.16.4). In fact, they only cover about 79% of the reduplications that were used as the base for their formulation.
- They are sometimes mutually exclusive: e.g. *berrak* ‘limpid, clear’ or *dāynyk* ‘scattered, dispersed’ would have to have *p* or *s* on one hand, and *m* “und kein anderer Laut” on the other, at the same time.

- The rules contain exceptions. The rule for *r* is just a list of specific base words and nothing more.
- The rule for *s* is optional. Only the distinction between *p* and *m* is categorical, but because of this, it is incorrect in 59% of cases.

A.2 Test

Undeterred by the obvious shortcomings of his rules, Müller 2004 put them to two similar tests. The first was a small scale try-on, and will be ignored here. The second was large scale, and I will focus on this one.

Müller prepared three corpora (p. 251f): one that contained 100 words which actually have reduplications in Turkish (corpus A), one with 94 words which do not (corpus B) and one with 24 nonsense words (corpus C), and asked 125 Turkish students at the Başkent and Hacettepe universities to reduplicate them.

The idea is a fragile one. Not being certain that phonetics is the only or even the main factor in the choice of the closing consonant, this test somewhat resembles asking present-day English speakers to build irregular past participles for a set of regular and a set of made up verbs. It seems that Müller's goal was to test whether his rules are correct, but it is not at all clear to me how he intended to husk this information from the results. In theory, a test so designed could be just as well aiming to answer several other questions, e.g. *Do Turkish speakers follow any specific set of rules when asked to reduplicate non-existent words?*, *Are modern rules compatible with the ones that had created Turkish reduplications?*, *How similar to the literary language is the linguistic intuition of students in Ankara?*, and perhaps many others. Whatever the results, their interpretation will always leave room for doubt if not conducted expertly.

(This is not to say that any experiment of this kind must be entirely pointless as a rule. A similar test can be designed in a way capable of yielding usefully interpretable results, and in fact has been by Sofu 2005 and Sofu/Altan 2009. See 1.1.2 for a short summary.)

Next, Müller compared the results to a prediction based on his rules. Rather amusingly, his prediction actually breaks his own rules in 31% of cases in corpus B and 29% of cases in corpus C – or 46% and 41%, respectively, if a lack of possibility is counted as breaking the rules. Examples in tab. A.1.

The answers of the interviewed students correspond with Müller's prediction in 84.2% for corpus A, in 57.7% for corpus B, and in 36.1% for corpus C. Müller did not publish all the questionnaires, and so it is not possible to compare them to the corrected predictions. Only the exact scores of 51 students are available. There is no correlation between the results for either pair of corpora, which may be found slightly surprising knowing that they were clearly separated in the questionnaires (Müller 2004: 252).

Given that the students were allowed to pick for each word one of the five answers (*m*, *p*, *r*, *s* and *?*), the results are significantly higher than pure chance. Yet, they are far too low to prove Müller's rules linguistically significant.

| Corpus | Word | Müller's prediction | Correction | Justification (Rule) |
|--------|----------------|----------------------|----------------------|---|
| B | <i>dilber</i> | <i>m</i> or <i>s</i> | <i>p</i> or <i>s</i> | <i>m</i> not in words without sibilants or <i>k</i> (E) |
| B | <i>miskin</i> | <i>p</i> | impossible | <i>m</i> and <i>p</i> not in words beginning with <i>m</i> - (D, E), <i>r</i> only in a fixed set of exceptions (F), <i>s</i> not in words containing sibilants (E) |
| B | <i>rekik</i> | <i>m</i> or <i>s</i> | <i>p</i> | <i>m</i> and <i>s</i> not in words beginning with <i>r</i> - (E), <i>m</i> and <i>s</i> mutually exclusive (<i>m</i> “und kein anderer Laut”) (E) |
| C | <i>čobal</i> | <i>s</i> | <i>p</i> | <i>s</i> not in words beginning with <i>č</i> - (E) |
| C | <i>darylyk</i> | <i>s</i> | <i>m</i> | <i>m</i> “und kein anderer Laut” (E) |
| C | <i>sejil</i> | <i>r</i> | <i>p</i> | <i>r</i> only in a fixed set of exceptions (F) |

Table A.1: Müller's prediction of the closing consonant for words which do not have reduplications in Turkish (p. 246f, 249f), and my corrections based on his own rules.

The facts that Müller presented the three corpora separately to his interviewees, and that he prompted their choices by so narrowly limiting the available answers, are probably sufficient from a psychological point of view to discard his results completely, which I choose here to refrain from because analyzing them leads to conclusions that might prove interesting for the linguistic methodology.

The rules cannot be simply ignored as entirely ‘wrong’ because they do actually match a greater part of the fixed set or Turkish reduplications. Clearly, however, their predictive power is far below what would be expected from *rules* in linguistics. At least three explanations are possible:

1. The rules capture the wrong factors. They are sufficiently accurate to explain the majority of the relatively small set of actually existing reduplications (Müller used 165 to formulate his rules), but not to make predictions.
2. The choice of the closing consonant is not entirely based on the phonetic shape of the base. Analogy to the most frequent reduplications might also play a role, and other factors, too.
3. There might simply not exist *one, synchronic* set of rules which all speakers of Turkish strictly follow to form new reduplications. (Especially that the method is no longer productive.)

At least partially, explanations 2 and 3 are certainly true, as demonstrated by some words which can be reduplicated with different closing consonants, e.g. *čimčij* ~ *čipčij* ‘completely raw’ or *jepjeni* ~ *jesjeni* ‘brand-new’.

In the following subsections, I will try to find a more definite answer with the help of artificial neural networks.

A.3 Simulation

Artificial neural networks (ANN) are known for their particular pattern extraction capabilities and the ability quite adequately to mimic or approximate human behaviour. Since their accuracy is to a certain degree random, but at the same time easier to control than that of human interviewees, I could use networks to simulate Müller’s 2004 test on 125 Turkish students, and to better understand his results.

I prepared three corpora.

Müller 2004: 252 mentions that his corpus A was composed of a hundred words which really have reduplications in Turkish. Unfortunately, he does not precise what these words exactly were. One can only guess it is the hundred he had initially extracted from Steuerwald 1972 or 1974 (p. 84f). He does also state that he used 165 words to formulate his rules (p. 120). However, the list – scattered on p. 86f, 109 and 119f and mixing different types of reduplications and sometimes multiple words inside one entry – contains in fact 166 usable examples. I settled with the latter corpus as the training data.

Corpora B and C are listed on p. 246f and 249f.

I tested several architectures and training algorithms and, found that I can achieve the best results with a multi-layer perceptron as implemented in the *neuralnet* package for R (Fritsch/Günther 2012, see also Günther/Fritsch 2010) trained with the resilient backpropagation algorithm without backtracking (*rprop-*).

In this model, the input data is required to be of constant length, but the words in Müller’s corpora, naturally, are not. I tried both artificially extending the shorter words (see A.3) and clipping the longer ones. I achieved a higher accuracy on the training data with the former approach, but it needs to be noted that the results with the corpora shortened to just the first three phonemes were only a little worse: with 1–50 hidden neurons, the average was 81.32% and 77.36%, respectively, and the maximum scores were 88% and 83.1%. (Note that 11 out of the 166 words appeared twice in the corpus with different closing consonants, e.g. *čimčij* ~ *čipčij* ‘completely raw’.)

With the shorter words extended, the top three total accuracies for all three corpora were achieved by networks with 15, 22 and 6 hidden neurons: 197.8%, 192.7% and 191.1% (out of the maximum 200%), respectively.

The accuracy of a network hinges on a number of settings. For best repeatability, I tried to keep the number of variables as low as possible. Apart from what was mentioned above, the main factors are:

- Encoding: see below.
- Error threshold: the minimal required accuracy during training; was always 0.1.
- Random seed: the state of the pseudo-random number generator used for initializing the weights before training. Unless stated otherwise, was always 1. See A.3.
- Other: I kept the default values; see Fritsch/Günther 2012 for details.

Encoding

Since my goal was to test Müller's *phonetic* rules for choosing the closing consonant, I could not simply turn every phoneme into a random unique number, but rather had to encode them in a way that would mirror their phonetic qualities.

First, I tested a system based on Li/MacWhinney's 2004 proposition for English. Dissatisfied with the accuracy, I prepared my own encoding, and found that it performs considerably better in a test where the network computes the phonetic shape of the Turkish past tense suffix for 170 monosyllabic verb roots, divided evenly between the training and test corpora. Details of the encoding and of the test can be found in Stachowski K. 2012b. Below, I will only outline the idea and the results.

In this encoding, each phoneme is represented by three numbers (three dimensions) which denote: place of articulation, manner of articulation, and the number of additional organs taking part in the articulation (vocal cords, nasal cavity, sides of the tongue, lips). Vowel length is not represented, as it appears that nothing depends on it in the Turkish phonology.

The entire system for Turkish is given in tab. A.2. Some insight into the distances between phonemes can also be obtained from the dendrogram in fig. A.1. Example: /p/ is represented as the ordered triple (1, 1, 1), which stands for (labial, stop, voiceless), /b/ as (1, 1, 2), /f/ as (1, 3, 1), &c.

Where shorter words needed to be extended to fit the template, they were completed with a dash represented as the triple (-5, -5, -5).

To test the encoding, I collected 170 unique monosyllabic Turkish verb stems, ordered them alphabetically, and used the odd ones as the training corpus and the even ones as the test corpus. I also tested two CVCC templates; one filled from the left (e.g. bak- 'to look'), and the other one from the right (e.g. ba-k).

The network's task was to compute the shape of the past tense suffix which, depending on the phonetic shape of the base, can take the form of one of the eight combinations of $d \sim t + y \sim i \sim u \sim ü$. As for the consonant, d is chosen iff the final phoneme of the stem is voiced; as for the vowel, y and u are used iff the final vowel of the stem is back, and u and $ü$ iff it is labial. Examples: *çyk.ty* '(s)he quit', *kes.ti* '(s)he cut', *vur.du* '(s)he hit', *ör.dü* '(s)he knit'.

I tested networks containing 1–100 hidden neurons. The best total accuracy without any template was 194.1% (100% on the training corpus and 94.1% on the test corpus). With a template, the best accuracy achieved was 197.6% (again, 100% on the training corpus). I consider this to be an acceptable approximation of human accuracy.

| | Labial | Alveolar | Palat-alveolar | Palatal | Guttural | Value |
|------------|--------|----------|----------------|---------|----------|-------|
| Stop | p b m | t d n | | | k g – | 1 |
| Affricate | | | ç ǰ – | | | 2 |
| Fricative | f v – | s z – | ş ž – | | h – – | 3 |
| Liquid | | – r l | | – j – | | 4 |
| High vowel | | | | – i ü | – y u | 5 |
| Low vowel | | | | – e ö | – a o | 6 |
| | 1 | 2 | 3 | 4 | 5 | Value |

Table A.2: Phonological encoding of Turkish. The third dimension is flattened in the table and represented by the ordering of the phonemes. Thus e.g. /p/ lies at $z=1$, /b/ at $z=2$, and /m/ at $z=3$.

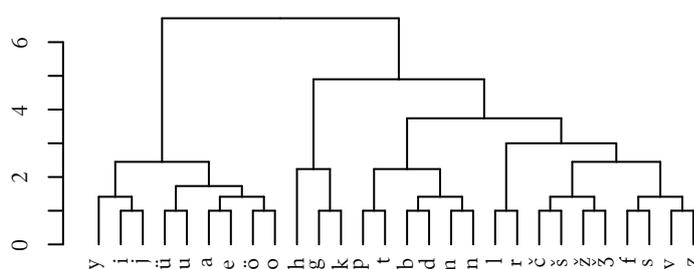


Figure A.1: Dendrogram of the distances between phonemes as defined by the numerical representation in tab. A.2. Note that a certain amount of bias is inevitable as the dendrogram reduces a three-dimensional (three-featured) representation to a two-dimensional plot.

Test

Müller tested 125 students and compared the results with his prediction (see A.2). Naturally, each student answered differently and scored differently. The average coincidence was 84.2% for corpus A, 57.7% for corpus B and 36.1% for corpus C. Unfortunately, Müller did not publish the exact answers of all the students.

Having determined the optimal training algorithm, encoding, error threshold and the number of hidden neurons, and leaving all the remaining settings at their default values, I used the random seed to simulate Müller's experiment.

For each of the three top scoring numbers of neurons (see A.3), I created 125 networks using every time a different random seed from the range one to one million. The average accuracies are given in tab. A.3.

Note that the numbers in the first row in tab. A.3 cannot be compared directly in the same way as those in the lower ones. Müller only used a corpus of 100 words, the majority of which the majority of the students he interviewed, knew by heart. I, not knowing what words he used exactly, was forced to train and test the network on a

| Corpus | Müller 2004 | ANN, 15 neurons | ANN, 22 neurons | ANN, 6 neurons |
|------------|----------------|--------------------|--------------------|-------------------|
| A/Training | 84.2% | 76.43% | 81.80% | 65.86% |
| B | 57.7% | 54.21% | 52.74% | 57.56% |
| C | 36.1% | 32.46% | 32.30% | 35.00% |

Table A.3: Comparison of the mean accuracy of the students interviewed by Müller 2004, and of different neural networks.

| Corpus | Müller 2004 | ANN, 15 neurons | ANN, 22 neurons | ANN, 6 neurons |
|------------|----------------|--------------------|--------------------|-------------------|
| A/Training | 0.0434* | 0.0548 | 0.0365 | 0.0908 |
| B | 0.1361* | 0.0616 | 0.0484 | 0.0860 |
| C | 0.2167* | 0.0861 | 0.0961 | 0.1236 |

Table A.4: Comparison of the standard deviation of the students interviewed by Müller 2004, and of different neural networks.

* Based only on the accuracies of 51 students (Müller 2004: 253f). The results of the remaining 74 were not published.

larger corpus of 166 words, which I only believe must have been almost the same as the corpus Müller based his rules on (see A.3). Also, while neural networks can quite well recognize the recurring general patterns, they do have some difficulties remembering exceptions, and out of the 166 words in this corpus, 11 appeared twice with different closing consonants.

The tests on the neural networks are random to a certain degree (see the standard deviations in tab. A.4), but this randomness is largely ironed out by the high number of tests. Overall, there is a good correlation between Müller's results and those of the networks: 0.9986 for the network with 15 hidden neurons, 0.9991 for the one with 22, and 0.9494 for the one with 6.

A.4 Interpretation

An interpretation of the simulation is not immediately obvious. The results, however, seem to be too convergent with Müller's to be ignored. Two questions arise:

1. Why are the average accuracies of neural networks so similar to the accuracy of students?, and
2. Why are these accuracies so low?

However the evolution of reduplication proceeded in the Turkic languages, phonetics surely did play some role in it, at least at some stages. Presently, the method is no longer productive in Turkish. All that modern speakers have at their disposal to create new forms is the unconscious knowledge about the phonetic patterns of the language as a whole and of the existing reduplications in particular, and of the frequency and recency of their use. The neural networks knew nothing about the latter two. Yet, the results are very similar.

Unfortunately, Müller 2004 chose not to publish the exact results of all the 125 students he had interviewed, and so they cannot be compared directly to the outputs of the networks. Knowing how close to one another the mean accuracies are, however, I assume that they must have been relatively similar, too.

At first, two explanations seem to be possible: either phonetic patterns were the main factor that determined the students' answers, leaving frequency and recency of use quite irrelevant in comparison, or all these three factors just happen to coincide, and each would have produced very similar results even in isolation.

Müller 2004 only published the scores of 51, i.e. of 40.8% of the 125 interviewed students (p. 253f). If anything can be deduced from this sample, it is that the standard deviation is the lowest in corpus A and the highest in corpus C. This is just like with the neural networks, only with students, the increase is much more rapid; see tab. A.4.

Too many pieces of data are missing for the conclusions to be anything but preliminary. I should like to offer the following remarks:

- Corpus A can probably be described in purely phonetic terms in ca. 80%. (Müller's rules cover ca. 79%, see A.1.) Humans can remember exceptions better than neural networks but most of them do not know all the existing reduplications. Hence the lower, albeit still generally high, results of the networks.
- A considerable part of corpus B can be reduplicated using the phonetic patterns extracted from the existing reduplications – hence the relative consistency of the networks' answers (see tab. A.4). People are more hesitant because they had never before heard the majority of the reduplications they created during the test and neither version really “sounded good” to them – hence the visibly higher standard deviation in humans. Müller's predictions not only often went against his own rules (see A.2), but also against the students' intuition and the networks' training – hence the generally somewhat low – but still significant – accuracy.
- Corpus C seems essentially to be an intensified version of B, both as far as the students' reactions and the networks' answers are concerned.
- It seems that the role played by frequency and recency of use is not, after all, marginal. In each interviewee, however, the two factors suggest slightly different answers and on the whole, they effectively almost balance and annul themselves.

Overall, the simulation appears to indirectly confirm that it is not possible to formulate synchronic and phonetic *rules* (as opposed to *tendencies*) describing the reduplications

found in modern Turkish. This is because their evolution was not linear; be it as relics or loanwords, the contemporary stock contains products of parallel evolutionary paths which did not conform to the same set of rules.

At the same time, the simulation suggests the potential of neural networks in linguistic research and their capability to approximate the results of questionnaires while being considerably easier to control.

Indices

Semantic groupings

The list below details which reduplications were assigned to which semantic units and ranges in 3.3.1.

Where possible, units are ambiguous so as to match the entire scope of the unit. They are set in bold and, where necessary, followed by one or more specifying and explaining meanings. These are followed by the range, set in small caps, and by a list of reduplications. Where one or more of these auxiliary meanings are shared with other units, references to them are given at the end of the entry.

Those units that translate into entire phrases in English are sorted by the first word of the phrase, but references to them are also given under every adjective and noun in the translation. For example, Yak. *čegejikēn* ‘with head high on a thin neck and wide eyes’ should be looked for under “w”, but it is also referred to under “head”, “high”, “thin”, “neck”, “wide”, and “eyes”.

The classifier “oldest” refers to the oldest attestations discussed in 2.1.

Like entries, ranges are also set in bold, and additionally in small caps, and followed by a list of units they encompass.

abundant; QUANTITY; Tksh. *gür*

accurate, faithful, true, real; PERCEPTION;

Az. *doynu*, Khak. *syn*, Kirg. *yras*, Kklp.

durys, *yras*, Kmk. *gerti*, *tüz*, Kzk. *durys*,

Oir. *čike*, Tat. *čyn*, *döres*, Tksh. *döru*,

Trkm. *düz*, Uigh. *toynu*, Uzb. *toyni*

across *see* opposite

acute *see* pointed

addled *see* rotten

affluent *see* plentiful

agile, brisk, clever, sprightly; CHARACTER;

Az. *gyvrag*, Kzk. *kuwnak*, Tksh. *kyvrag*

alert, keen, sensitive, sharp-eyed, snappy;

CHARACTER; Khak. *sirgek*, Kzk. *sergek*

alien *see* foreign

alive *see* live

all, everyone, everything, whole; QUANTITY;

Az. *bütün*, Bshk. *bötön*, Kar.E *bütün*,

tamam, Kirg. *bütün*, *tügöl*, Kmk. *bütün*,

Kzk. *bütün*, *tügel*, Ott. *bütün*, *götürü*, Tksh.

bütün, Trkm. *bütün*, Uzb. *bütün*, Yak.

bütün

alone *see* lone

amiss *see* awkward

angry, furious; CHARACTER; Gag. *kызгын*

APPEARANCE bearded, blurry, clean, decor-

ated, dirty, foggy, hairy, kind, limpid,

naked, nice, rough, shapely, shiny,

- ugly, with nails/claws, with (nice) skin/complexion
- appropriate**, convenient, fit, just, very, quite; PERCEPTION; QUALITY, Bshk. *šaktaj*, Kirg. *čak*, Kklp. *šak*, Kzk. *šak*, Tksh. *ujgun*; see also worthy
- around**, surrounding; LOCATION; Gag. *čevre*, *dolaj*, Kar.E *čövre*, Ott. *čevre*, *dolajinže*, Tksh. *čevre*
- arranged** see orderly
- at night** see nighttime
- auburn** see redhead
- awkward**, amiss; PERCEPTION; Yak. *toku*, *tolös*
- awl** see pointed
- awry** see crooked
- bad**, crappy, improper, shitty, unfit; QUALITY; Bshk. *nasar*, Kirg. *žaman*, Kklp. *žaman*, Kzk. *žaman*, Tat. *načar*, Tksh. *bok*, *kötü*, Yak. *kufayan*, *negej*
- bald** see naked
- bare** see naked
- base** see with a broad edge and a narrow base
- basin** see depression
- baulk** see paralyzed
- bearded**; APPEARANCE; Kirg. *sakaldū*
- beautiful** see nice
- big**, great, huge, large; SIZE; Az. *jekä*, Kirg. *čoŋ*, Kklp. *ülken*, Kzk. *ülken*, Oir. *kozyr*, Ott. *koža*, *büjük*, Tksh. *iri*, *koža*, *kožaman*, Trkm. *ullakān*, Uzb. *katta*
- bit**, morsel; QUANTITY; SIZE, Kklp. *tomalak*
- bitter**, hot, sour; TASTE; Bshk. *äse*, Kirg. *ačū*, Tat. *äče*, Tksh. *ažy*, Trkm. *āžy*, Yak. *ahī*
- black**, dark; COLOUR; oldest *kara*, *karaju*, Az. *gara*, *garalyg*, *garanlyg*, Bshk. *kara*, *karajyy*, Dolg. *kara*, *karaja*, Gag. *kara*, Kar.E *kara*, *karavlyk*, Kar.W *kara*, Khak. *hara*, *harashy*, Kirg. *kara*, *karangy*, Kklp. *kara*, *karajyy*, Kmk. *kara*, *karangy*, Kzk. *kara*, *karajyy*, Oir. *kara*, Ott. *kara*, *sijāh*, Shor *kara*, Tat. *kara*, *karangy*, Tksh. *kara*, *karanlyk*, *koju*, *sijāh*, Trkm. *gara*, *garangky*, Tuv. *kara*, Uigh. *qara*, *qarajyy*, Uzb. *qāra*, *qārajmāq*, *qārānyi*, *sijā*, Yak. *χara*
- blind**; NATURE; Tksh. *kör*
- blue**; COLOUR; oldest *kök*, Az. *göj*, *göjlük*, Bshk. *kük*, *zāngär*, Dolg. *küök*, Gag. *māvi*, Kar.E *kök*, *mavy*, Kirg. *kök*, Oir. *kök*, Ott. *gök*, *māvi*, Shor *kök*, Tat. *kük*, *zānger*, Tksh. *gök*, *lāživert*, *māvi*, Trkm. *gök*, Tuv. *kök*, Uigh. *kök*, Uzb. *kök*, Yak. *küöχ*
- blunt**; SHAPE; Kar.E *tokal*
- blurry**, fuzzy; APPEARANCE; Tksh. *bulanyk*
- bright**, clear, light, open; COLOUR; STATE, Az. *ačyg*, *ajdyn*, Bshk. *asyk*, *jakty*, Gag. *ačyk*, *ajdynnyk*, Kar.E *ačyk*, *ačmak*, Khak. *ajas*, *aryg*, *čaryh*, Kklp. *ašyk*, *žakty*, *žeñil*, Kmk. *ačyk*, Kzk. *ašyk*, *žaryk*, Ott. *ačyk*, Tat. *ačyk*, *ajaz*, *jakty*, *žiñel*, Tksh. *ačyk*, *ajdyn*, *ajdynlyk*, Trkm. *ājdyn*, *jagty*, *jeñil*, Uigh. *aškara*, *joruq*, *süzük*, Uzb. *āčiq*, *jāruy*, *jen-gil*, Yak. *syrdyk*
- brisk** see agile
- broad** see wide, and with a broad edge and a narrow base
- brown**; COLOUR; Bshk. *horo*, *körän*, Yak. *saharχaj*
- buckle** see bulge
- bulge**, buckle, convex, full; SHAPE; Kklp. *tompak*
- by day** see daytime
- calm**, peaceful, quiet; CHARACTER; STATE, Bshk. *tynys*, Khak. *amyr*, Kirg. *žaj*, Kklp. *tynyk*, Kmk. *parahat*, Kzk. *bajsaldy*, *tynyk*, *tynyš*, *žuwas*, Tat. *tymzyk*, *tyn*, *tynyč*, Trkm. *arkajyn*, Uigh. *šük*, *teč*, *tin*, *tinč*; see also meek
- calumny** see slander
- certain** see obvious
- CHARACTER** agile, alert, angry, calm, consentient, dull, friendly, generous, happy, kind, meek, mild, quiet, rabid, right-hand, tense, vigorous, weep, well-behaved

- charge** *see* free of charge
- cheap**, inexpensive; PERCEPTION; Kzk. *arzan*, Tksh. *užuz*, Yak. *čepčeki*
- chic** *see* decorated
- chip** *see* splinter
- claws** *see* with nails, claws
- clean**, neat; APPEARANCE; oldest *aryg*, Az. *duru*, *durulug*, *tämiz*, *tämizlik*, Bshk. *tađa*, Gag. *pak*, Kar.E *temiz*, Khak. *aryg*, Kirg. *taza*, Kklp. *taza*, Kmk. *taza*, Ott. *ary*, *temiz*, Tat. *ak*, *čista*, Tksh. *temiz*, Trkm. *arassa*, *tämiz*, Uigh. *taza*, *taziliq*, Uzb. *tāza*, Yak. *seber*, *yrās*
- clear** *see* bright, limpid, *and* obvious
- clever** *see* agile
- close** *see* close friend, near, *and* respected
- close friend**; OTHER; Kirg. *ynak*
- coarse** *see* rough
- cold**; TOUCH; oldest *savuk*, Az. *buz*, Bshk. *halkyn*, Gag. *sūk*, Kirg. *muzdak*, *sūk*, Kzk. *muzdaj*, Tat. *salkyn*, Tksh. *buz*, *serin*, *sōuk*, Trkm. *sovuk*, Yak. *tymn̄y*
- COLOUR** black, blue, bright, brown, green, limpid, motley, orange, pale, pink, purple, red, redhead, white, yellow
- colourful** *see* motley
- common** *see* usual
- commotion** *see* turmoil
- complete**, entire, prepared, ready; STATE; Kirg. *dajar*, *tyjpyl*, *žylmakaj*, Kklp. *tajar*, Kmk. *gazir*, Kzk. *äzir*, *dajar*, *dajyn*, Ott. *hāzyr*, *tamām*, Tksh. *tamam*, Trkm. *tajjar*, Uzb. *šijdam*, *tajjār*, *tola*; *see also* full
- complexion** *see* with (nice) skin complexion
- confused** *see* mixed
- conical** *see* pointed
- consentient**; CHARACTER; Gag. *kajyl*
- considerable**, significant; PERCEPTION; Trkm. *esli*
- contiguous**, continuous, osculant, permanent; NATURE; oldest *tutčy*
- continuous** *see* contiguous
- contrariwise** *see* opposite
- contrary** *see* opposite
- convenient** *see* appropriate
- convex** *see* bulge
- corpulent** *see* fat
- correct** *see* direct
- cracked**, fractured; STATE; Kzk. *synyk*, Ott. *kyryk*, Tksh. *jaryk*
- crappy** *see* bad
- creased** *see* wrinkled
- crippled** *see* paralyzed
- crooked**, awry; SHAPE; Tat. *käkre*, Tksh. *ejri*
- crude** *see* raw
- cry** *see* weep
- curly**; SHAPE; Kzk. *bujra*
- cute** *see* sweet
- damp** *see* wet
- dappled** *see* motley
- dark** *see* black
- day** *see* daytime
- daytime**, by day, day; LOCATION; Tksh. *gündüz*, Uigh. *kündüz*
- deaf**; NATURE; Kmk. *sangyrav*, Yak. *žülej*
- dear** *see* respected
- decorated**, chic, fancy; APPEARANCE; Az. *bäzäk*, *bäzäkli*
- deep**; SIZE; Bshk. *tärän*, Kklp. *šukyr*, Kzk. *šuyyl*, Tksh. *derin*, Yak. *dirin*
- deletion**; OTHER; Gag. *silme*
- demolished** *see* destroyed
- dense** *see* thick
- depression**, basin; SHAPE; Kirg. *čuŋkurčak*
- desert**, lorn; LOCATION; Kirg. *ēn*, Tksh. *yssyz*
- deserving** *see* worthy
- destitute** *see* poor
- destroyed**, demolished, gone; STATE; Kirg. *žylas*
- dexter** *see* right-hand
- different**, other; PERCEPTION; Az. *bašga*, Kar.E *baška*, Tksh. *baška*, Yak. *čyğa atyn*
- difficult** *see* heavy
- direct**, correct, straight, true; LOCATION; PERCEPTION, oldest *köni*, Az. *doyru*, Bshk.

- tängül, töd, tura, tütä*, Gag. *dōru*, Kar.E *tüz*, Khak. *čike*, Kklp. *tuwra*, Kmk. *tuwra*, *tüz*, Kzk. *tuwra*, Oir. *čike*, Trkm. *gōni*, Uzb. *tikka, toʻyri*
- dirty**, foul, icky, unclean, vile; APPEARANCE; TOUCH, Az. *bulašyg*, Bshk. *byrsak*, Gag. *kirli*, Tksh. *kirli, pis*, Yak. *byrtaχ*
- disorder** *see* mixed
- disordered** *see* mixed
- dry**; TOUCH; Az. *guru*, Bshk. *koro*, Gag. *kuru*, Kar.E *kuru*, Kirg. *kurgak*, Kmk. *kuru*, Kzk. *kuryak, kuw*, Ott. *kuru*, Tat. *kory*, Tksh. *kuru*, Trkm. *gūry*, Uzb. *quruq*
- dull**, CHARACTER; PERCEPTION; Bshk. *bojok*
- durable** *see* sturdy
- early**; LOCATION; Kzk. *erte*, Tksh. *erken*
- easy**, light; NATURE; PERCEPTION, Bshk. *jejel*, Kirg. *ojoj*, Kklp. *aņsat, žejil*, Kmk. *jengil*, Kzk. *oņaj*, Tat. *žinel*, Tksh. *kolaj*, Trkm. *aņsat, jejel*, Uzb. *āsān, jengil*
- edge** *see* with a broad edge and a narrow base
- elastic**, flexible; NATURE; Bshk. *šyjk*
- empty**; STATE; Az. *boš*, Bshk. *buš*, Kar.E *boš*, Kar.W *bos ~ boš*, Kmk. *boš*, Kzk. *bos*, Ott. *boš*, Tat. *buš*, Tksh. *boš*, Trkm. *takyr*, Tuv. *kurug*, Uigh. *boš*, Uzb. *boš, quruq, šjdam*
- entire** *see* complete
- equal** *see* same
- erect** *see* vertical
- even**, level, smooth, straight; SHAPE; TOUCH, oldest *tüz*, Az. *düz*, Bshk. *šyma, takyr, tigeδ*, Gag. *düz*, Khak. *čylbyraņ, tüs*, Kirg. *syjda, tegiz, tüz, žylma, žylmakaj*, Kklp. *daņnyl, tegis, teņ*, Kmk. *tegiš*, Kzk. *tegis*, Ott. *düz*, Tksh. *düz, düzgün*, Trkm. *dogry, düz, tekiz*, Tuv. *deski*, Uigh. *tekis, tüz*, Uzb. *silliq, taqir, tekis*, Yak. *kiligir, könö*
- everyone** *see* all
- everything** *see* all
- exact** *see* just
- exit** *see* lack of exit
- explicit** *see* obvious
- extraordinary** *see* unusual
- eyes** *see* with head high on a thin neck and wide eyes, *and* with narrow eyes
- faithful** *see* accurate
- famous**, renowned, well-known; PERCEPTION; Tksh. *ünlü*
- fancy** *see* decorated
- far**; LOCATION; Kirg. *alys*, Kklp. *uzak*, Kzk. *alys*, Tksh. *yarak*
- fast** *see* quick
- fat**, corpulent, fatty, greasy, stout; SHAPE; Khak. *čōn*, Kirg. *semiz, žōn*, Kklp. *žuwan*, Kzk. *semiz, žuwan*, Tat. *simez*, Trkm. *semiz*, Yak. *suon*; *see also* squat
- fatty** *see* fat
- fear** *see* fearsome
- fearsome**, fear, formidable, horror; PERCEPTION; Kklp. *sur*, Yak. *sür*
- feeble** *see* thin
- few**, little; QUANTITY; Bshk. *äδ*, Tat. *az*
- fierce** *see* rabid
- file** *see* in file
- fine** *see* small
- firm** *see* tight
- fit** *see* appropriate
- flat**; SHAPE; oldest *jazy*, Az. *jasty*, Kirg. *tüz*, Kzk. *žalpak*, Ott. *jassy*, Tksh. *jassy*, Uzb. *japalāq*
- flexible** *see* elastic
- foggy**, misty; APPEARANCE; Yak. *budān*
- fool**, idiot; PERCEPTION; Tksh. *ibiš*
- foreign**, alien; PERCEPTION; Tksh. *jabanžy*
- forked** *see* with a broad edge and a narrow base
- formidable** *see* fearsome
- foul** *see* dirty
- fractured** *see* cracked
- free**, unimpeded, unrestrained; OTHER; Trkm. *arkajyn*, Yak. *kōņüllük*; *see also* free of charge
- free of charge**, free; OTHER; Kar.E *bedava*, Kmk. *gavajyn*, Kzk. *tegin*, Tksh. *bedāva*
- frequent** *see* thick
- fresh** *see* new, *and* vigorous

- friend** *see* close friend
- friendly**; CHARACTER; oldest *alčak*
- frozen**; STATE; Bshk. *tuj*
- full**, complete, satisfied; STATE; oldest *tolu*, *tolun*, Az. *dolu*, Bshk. *tuly*, Gag. *dolu*, Kar.E *tok*, *tolu*, Kar.W *tolu*, Kirg. *tok*, *tol-tura*, *toluk*, Kzk. *tolyk*, *tutas*, Ott. *dolu*, Tat. *tuly*, Tksh. *dolu*, *tok*, Trkm. *dōly*, Uzb. *tola*, Yak. *toloru*; *see also* bulge
- furious** *see* angry
- futile**, in vain, vain; PERCEPTION; Kar.E *bošyna*, Kirg. *beker*, Oir. *temej*, Shor *tegen*, Uzb. *quruq*
- fuzzy** *see* blurry
- generous**; CHARACTER; Tat. *jumart*, Yak. *salaj*
- gentle** *see* soft
- gingerly** *see* redhead
- glossy** *see* shiny
- gone** *see* destroyed
- good**; QUALITY; oldest *ädgü*, Bshk. *aryw*, *jakšy*, Kar.W *jahši* ~ *jahšy*, Kirg. *sonun*, *žakšy*, Kklp. *žakšy*, Kzk. *täwir*, *žaksy*, Ott. *ej*, Trkm. *gejsi*, *govy*, Tuv. *nogān*
- gradually** *see* slowly
- greasy** *see* fat
- great** *see* big
- green**; COLOUR; oldest *jašyl*, Az. *göj*, *göjlük*, *jašyl*, Bshk. *jäšel*, Gag. *ješil*, Kar.E *ješil*, Kar.W *ješil* ~ *ješil*, Kirg. *žašyl*, Kklp. *žasyl*, Kmk. *jašil*, Kzk. *žasyl*, Oir. *žažyl*, Ott. *ješil*, Tat. *jäšel*, Tksh. *jašyl*, *ješil*, Trkm. *gök*, *jāšyl*, Uigh. *ješil*, *kökläk*, Uzb. *jašil*, *kök*, Yak. *küöχ*
- grey** *see* pale
- hairy**; APPEARANCE; Gag. *tülü*
- happy**; CHARACTER; Tat. *ak*, Uigh. *saq*, Yak. *žollōχ*
- hard** *see* harsh, heavy, *and* sturdy
- harsh**, hard; PERCEPTION; Gag. *sert*, Tksh. *sert*
- hasty**, hurried; PERCEPTION; Az. *täläsik*, Kzk. *žyldam*
- healthy**, sound, unscathed; STATE; oldest *äsän*, Az. *say*, *saylam*.*saylamlyg*, *saylyg*, Bshk. *haw*, Gag. *sā*, Kar.E *saglam*, Kirg. *sō*, Kklp. *saw*, Kzk. *saw*, Tksh. *sā*, Uigh. *saq*
- head** *see* with head high on a thin neck and wide eyes
- heavy**, difficult, hard; NATURE; PERCEPTION, Kirg. *ōr*, Kklp. *awyr*, Kzk. *awyr*, Tksh. *ā(yr)*, *zor*, Yak. *yaraxan*
- hefty**, portly, stalwart; SHAPE; SIZE, Kirg. *dardaj*, Kzk. *dardaj*
- high** *see* tall, *and* with head high on a thin neck and wide eyes
- hole**, holey, hollow, leaky, pit, sunken; SHAPE; Kzk. *šunyyyl*, *šunkyr*, Yak. *teheyes*
- holey** *see* hole
- hollow** *see* hole
- holy**, saint; PERCEPTION; Yak. *sibetiej*, *yrās*
- horizontal**, lying; LOCATION; SHAPE, Kzk. *žatyk*
- horror** *see* fearsome
- hot** *see* bitter, *and* warm
- huge** *see* big
- hurried** *see* hasty
- icky** *see* dirty
- identical** *see* same
- idiot** *see* fool
- immediately** *see* now
- improper** *see* bad
- in file**, LOCATION; SHAPE; Kirg. *žanaša*
- in vain** *see* futile
- inclined** *see* sloping
- incorrect** *see* wrong
- indistinct** *see* weak
- inexpensive** *see* cheap
- insipid** *see* savourless
- itch**, mange, mangy, scab; OTHER; Tksh. *ujuz*
- jobless**, unemployed; OTHER; Tksh. *išsiz*
- juicy**, saucy; TASTE; Tksh. *žyvyk*
- just**, exact, precise; PERCEPTION; QUALITY, Bshk. *taman*, *tewāl*, Kirg. *tak*, Kzk. *däl*, Tksh. *tam*, Trkm. *dogry*; *see also* appropriate

- keen** *see* alert
- kind**, nice; APPEARANCE; CHARACTER, Kirg. *tatynakaj*
- lack of exit**; OTHER; Bshk. *tokon*
- lame**, limping; NATURE; Yak. *doyolon*
- large** *see* big
- leaky** *see* hole
- level** *see* even
- lie**; PERCEPTION; Kzk. *ötirik*, Yak. *symya*
- lifeless** *see* withered
- light** *see* bright, easy, *and* mild
- limpid**, clear, transparent; APPEARANCE; COLOUR; NATURE; oldest *süzök*, Gag. *duruk*, Khak. *aryg*, Kirg. *möldür*, Kklp. *möldir*, *tynyk*, Kzk. *möldir*, *tunyk*, Ott. *duru*, Tksh. *berrak*, *duru*, Trkm. *dury*, Uigh. *tiniq*, *süzük*, Yak. *ñeñkir*
- limping** *see* lame
- liquid**; NATURE; Az. *duru*
- little** *see* few, *and* small
- live**, alive, lively; STATE; oldest *tirig*, Az. *diri*, Gag. *diri*, Kirg. *tirü*, Ott. *diri*, Tksh. *diri*, *žanly*, Trkm. *dīri*
- lively** *see* live
- LOCATION** around, daytime, desert, direct, early, far, horizontal, in file, lone, moment, near, nighttime, noon, now, observe, outer, protruding, vertical, when, where
- log** *see* paralyzed
- lone**, alone, lonely; LOCATION; PERCEPTION; QUANTITY; Az. *jalgyz*, *jalnyz*, *jalnyzža*, Bshk. *janγyδ*, Kar.E *jalyņyz*, Kirg. *žalgyz*, Kklp. *žalyyz*, Kmk. *jangyz*, Ott. *jalnyz*, Tksh. *jalnyz*, Uzb. *jālyiz*, Yak. *soyotoχ*
- lonely** *see* lone
- long**, tall; SHAPE; SIZE, oldest *uzun*, Gag. *uzun*, Kar.E *uzun*, Khak. *uzun*, Kirg. *uzun*, Kklp. *uzun*, Kzk. *uzyn*, Ott. *uzun*, Tksh. *uzun*, Trkm. *uzyn*, Tuv. *uzun*, Uigh. *uzun*, Yak. *talarχaj*, *uhun*
- loose**, sparse; NATURE; QUANTITY, Kzk. *seldir*
- lopped**, trimmed; SHAPE; Kklp. *tokalak*
- lorn** *see* desert
- low**; SIZE; Bshk. *täläšäk*, *täpäš*, *täpäšäk*, *tübän*, Kzk. *alsa*, *tapal*
- lump-sum**; OTHER; Tksh. *götürü*
- lying** *see* horizontal
- mange** *see* itch
- mangy** *see* itch
- master**; QUALITY; Kzk. *šeber*
- mature** *see* ripe
- meek**, calm; CHARACTER; Az. *farayat*, Kklp. *synyk*, Kzk. *momyn*, *žuwas*, Tat. *juaš*
- mild**, light; CHARACTER; oldest *alčak*, Kirg. *žejil*, Kklp. *žejil*, Tat. *žijel*, Trkm. *jenil*, Uzb. *jengil*
- miserable**, wretched; STATE; PERCEPTION; QUALITY; Tksh. *perišān*, *sefil*
- misty** *see* foggy
- mixed**, confused, disorder, disordered; STATE; Az. *garyšyg*, *garyšyglyg*, Tksh. *dācynyk*, *perišān*
- moment**; LOCATION; Az. *dām*
- more so** *see* particularly
- morsel** *see* bit
- motley**, colourful, dappled, spotted; COLOUR; Bshk. *sybar*, Kzk. *ala*, *šubar*, Oir. *čokur*, Ott. *alaža*, Tat. *čuar*, Tksh. *alaža*, *renkli*, Yak. *čuoyur*, *erien*
- nails** *see* with nails, claws
- naked**, bald, bare, nude; APPEARANCE; Az. *lüt*, Bshk. *kak*, *šärä*, Gag. *čyplak*, Kar.E *čyplak*, Kirg. *takyr*, *takyrčak*, *žyłaŋač*, Kklp. *takyr*, Kmk. *takyr*, Kzk. *žalaŋaš*, Ott. *čyplak*, Tat. *jalangač*, *šärä*, Tksh. *čyplak*, *dazlak*, *kel*, *urjan*, *žavlak*, *žybyl*, Trkm. *takyr*, Uigh. *taqir*, Uzb. *jajdâq*, *jalanyâč*, *šijdam*, *taqir*, Yak. *sygyñaχ*, *tarayaj*
- narrow** *see* tight, with a broad edge and a narrow base, *and* with narrow eyes
- NATURE** blind, contiguous, deaf, easy, elastic, heavy, lame, limpid, liquid, loose, purple blind, pure, sturdy, thick
- neck** *see* with head high on a thin neck and wide eyes

- near**, close; LOCATION; Bshk. *jakyn*, Khak. *čagyn*, Kklp. *žakyn*, Kzk. *tajaw*, *žakyn*, Tat. *jakyn*, Tksh. *jakyn*, Yak. *čugas*
- neat** see clean
- nervous** see tense
- new**, fresh, recent; STATE; Az. *jeni*, *täzä*, *täzäžä*, Bshk. *jangy*, *ör-jangy*, Gag. *jeni*, Khak. *nā*, Kirg. *žany*, Kklp. *žaŋa*, Kmk. *jangy*, Kzk. *žakyn*, *žaŋa*, Oir. *žany*, Ott. *tāze*, Tksh. *jeni*, *tāze*, Trkm. *jangy*, *tāze*, Tuv. *čā*, Uigh. *jeŋi*, Uzb. *jangi*
- nice**, beautiful; APPEARANCE; PERCEPTION; QUALITY; Bshk. *matur*, Kirg. *sulū*, Kklp. *suluw*, Kzk. *ädemi*, *äsem*, Tat. *čibār*, Tksh. *güzel*, Trkm. *geŋsilik*, Uzb. *čirâjli*; see also kind
- night** see nighttime
- nighttime**, at night, night; LOCATION; Tksh. *geže*
- noon**; LOCATION; Gag. *ülen*
- now**, immediately, quickly, soon; LOCATION; Bshk. *tiđ*, Kar.E *tez*, Tuv. *dorān*, Yak. *sib-ilgin*, *sotoru*, *türgennik*
- nude** see naked
- obedient** see well-behaved
- oblique**, slanted; SHAPE; Bshk. *salyš*, Khak. *hyjyr*
- obverse**; LOCATION; Kzk. *öŋdi*
- obvious**, certain, clear, explicit, well-known; PERCEPTION; Gag. *belli*, *koža*, *kožamiti*, Kar.E *belli*, Kirg. *ačyk*, *dajyn*, *dāna*, Kklp. *anyk*, Kmk. *belgili*, Kzk. *ajkyn*, *anyk*, *begili*, Ott. *belli*, Tksh. *ašikar*, *belli*, Trkm. *belli*, Uigh. *aškara*, *očuq*, Uzb. *tajin*
- often** see thick
- old**; STATE; Gag. *koža*, *kožamiti*, Tksh. *eski*, *koža*; see also past
- only** see separate
- open** see bright
- opposed** see opposite
- opposite**, across, contrariwise, contrary, opposed, the other way round; PERCEPTION; Khak. *togyr*, Kirg. *sūk*, Ott. *karšy*, Yak. *kyžy*, *tiere*, *utary*
- orange**; COLOUR; Tksh. *turunžu*
- orderly**, arranged; STATE; Kzk. *žyjnaky*
- ordinary** see usual
- osculant** see contiguous
- other** see different
- OTHER** close friend, deletion, free, free of charge, itch, jobless, lack of exit, lump-sum, right-hand, same, separate, slowly, splinter, stone, together, turmoil, unleavened
- outer**, outsider; LOCATION; Yak. *tastyŋ*
- outsider** see outer
- pale**, grey, white, yellow; COLOUR; oldest *kök*, Az. *boz*, Khak. *hū*, *hyr*, Kirg. *boz*, *kū*, *sur*, Kklp. *boz*, *kuw*, *sur*, Kzk. *kuw*, *sur*, Oir. *boro*, Tat. *ak*, *sory*, Tksh. *boz*, Trkm. *čāl*
- paralyzed**, baulk, crippled, log, tree-stump; STATE; Tksh. *kütük*, *kötürüm*
- particularly**, more so; PERCEPTION; Trkm. *beter*
- past**, old; STATE; Kirg. *kačanky*, Kzk. *bajayy*
- peaceful** see calm
- PERCEPTION** accurate, appropriate, awkward, cheap, considerable, different, direct, dull, easy, famous, fearsome, fool, foreign, foolish, harsh, hasty, heavy, holy, just, lie, lone, miserable, nice, obvious, opposite, particularly, poor, proper, quick, respected, serious, soft, sudden, thin, unexpected, unpleasant, unusual, usual, weak, wise, worthy
- permanent** see contiguous
- pink**; COLOUR; Tksh. *pembe*, Yak. *tetekej*
- pit** see hole
- plain** see pure
- plentiful**, affluent; QUANTITY; Yak. *delej*
- pockmarked** see rough
- pointed**, acute, awl, conical, sharp; SHAPE; Az. *biz*, Kzk. *šošak*, *süjir*, Ott. *sivri*, Tksh. *sivri*, Yak. *syty*
- polite** see soft
- ponderous** see serious

- poor**, destitute; PERCEPTION; Az. *joxsul*, *joxsullug*, Yak. *žadany*
- portly** *see* hefty
- precise** *see* just
- prepared** *see* complete
- proper**; PERCEPTION; QUALITY; Khak. *orta*; Kirg. *tūra*, Kzk. *durys*, Yak. *könö*
- protruding**, LOCATION; SHAPE; oldest *kötgi*
- puny** *see* thin
- purbblind**, with narrow eyes; NATURE; Yak. *simigir*
- pure**, plain, simple; NATURE; Ott. *sāfi*, Tksh. *sāde*
- purple**, violet; COLOUR; Gag. *mor*, Kar.E *mor*, Ott. *mor*, Tksh. *mor*
- QUALITY** appropriate, bad, good, just, master, miserable, nice, proper, rogue, slander, thin, unusual, usual, weak, worse, wrong
- QUANTITY** abundant, all, bit, few, lone, loose, plentiful, separate, sparse, thick
- quick**, fast; PERCEPTION; Kar.E *čebik*, Kklp. *tez*, Kzk. *tez*, *žyldam*, Ott. *čabuk*, Tuv. *türgen*
- quickly** *see* now
- quiet**, silent; CHARACTER; STATE, Bshk. *äkren*, *šym*, *tymyk*, *tyn*, Khak. *symsyryh*, Uigh. *šük*, *tik*; *see also* calm
- quite** *see* appropriate
- rabid**, fierce, wicked; CHARACTER; Yak. *kütür*
- rare** *see* sparse
- raw**, crude; STATE; Bshk. *sej*, Kzk. *šijki*, Tat. *či*, *jüeš*, Tksh. *čij*, Yak. *sikej*
- ready** *see* complete
- real** *see* accurate
- recent** *see* new
- red**; COLOUR; oldest *kyrmyzy*, *kyzyl*, Az. *gyrmyzy*, Bshk. *kyzyl*, Dolg. *kyhyl*, *kytarkaj*, Gag. *kyrmyzy*, Kar.E *al*, Kar.W *kyzyl*, Khak. *hyzyl*, Kirg. *kyzyl*, Kklp. *kyzyl*, Kmk. *kyzyl*, Kzk. *kyzyl*, Oir. *kyzyl*, Ott. *kyrmyzy*, *kyzyl*, Shor *kyzyl*, Tat. *kyzyl*, Tksh. *kyrmyzy*, *kyzyl*, Trkm. *gyrmyzy*, Tuv. *kyzyl*, Uigh. *qizil*, Uzb. *qizarmâq*, *qizil*, Yak. *kugas*, *kyhyl*
- redhead**, auburn, gingerly, sorrel; COLOUR; Kzk. *žijren*, Yak. *čačarçaj*
- regular** *see* shapely
- renowned** *see* famous
- respected**, close, dear; PERCEPTION; Trkm. *gadyrly*, Yak. *ytyk*
- right-hand**, dexter, right; CHARACTER; OTHER, Tksh. *sā*
- right** *see* right-hand
- rip** *see* splinter
- ripe**, mature; STATE; Tksh. *olgun*
- rogue**, swindler, thief; QUALITY; Yak. *tüökün*
- rolled**; SHAPE; Tksh. *dürü*
- rotten**, addled; STATE; Bshk. *serek*, Ott. *čürük*, Tksh. *čürük*, *žylk*, Yak. *sytygan*
- rough**, coarse, pockmarked, uneven; APPEARANCE; TOUCH, Az. *kobud*, Tat. *kytyršy*, Uzb. *čötir*
- round**; SHAPE; oldest *tägirmä*, Az. *girdä*, *jumru*, *jumruža*, Bshk. *jomro*, *tumalak*, *tünjäräk*, Dolg. *tögürük*, Kar.E *jumarlak*, *tomalak*, *tögerek*, Khak. *tērpek*, *tigilek*, *toglah*, Kirg. *döngölök*, *tegerek*, *žumuru*, Kklp. *domalak*, *tomalak*, *žumyry*, Kzk. *domalak*, *žumyr*, Ott. *degirmi*, *top*, Tat. *či*, *jomry*, *tügäräk*, Tksh. *dejirmi*, *jumru*, *juvarlak*, *tekerlek*, *top*, *toparlak*, Trkm. *tegelek*, *togalak*, Tuv. *borbak*, *tögerik*, Uigh. *jumulaq*, Uzb. *dumalâq*, Yak. *tögürük*
- saint** *see* holy
- salty**; TASTE; Tksh. *tuzlu*, Yak. *tüstāχ*
- same**, equal, identical, similar; OTHER; Bshk. *tiŋ*, Khak. *tiŋ*, *tōj*, Kirg. *birdej*, *okšoš*, *teŋ*, Oir. *teŋ*, *tüŋej*, Tat. *tiŋ*, Tksh. *ajny*, Trkm. *deŋ*, Tuv. *deŋ*, Uigh. *baravär*, *täŋ*, Uzb. *baravar*
- satisfied** *see* full
- saucy** *see* juicy
- savourless**, insipid; TASTE; Yak. *ñulun*
- scab** *see* itch
- scrawny** *see* thin

sensitive *see* alert

separate, only, singular; QUANTITY; OTHER, Kirg. *žeke*, Tksh. *ajry*

serious, ponderous; PERCEPTION; Kzk. *sal-makty*

settled *see* still

shallow; SHAPE; SIZE; Bshk. *haj*; Yak. *čyčās*; *see also* small

SHAPE blunt, bulge, crooked, curly, depression, even, fat, flat, hefty, hole, horizontal, in file, long, lopped, oblique, pointed, protruding, rolled, round, shallow, slender, sloping, slouching, squat, standing up, tall, thin, tight, vertical, wide, with a broad edge and a narrow base / split / forked, with head high on a thin neck and wide eyes

shapely, regular; APPEARANCE; Tksh. *düzgün*

sharp-eyed *see* alert

sharp *see* pointed

shiny, glossy, sparkling; APPEARANCE; Bshk. *jaltyr, jyltyr*, Gag. *jalabyk*, Kzk. *žyltyr*, Ott. *sylrak*, Yak. *kilekij*

shitty *see* bad

short; SIZE; Bshk. *kyθka, nakyθ, tokor, tölöš*, Kklp. *kyska*, Kzk. *šolak*, Tat. *kyska*, Tksh. *kysa*

significant *see* considerable

silent *see* quiet

similar *see* same

simple *see* pure

singular *see* separate

SIZE big, bit, deep, hefty, long, low, shallow, short, slender, small, tall, thin, tight, wide

skin *see* with (nice) skin complexion

skinny *see* thin

slander, calumny; QUALITY; Yak. *χobū*

slanted *see* oblique

slender, svelte, tall; SHAPE; SIZE, Yak. *χožoyor*

slim *see* thin

sloping, inclined; SHAPE; Kzk. *žatyk*

slouching; SHAPE; Bshk. *bökrö*

slowly, gradually; OTHER; Khak. *agrin*

small, fine, little, shallow; SIZE; Az. *balaža*, Bshk. *bäläkāj*, Dolg. *kyra*, Kirg. *kičine*, Kklp. *kiškene, kiškentaj*, Kzk. *kiškentaj, majda, šayyn, usak*, Tksh. *küçük, ufak*, Yak. *naryn, oččuguj, kyra*

smooth *see* even

snappy *see* alert

sob *see* weep

sober; STATE; Khak. *sah*

soft, gentle, polite, suave, subtle, tender; PERCEPTION; TOUCH, oldest *jumšak*, Az. *jumšag*, Bshk. *jomšak*, Kirg. *sylyk, synyk, žylma, žumšak*, Kklp. *synyk, žumsak*, Kzk. *žumšak*, Tat. *juaš*, Tksh. *jumušak*, Yak. *symnayas*

solid *see* sturdy

soon *see* now

sorrel *see* redhead

sound *see* healthy

sour, tart; TASTE; Az. *turš, turšlug*, Tksh. *ekši*, Trkm. *turšy*, Yak. *ahy*; *see also* bitter

spacious *see* wide

sparkling *see* shiny

sparse, rare; QUANTITY; Bshk. *hiräk*; *see also* loose

splinter, chip, rip; OTHER; Az. *söküntü*, Gag. *sökük*

split *see* with a broad edge and a narrow base

spotted *see* motley

sprightly *see* agile

squat, fat, stubby; SHAPE; Kirg. *balpak*, Tksh. *güdük*

stagnant *see* still

stale; STATE; Tksh. *bajat*, Yak. *nijsik*

stalwart *see* hefty

standing up; SHAPE; Tuv. *tura*

STATE bright, calm, complete, cracked, destroyed, empty, frozen, full, healthy, live, miserable, mixed, new, old, orderly, paralyzed, past, quiet, raw, ripe, rotten, sober, stale, still, thawed, withered, young

steep *see* vertical

- still**, settled, stagnant; STATE; Bshk. *tonok*, Tksh. *durgun*, Uzb. *tin*
- stone**; OTHER; Kzk. *tastaj*
- stout** *see* fat
- straight** *see* direct, *and* even
- strong** *see* sturdy
- stubby** *see* squat
- sturdy**, durable, hard, solid, strong, tough; NATURE; TOUCH, Bshk. *katy*, Gag. *katy*, Khak. *tyŋ*, Kklp. *katty*, Kzk. *berik*, *katty*, *nyk*, Ott. *katy*, Tat. *katy*, *taza*, Tksh. *katy*, *sālam*, *topač*, Trkm. *dajav*, Uigh. *küčlük*
- suave** *see* soft
- subtle** *see* soft
- sudden**; PERCEPTION; Gag. *ansyz &c.*, Kar.E *ansyz &c.*, *kenete*, Kar.W *keñeŋa* ~ *kenete*, Khak. *kinetŋ*, Ott. *ansyz &c.*, Tksh. *ansyz &c.*, Tuv. *henerten*, Uzb. *tosatdan*, *tosindan*
- sunken** *see* hole
- surrounding** *see* around
- svelte** *see* slender
- sweet**, cute; TASTE; oldest *süčig*, Az. *širin.širinlik*, *šit*, *šitlik*, Khak. *tadylyg*, Kirg. *širin*, Kzk. *tätti*, Ott. *tatly*, Tksh. *širin*, *tatly*, Trkm. *sūži*
- swindler** *see* rogue
- tall**, high; SHAPE; SIZE, Kirg. *bijik*, Kklp. *bijik*, Kzk. *bijik*, Oir. *uzun*, Yak. *ürdük*; *see also* long, *and* slender
- tart** *see* sour
- TASTE** bitter, juicy, salty, savourless, sour, sweet, tasty, unpalatable
- tasty**; TASTE; Kzk. *dämdi*, Tksh. *tatly*
- tender** *see* soft
- tense**, nervous, tight; CHARACTER; Tksh. *gergin*
- tepid** *see* warm
- thawed**; STATE; Yak. *ūllayas*
- the other way round** *see* opposite
- thick**, dense, frequent, often; NATURE; QUANTITY, Bshk. *kujy*, Kirg. *kojū*, Kklp. *kojyw*, *šymyr*, *žuwān*, Kzk. *tyyyz*, *žiji*, Tksh. *gür*, *jō(un)*, *kalyn*, *koju*, *syk*, *tykyz*, Yak. *χojū*
- thief** *see* rogue
- thin**, feeble, puny, scrawny, skinny, slim, weak; SHAPE; SIZE; PERCEPTION; QUALITY, Bshk. *joka*, *kak*, *nādek*, *neskā*, *šyjk*, Kar.E *aryk*, *inčće*, Kar.W *aryh*, *inčka* ~ *inčće*, Kirg. *ičke*, *žuka*, Kklp. *žuka*, Kzk. *aryk*, *seldir*, *žuka*, Tat. *aryk*, Tksh. *inže*, *syska*, *žylyz*, Tuv. *čingē*, Yak. *čarās*, *kuhayan*, *naryn*, *sīhes*, *siñiges*; *see also* with head high on a thin neck and wide eyes
- tight**, firm, narrow; SHAPE; SIZE, Bshk. *tar*, *tyyyδ*, Kzk. *tar*, *tyyyz*, *žijiške*, Ott. *kyv-rak*, *syky*, Tat. *tygyz*, Tksh. *dar*, *syky*, Yak. *yksary*; *see also* tense
- together**; OTHER; Ott. *berāber*, Tksh. *berāber*
- TOUCH** cold, dirty, dry, even, rough, soft, sturdy, warm, weak, wet, wrinkled
- tough** *see* sturdy
- transparent** *see* limpid
- tree-stump** *see* paralyzed
- trimmed** *see* lopped
- true** *see* accurate, *and* direct
- turmoil**, commotion, uproar; OTHER; Uzb. *topālān*
- ugly**; APPEARANCE; Tksh. *čirkin*
- unclean** *see* dirty
- unemployed** *see* jobless
- uneven** *see* rough
- unexpected**; PERCEPTION; Yak. *deŋ*
- unfit** *see* bad
- unimpeded** *see* free
- unleavened**; OTHER; Bshk. *sösö*
- unpalatable**; TASTE; Bshk. *tāmheδ*
- unpleasant**; PERCEPTION; Kirg. *sūk*
- unrestrained** *see* free
- unscathed** *see* healthy
- unusual**, extraordinary, wonderful; PERCEPTION; QUALITY, Yak. *žikti*
- upright** *see* vertical
- uproar** *see* turmoil
- usual**, common, ordinary; PERCEPTION; QUALITY, Tksh. *bajāγ*
- vain** *see* futile

- vast** *see* wide
- vertical**, erect, steep, upright; LOCATION; SHAPE, Az. *dik*, Bshk. *tekä*, Kar.E *tyk*, Kirg. *tik*, Kklp. *tik*, *tikke*, Kzk. *tike*, Tksh. *dik*, Yak. *bosxo*
- very** *see* appropriate
- vigorous**, fresh; CHARACTER; Tksh. *dinč*
- vile** *see* dirty
- violet** *see* purple
- warm**, hot, tepid; TOUCH; Bshk. *jyly*, Kzk. *žyly*, Tat. *žyly*, Tksh. *syžak*, *ylyk*, Trkm. *yssy*, Yak. *čylās*, *itī*
- weak**, indistinct, worn; PERCEPTION; QUALITY; TOUCH; Tksh. *silik*, *zajyf*; *see also* thin
- weep**, cry, sob; CHARACTER; Yak. *sojū*
- well-behaved**, obedient; CHARACTER; Kzk. *momakan*, *žuwas*, Tksh. *uslu*
- well-known** *see* famous, *and* obvious
- wet**, damp; TOUCH; Bshk. *jeüeš*, Gag. *jaš*, Kar.E *sylak*, Ott. *jaš*, *syklam*, Tat. *jüeš*, Tksh. *jaš*, *yslak*, Yak. *niččeyej*
- when**; LOCATION; Kirg. *kačan*, *kačanky*
- where**; LOCATION; Kirg. *kajdagy*
- white**; COLOUR; oldest *ak*, *jürüj*, Az. *ay*, Bshk. *ak*, Dolg. *ürüj*, Gag. *ak*, *bijaz*, Kar.E *ak*, *bijaz*, Kar.W *ah* ~ *ak*, Kirg. *ak*, *apakaj*, *appak*, Kklp. *ak*, Kmk. *ak*, Kzk. *ak*, Oir. *ak*, Ott. *ak*, *bejaz*, Shor *ak*, *apagaš*, Tat. *ak*, Tksh. *ak*, *bejaz*, Trkm. *āk*, Tuv. *ak*, Uigh. *aq*, Uzb. *âq*, Yak. *mañan*, *ürüj*, *ürüññük*; *see also* pale
- whole** *see* all
- wicked** *see* rabid
- wide**, broad, spacious, vast; SHAPE; SIZE, Kirg. *keñ*, *kenen*, Tksh. *geniš*, Trkm. *gīj*, Yak. *ketit*, *kieñ*, *tenigir*; *see also* with head high on a thin neck and wide eyes
- wise**; PERCEPTION; Kirg. *keñ*, *kenen*
- with (nice) skin, complexion**; APPEARANCE; Kzk. *öñdi*
- with a broad edge and a narrow base**, split, forked; SHAPE; Yak. *tereger*
- with head high on a thin neck and wide eyes**; SHAPE; Yak. *čegejikēn*
- with nails, claws**; APPEARANCE; Kirg. *tyr-maktaj*
- with narrow eyes** *see* purblind
- withered**, lifeless; STATE; Tksh. *ölgün*
- wonderful** *see* unusual
- worn** *see* weak
- worse**; QUALITY; Az. *betär*, Kar.E *beter*, Tksh. *beter*
- worthy**, appropriate, deserving; PERCEPTION; Uzb. *lâjiq*
- wretched** *see* miserable
- wrinkled**, creased; TOUCH; Tksh. *buruš*, *bur-ušuk*
- wrong**, incorrect; QUALITY; Tksh. *janlyš*
- yellow**; COLOUR; oldest *saryg*, Az. *sary*, *sarylyg*, Bshk. *hary*, Gag. *sary*, Kar.E *sary*, Kar.W *sary*, Khak. *saryg*, Kirg. *sary*, Kklp. *melle*, *sary*, Kmk. *sari*, Kzk. *sary*, Oir. *sary*, Ott. *sary*, Shor *saryg*, Tat. *sary*, Tksh. *sary*, Trkm. *sāry*, Tuv. *saryg*, Uigh. *seriq*, Uzb. *sariq*, Yak. *arayas*, *saharçaj*; *see also* pale
- young**; STATE; Bshk. *jäš*, Kirg. *žäš*, Kklp. *žas*, Kzk. *žas*, Tat. *jäš*, Tksh. *genč*, Uigh. *jaš*, Uzb. *jangi*, Yak. *eder*

Subject index

The followig index only contains a selection of the subjects discussed in the present book. It generally omits subjects that appear so often that the list of pages would have to be of comparable length to the list of pages where they are not mentioned. These are such subjects as *adjective*, *anlaut*, *closer*, *consonant*, *intensification*, *reduplication*,

semantics, Turkish, &c. Some of them, however, are included in the way of holders for subentries, so that e.g. *multiple intensification* does not need be looked for under “m”. Also words that appear in the translations of examples are not included. Entries for languages only contain selected pages and generally omit those where just one or two examples from the given language are mentioned. Likewise, non-Turkic languages are not listed, except for “Mongolic” and “Tungusic”.

A

Abdal, 12
 accent, *see* stress
 adverb, 1, 3, 5, 8, 68, 71, 89, 101, 102, 104, 108, 122, 139, 145, 155, 217, 255, 259
 alliteration, 8, 20, 137, 150
 Altaic, 9, 186, 187, 198, 215, 230–235, 237, 244, 247, 248, 250, 253, 255, 260, 261
 anlaut
 emancipated, 24, 38, 77, 123, 136, 137, 151, 171–173, 237, 259
 nonmatching, *see* head, nonmatching
 appending, 3, 73, 213, 255
 areal influence, 47, 52, 54, 55, 131, 158, 180, 184, 188, 195, 198, 202, 212, 215, 233, 234, 238–244, 247, 248, 251, 252, 255, 257, 260, 261
 Armeno-Kipchak, 12, 55
 assimilation, 25, 61, 99, 152, 158, 163, 164, 167, 235, 236, *see also* simplification
 auslaut, 4, 5, 31, 152, 214, 264
 Azeri, 27–34, 42, 45, 82, 108, 119, 121, 136, 145, 159, 161, 164, 169, 177, 178, 181, 183, 186, 187, 190, 193, 196, 200, 202–205, 207–209, 212, 213, 215, 217, 221, 224, 238, 252, 252, 253, 253, 257

B

base
 non-adjective, *see* adverb; noun; pro~~n~~
 noun; verb

 obsolete, 53, 60, 65–67, 71, 118, 137, 143, 161, 174, 176, 177, 258, 259, 259
 prepended, 27, 31, 55, 85–88, 175, 177, 178, 181
 Bashkir, 34–39, 159, 161, 169, 171, 183, 184, 187, 189, 190, 193–195, 210, 226, 248, 251, 260
 binom, *see* compound
 borrowing, 9, 24, 31, 44–47, 51, 54, 55, 57, 60, 65, 66, 80, 82, 83, 119, 120, 135, 139, 150, 158, 160, 164, 168, 204, 210–212, 216, 228, 229, 234, 237, 254, 258, 259, 272

C

calquing, 228, 229, 254
 Chaghatai, 7, 12, 244, 254
 Chulym, 12
 Chuvash, 195, 211, 212, 216, 239, 241
 closer
 alternative, 10, 17, 18, 26, 33, 34, 38, 39, 46, 56, 61, 78, 83, 89, 100, 107–109, 121, 122, 126, 138, 143, 145, 153, 154, 158–160, 162, 163, 165, 180, 182, 189, 196, 201, 206, 207, 213, 228, 245, 257
 diversification, 7, 9, 20, 48, 52, 82, 145, 153, 182–186, 202–213, 231, 237, 238, 247–255, 257, 258, 260, 261
 peculiar, 33, 39, 46, 51, 56, 60, 67, 82, 83, 121, 138, 144, 151, 153, 161–168, 180, 181
 pp, 4, 6, 19, 20, 26, 38, 39, 48, 56, 57, 70, 78, 83, 85, 86, 88, 89, 95, 107, 138, 140, 143, 145, 152–154, 159,

- 169–171, 174–176, 180, 182, 217, 236, 245
- rules, 7–11, 20, 48, 90, 107–110, 161, 164–166, 181, 184, 188–191, 207, 208, 213, 215, 227, 228, 253, 255–258, 260, 261, 263–268, 270, 271, *see also* reduplication, description
- type, 168
- types, 3, 5, 6, 131, 150, 154, 231
- vs. C_1 , 4, 5, 161–166, 189, 190, 257
- vs. C_2 , 4, 5, 26, 33, 39, 53, 67, 83, 121, 131, 143, 144, 150, 164–168, 189, 190, 207, 245, 257
- closing consonant, *see* closer
- cluster
- clustering, 242, 243
 - consonantal, 10, 87, 88
 - geographical, 202, 254, 255
- coda, *see* auslaut
- colour, 7, 20, 27, 68, 114, 133, 140, 196, 198, 216, 218–220, 222, 226, 227, 235, 240, 242, 247, 260, 275
- composition, *see* compound
- compound, 5, 20, 25, 31, 32, 38, 41, 45, 54, 85, 87, 88, 118, 119, 119, 125, 135–138, 152, 160, 171–173, 175, 177, 178, 178, 182, 235, 259, *see also* base, prepended
- contraction, *see* simplification
- correlation, 13, 110, 180, 194, 224, 227, 234, 239, 242, 260, 265, 270
- Crimean Tatar, 12, 45, 54, 55
- D**
- derivative, 27, 31–34, 36, 45, 47, 52, 56, 66, 77, 83, 102, 137, 144, 150, 151, 173, 174, 229, 230, 232, 252, 253
- diachrony, 1, 12, 187, 191, 223, 228–230, 253–256, 260
- diminutive, 25, 31, 32, 70, 72, 77–79, 85–87, 103, 118, 168, 169, 175, 176, 182, 217
- diphthong, 41, 145, 154, 179, 180, 214
- distance, 198, 201, 215, 223, 243, 260, 268, 269
- Dolgan, 39–42, 98, 125, 145, 161, 175, 179, 183, 186, 193, 195, 197, 198, 203, 234, 238, 244, 248, 250, 252, 254, 260
- doubling, 3, 5, 6, 8, 20, 34, 38, 72, 76, 79, 86, 102, 168–172, 174, 175, 181, 214, 235, 255, 256, *see also* closer, *pp*; lengthening; intensification, multiple; reduplication, multiple
- E**
- Eastern Karaim, 203
- elision, *see* simplification
- enclitic, 235
- entropy, 182–188, 202, 205, 209, 232, 245, 250–252
- erosion, *see* intensification, erosion
- evolution, *see* reduplication, evolution; semantic evolution
- F**
- Fuyu, 12
- G**
- Gagauz, 42–47, 122, 159–162, 164, 183, 187, 190, 193, 196, 203, 208, 209, 238, 252, 253, 257
- genealogy, *see* inheritance
- H**
- head, 4, 5, 41, 214, 215
- nonmatching, 23, 41, 52–54, 61, 95, 97, 98, 100, 107, 118, 135, 136, 138, 151, 178, 179, 213, 250, 259
- I**
- influence, *see* areal influence
- inheritance, 45, 57, 101, 158, 169, 170, 184, 187, 188, 190, 191, 194, 198, 200, 202–204, 208–212, 215, 216, 219, 221, 222, 227–230, 233, 234, 236–244, 247, 248, 250, 254, 255, 257, 260, 261

intensification

- erosion, 32, 38, 78, 103, 170
- multiple, 25, 78, 79, 87, 131, 168, 169, 174–176, 178, 181, 182, 251, 259, *see also* base, prepended; doubling; diminutive; lengthening; reduplication, multiple

K

- Karaim, 18, 45, 47–57, 158, 159, 161, 169, 177, 178, 180, 181, 183, 187, 188, 193–195, 203, 204, 210, 248, 251, 260
 - Eastern, 162, 164, 181, 184, 190, 203, 208, 209, 211, 252, 260
 - Western, 158, 184, 189, 210
- Karakalpak, 57–61, 65, 161, 169, 171, 178, 179, 183, 187, 193–195, 213, 236, 251, 256
- Kazakh, 60–67, 121, 161, 169, 171, 183, 187, 193–195, 200, 221, 236, 251, 252, 260
- Khakas, 68–71, 83, 85, 102, 103, 127, 161, 172, 182, 183, 186, 193, 196, 236, 239, 253
- Khalaj, 12, 241
- Kirghiz, 71–79, 160, 161, 169, 171, 172, 175, 178, 183, 187, 193–195, 239, 251, 260
- Kumyk, 33, 80–83, 121, 158, 161, 169, 171, 172, 183, 186, 187, 189, 193–195, 202, 210, 248, 251, 260

L

- lengthening, 6, 20, 41, 47, 70, 83, 86–88, 90, 98, 100, 140, 145, 151, 152, 154, 159, 169, 170, 174–176, 180, 182, 231, 235, 236, 245, 252, *see also* closer, *pp*; doubling
- loanword, *see* borrowing

M

- Mongolic, 4, 40, 131, 145, 160, 180, 181, 186, 230–234, 236–238, 244, 247, 248, 253

- mora, 1, 3, 5, 98, 214–216, 231, 255, 260
- multidimensional scaling, 201, 223, 243

N

- neural network, 269, 270
- nomina, *see* noun
- noun, 5, 16, 26, 27, 33, 39, 57, 61, 67, 71, 72, 79, 101, 107, 108, 119, 122, 125, 127, 132, 136, 139, 144, 145, 152, 155, 173, 176, 182, 217, 235, 259, 273

O

- Oghuz, 252, 253
- Oirot, 68, 70, 71, 83–89, 102, 103, 161, 169, 177, 178, 182, 183, 186, 193, 196, 253
- onset, *see* anlaut
- Ottoman, 7, 47, 89–101, 122, 158, 159, 161, 162, 164, 169, 174, 175, 183, 184, 186–188, 190, 203, 204, 208, 209, 245–248, 251–254, 261

P

- particle, 7, 8, 24, 80, 90, 125, 169, 170, 173, 226, 236, 237
- pluralization, 26, 27, 33, 136, 176, 217, 232, 252
- prefix, 9, 20, 46, 68, 69, 90, 104, 133, 215, 233
- prepending, 1, 3, 4, 88, 96, 102, 168, 177, 178, 213, 231, 255, *see also* base, prepended
- pronoun, 72, 77, 78, 150, 178, 217, 251

R

- reduplication
 - apparent, 25, 26, 45, 60, 65, 87, 96, 117–120, 131, 135–138, 150, 160, 161, 259
 - description, 3–11, 20, 21, 27, 28, 34, 35, 40, 42, 43, 48, 62, 68, 72, 73, 80, 84, 90, 102, 104, 108–110, 123, 127, 132, 133, 139, 140, 145, 146,

- 150, 211, 213–215, 255, 256, *see also* closer, rules
- diversification, 182–188, 191–212, 237, 238, 249–253
- evolution, 184, 186–188, 195, 198, 201–203, 215, 235–248, 254, 255, 257, 260, 261, 271, 272
- family, 32, 41, 65–68, 70, 71, 83, 85, 86, 88, 96, 103, 137, 145, 152, 169, 173–175, 177, 182, 192, 196, 253, 259
- multiple, 76, 103, 150, 171, *see intensification*, multiple
- origin, 235–238, 247, 248, 253, 254, 260
- types, *see* closer, types
- repetition, *see* doubling
- S**
- Salar, 12
- semantic evolution, 42, 47, 61, 66, 67, 77–79, 83, 89, 102, 104, 108, 122, 127, 131, 139, 144, 154, 159, 160, 217, 221–224
- Shor, 70, 83, 85, 86, 88, 102–104, 171, 182, 183, 193, 196, 253
- shortening, 5, 8, 41, 42, 46, 47, 71, 78, 85, 97, 98, 100, 108, 109, 121, 123, 125, 126, 132, 135, 145, 154, 178–180, 213, 231, 235, 236, 255, 256
- simplification, 6, 53, 56, 69, 85–88, 96, 131, 169, 170, 176, 231, 235, 258, *see also* assimilation
- stress, 8, 34, 73, 89, 97, 100, 104, 120, 135, 140, 170, 235, 237, 264
- substantive, *see* noun
- superlative, 8, 36, 60, 84
- syllable, 4, 5, 7, 7, 8, 10, 34, 40, 42, 52, 53, 62, 68, 72, 73, 84, 85, 90, 100, 102, 120, 132, 138, 143, 145, 150, 177, 190, 207, 213, 214, 216, 231, 233, 235, 237, 255, 260, 263, 264
- mono-, 24, 100, 104, 171, 189, 207, 208, 212, 216, 233, 256, 264, 268
- poly-, 69, 70, 89, 96, 100, 170, 189, 207, 208, 216, 237
- synchrony, 1, 9, 12, 108, 109, 177, 188–191, 213, 227, 228, 231, 249–253, 256, 258, 260, 264, 266, 271
- T**
- tail, 4, 5, 214, 215
- Tatar, 104–107, 159–161, 169, 183, 184, 187, 189, 190, 193–195, 200, 210, 221, 226, 248, 251, 260
- Tofalar, 12
- Tungusic, 39, 40, 186, 195, 230, 232–234, 237, 238, 244, 247, 248, 252, 253
- Turkish, 203
- Turkmen, 25, 122–127, 159, 161, 164, 172, 183, 184, 186, 187, 189, 193, 196, 202–209, 212, 213, 215, 224, 238, 252, 253, 257
- Tuvinian, 68, 83, 102, 127–132, 161, 169, 181, 183, 186, 193, 196, 249, 253
- U**
- Uighur, 19, 132–139, 159, 161, 169, 172, 178, 179, 181, 183, 186, 188, 192, 193, 202, 239, 250, 251, 256
- umlauting, 132, 133, 136, 138, 178, 179, 250
- Urum, 55, 211
- Uzbek, 132, 139–145, 159–161, 164, 169, 171, 172, 181, 183, 184, 187–190, 192, 193, 202–209, 213, 248, 250, 251
- V**
- verb, 48, 57, 67, 128, 128, 129, 131, 132, 139, 144, 144, 145, 177, 181, 217, 250, 251, 253, 265, 268
- W**
- Western Yughur, 12
- Y**
- Yakut, 7, 39–41, 98, 108, 125, 145–155, 158, 159, 161, 163, 164, 169, 171, 173, 175, 179, 182–184, 186–190, 193,

195, 197, 198, 200, 203, 208–210, Yughur, *see* Western Yughur
 213, 215, 222–224, 227, 231, 232,
 234, 238, 240, 244, 248, 250, 252,
 254–256, 260, 261

Symbols and abbreviations

~ = alternation | → = word formation | > = change or borrowing | ≧ = bifurcation | Ø = contamination | ≡ = without change | :: = affinity | **Ar.** = Arabic | **arch.** = archaic | **Az.** = Azeri | **Bon.** = Bonan | **Bshk.** = Bashkir | **Bur.** = Buriat | **child.** = children's | **Chuv.** = Chuvash | **CMo.** = Common Mongolic | **CTat.** = Crimean Tatar | **Dag.** = Dagur | **dial.** = dialectal | **DIMIN.** = diminutive | **Engl.** = English | **Germ.** = German | **honor.** = honorific | **INF.** = infinitive | **intens.** = intensification | **KB** = *Kutadgu Bilig* | **Khor.** = Khorchin | **Kirg.** = Kirghiz | **Kklp.** = Karakalpak | **Klmk.** = Kalmuk | **Kmk.** = Kumyk | **Krm.** = Krymchak | **Kzk.** = Kazakh | **lit.** = literally | **Ma.** = Manchu | **Mgh.** = Moghol | **MMo.** = Middle Mongol | **MPers.** = Middle Persian | **Mng.** = Monguor | **Mo.** = Mongolic | **Nog.** = Nogai | **Nor.** = Norwegian | **NSib.** = North Siberian | **Oir.** = Oirots | **Ord.** = Ordos | **OTkc.** = Old Turkic | **Ott.** = Ottoman | **Pers.** = Persian | **PMo.** = Proto-Mongolic | **Pol.** = Polish | **PTkc.** = Proto-Turkic | **redup.** = reduplication | **Russ.** = Russian | **Sag.** = Sagal | **San.** = Santa | **SSib.** = South Siberian | **SYug.** = Shira Yughur | **Tat.** = Tatar | **Tksh.** = Turkish | **Trkm.** = Turkmen | **Tuv.** = Tuvian | **Uigh.** = Uighur | **Uzb.** = Uzbek | **WMo.** = Written Mongolian | **WYug.** = Western Yughur | **Xlx.** = Khalkha | **Yak.** = Yakut

References

- Abdülbâkî 1934, Yunus Emrede öz türkçe kelimeler. – *Türkiyat Mecmuası* 4, 265–277
- Abduldaev et al. 1987 = Абдулдаев Э.А. et al. (ed.), *Грамматика киргизского литературного языка*. 1: *Фонетика и морфология*, Фрунзе
- Adamović M. 2001, *Das Türkische des 16. Jahrhunderts. Nach den Aufzeichnungen des Florentiners Filippo Argenti (1533)* (= *Materialia Turcica: Beiheft* 14), Göttingen
- Ahmerov 1958 = Ахмеров К.З., Краткий очерк грамматики башкирского языка – BškRS58: 743–802
- Ahmerov et al. 1958 = Ахмеров К.З., Башев Т.Г., Каримова Г.Р., Юлдашев А.А. (ed.), *Башкортса-русса һүзлек. Башкирско-русский словарь*, Москва
- AiM = *Alem-i Medeniye*, medeniye.org/lugat
- Akobirov/Magrufov/Hodžahanov 1959 = Акобиров С.Ф., Магруфов З.М., Ходжаханов А.Т. (ed.), *Узбекско-русский словарь. Ўзбекча-русча лугат*, Москва
- Akşan Y. et al. 2012, Construction of the Turkish National Corpus (TNC). – *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)* [www.lrec-conf.org/proceedings/lrec2012/papers.html]

- Alderete J., Beckman J., Benua L., Gnanadesikan A., McCarthy J., Urbanczyk S. 1999, Reduplication with Fixed Segmentism. – *Linguistic Inquiry* 30/3, 327–364
- Alderson A.D., İz F. 1959, *The Concise Oxford Turkish Dictionary*, Oxford
- AltTS = Gürsoy-Naskali/Duranlı 1999
- Anderson G.D.S. 1998, *Xakas* (= *Languages of the World/Materials* 251), München
- Aqtay G., Jankowski H. 2011, *Słownik kazachsko-polski*, Kraków
- Aqtay G. 2009, *Eliyahu Ben Yosef Qilci's Anthology of Crimean Karaim and Turkish Literature* (= *Yıldız Dil ve Edebiyat Dizisi* 8), 1: *Introduction, Text and Indexes*, İstanbul
- Argenti 1533 = Adamović 2001
- Aşmarin 1928–50 = Ашмарин Н.И., *Чăваш сăмахёсен кёнеки. Thesaurus Linguae Tschuvaschorum*, Казань – Чебоксары
- Avrorin 1959–61 = Аврорин В.А., *Грамматика нанайского языка*, Москва – Ленинград
- Avrorin 2000 = Аврорин В.А., *Грамматика маньчжурского письменного языка*, Санкт-Петербург
- AzRS = Taġiyev et al. 2006
- Babuškin 1975 = Бабушкин Т.Ф., Имя прилагательное. – Баскаков Н.А. (ed.), *Грамматика хакасского языка*, Москва, 82–95
- Balakaev 1959 = Балакаев М.Б., *Современный казахский язык. Синтаксис словосочетания и простого предложения*, Алма-Ата
- Balcı A. 2006, Reduplication – Balpınar Z. (ed.), *Turkish Phonology, Morphology and Syntax (Türkçe Ses, Biçim ve Tümce Bilgisi)* (= *TC. Anadolu Üniversitesi Yayını 1533 = Açıköğretim Fakültesi Yayını* 814), Eskişehir
- Balcı A. et al. 2004, *Nouveau dictionnaire ouzbek-français. Yangı o'zbekcha-fransuzcha lug'at*, Toshkent
- Bammatov 1960 = Бамматов З.З. (ed.), *Русско-кумыкский словарь*, Москва
- Bammatov 1969 = Бамматов З.З. (ed.), *Кумыкско-русский словарь*, Москва
- Baskakov 1951–52 = Баскаков Н.А., *Каракалпакский язык, 1: Материалы по диалектологии (тексты и словарь), 2: Фонетика и морфология*, Москва
- Baskakov 1958 = Баскаков Н.А. (ed.), *Каракалпакско-русский словарь*, Москва
- Baskakov 1960 = Баскаков Н.А., *Тюркские языки*, Москва
- Baskakov 1972 = Баскаков Н.А., *Диалект кумандинцев (Куманды-кижи). Грамматический очерк, тексты, переводы и словарь*, Москва
- Baskakov 1978 = Баскаков Н.А., *Язык прииссыкульских уйгуров*, Алма-Ата
- Baskakov 1985 = Баскаков Н.А., *Диалект лебединских татар-чалканцев (Куу-кижи). Грамматический очерк, тексты, переводы, словарь*, Москва
- Baskakov et al. 1991 = Baskakov N.A. (ed.), Gaydarci G.A., Koltza E.K., Pokrovskaya L.A., Tukan B.P., *Gagauz Türkçesinin Sözlüğü*, Ankara [transl. İ. Kaynak and A.M. Doğru]
- Baskakov/Inkižekova-Grekul 1953 = Баскаков Н.А. (ed.), Инкижекова-Грекул А.И., *Хакасско-русский словарь*, Москва
- Baskakov/Karryev/Намзаев 1968 = Баскаков Н.А., Каррыев Б.А., Хамзаев М.Я. (ed.), *Туркменско-русский словарь*, Москва
- Baskakov/Nasilov 1939 = Баскаков Н.А., Насилов В.М., *Уйгурско-русский словарь*, Москва [after Jarring 1964]

- Baskakov/Zajončkovskij/Šaršal 1974 = Баскаков Н.А., Зайончковский А., Шапшал С.М., *Караимско-русско-польский словарь*, Москва
- Bektaev 2001 = Бектаев Қ., *Қазақша-орысша сөздік. Казахско-русский словарь*, Алматы
- Benzing J. 1956, *Die tungusischen Sprachen. Versuch einer vergleichenden Grammatik* (= *Abhandlungen der geistes- und sozialwissenschaftlichen Klasse* 1955/11), Wiesbaden
- Benzing J. 1985, *Kalmückische Grammatik zum Nachschlagen*, Wiesbaden
- Berta Á. 1989, *Lautgeschichte der tatarischen Dialekte*, Szeged
- Berta Á. 1998a, Tatar and Bashkir – *TkcLangs*, 283–300, 284
- Berta Á. 1998b, West Kipchak Languages. – *TkcLangs*, 301–17, 306
- von der Berswordt I. 1839, *Neueste Grammatik der türkischen Sprache für Deutsche zum Selbstunterricht*, Berlin
- Bese L. 1960, Einige Bemerkungen zur partikulären Reduplikation im Mongolischen. – *Acta Orientalia Academiae Scientiarum Hungaricae* 11/1, 43–49
- Birjukovič 1984 = Бирюкович Р.М., *Лексика чулымско-тюркского языка. Пособие к спецкурсу*, Саратов
- Blevins J. 2004, *Evolutionary Phonology. The Emergence of Sound Patterns*, Cambridge
- Bodrogligeti A.J.E. 2003, *Academic reference grammar of modern literary Uzbek*, Los Angeles
- Bombaci A., Padre Pietro Ferraguto e la sua *Grammatica turca* (1611). – *Annali del Regio Istituto Superiore Orientale di Napoli*, N. S. 1, 205–36
- Borovkov 1959 = Боровков А.К., *Краткий очерк грамматики узбекского языка* – UzbRS59: 677–715
- Boryś W. 2005, *Słownik etymologiczny języka polskiego*, Kraków
- Brockelmann K. 1954, *Osttürkische Grammatik der islamischen Literatursprache Mittelasien*, Leiden
- Bruck/Apenčenko 1964 = Брук С.И., Апенченко В.С. (ed.), *Атлас народов мира*, Москва
- BškRS58 = Ahmerov et al. 1958
- BškRS96 = Uraksin 1996
- BTS = *Büyük Türkçe Sözlük*, tdkterim.gov.tr/bts
- Budagov 1869–71 = Будаговъ Л., *Сравнительный словарь турецко-татарских нарѣчій, со включеніемъ употребительнѣшихъ словъ арабскихъ и персидскихъ и съ переводомъ на русскій языкъ*, Санкт-Петербургъ
- Budagov 1987 = Будагов Э., *Основы грамматики современного азербайджанского языка. Морфология*, Баку
- Bulatova 1999 = Булатова Н.Я., *Язык сахалинских эвенков*, Санкт-Петербург
- Butanaev 1999 = Бутанаев В.Я., *Хоорай-орыс тархын-этнография сөстигі*, Ағбан тура
- Cincius 1982 = Цинциус В.И., *Негидальский язык. Исследования и материалы*, Ленинград
- Bußmann H. ²1990, *Lexikon der Sprachwissenschaft*, Stuttgart
- Čispijakov 1992 = Чиспияков Э.Ф., *Учебник шорского языка*, Кемерово
- Clark L. 1998, *Turkmen Reference Grammar*, Wiesbaden
- Clauson G. 1972, *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*, Oxford
- Clauson G. 2002, *Studies in Turkic and Mongolic Linguistics*, London – New York
- Comidas de Carbognano C. 1794, *Primi principj della gramatica turca ad uso dei missionari apostolici di Constantinopoli*, Roma

- Crystal D. 2008, *A dictionary of linguistics and phonetics*, Malden, MA – Oxford – Victoria
- Daibova 1981 = Даибова К.Х., *Къумукъча-русча ва русча-къумукъча фразеология сѣзлюгю. Кумыкско-русский и русско-кумыкский фразеологический словарь*, Махачкала
- DAK = Tryjarski 1968–72
- Dankoff R., Kelly J. 1982, *Maḥmūd al-Kāšgarī. Compendium of the Turkic dialects (Dīwān Luyāt at-Turk) (= Sources of Oriental Languages and Literatures 7)*, 1, Harvard
- Davids A.L. 1836, *Grammaire turke : précédée d'un discours préliminaire sur la langue et la littérature des nations orientales avec un vocabulaire volumineux des dialogues, un recueil d'extraits en prose et en vers, et enrichie de plusieurs planches lithographiques extraites de manuscrits anciens et modernes*, Londres [transl. S. Davids]
- De Jong F. 2007, *A Grammar of Modern Uyghur*, Utrecht
- Demircan Ö. 1987, Emphatic Reduplications in Turkish. – Boeschoten H.E., Verhoeven, L.Th. (ed.), *Studies on modern Turkish. Proceedings of the third conference on Turkish linguistics*, Tilburg, 24–41 [after Müller 2004]
- Demircan Ö. 1989, Türkçe'de Pekiştirme İşlemi. – *Dilbilim* 8: 147–164
- Deny J. 1921, *Grammaire de la langue Turque (Dialecte Osmanli)*, Paris [reprint Vaduz 1986]
- Deny J. 1938, Existe-t-il des préfixes en turc ? – *Bulletin de la Société de linguistique de Paris* 39, 51–65
- Deny J., Grønbech K., Scheel H., Togan Z.V. (ed.), *Philologiae Turcicae Fundamenta*, Wiesbaden
- Derleme Sözlüğü (= Türk Dil Kurumu Yayınları 211/1)*, Ankara ²1993
- Dilçin C. 1991, *Mes'ūd bin Aḥmed: Süheyl ü Nev-Bahār. İnceleme – Metin – Sözlük (= Atatürk Kültür Merkezi Yayını 51)*, Ankara
- Di Sciullo A.M., Affixes at the Edge. – *Canadian Journal of Linguistics / Revue canadienne de linguistique* 50, 83–117
- Dmitriev 1940 = Дмитриев Н.К., *Грамматика кумыкского языка*, Москва – Ленинград
- Doerfer G. 1963–75, *Türkische und mongolische Elemente im Neupersischen: unter besonderer Berücksichtigung älterer neupersischer Geschichtsquellen, vor allem der Mongolen- und Timuridenzeit*, Wiesbaden
- Doerfer G. 1967, *Türkische Lehnwörter im Tadschikischen (= Abhandlungen für die Kunde des Morgenlandes 37, 3)*, Wiesbaden
- Doerfer G. 1975–76, Das Vorosmanische (die Entwicklung der oghusischen Sprachen von den Orchoninschriften bis zu Sultan Veled). – *Türk Dili Araştırmaları Yıllığı. Belleten*, 81–131
- Doerfer G. 1982, Nomenverba im Türkischen. – *Studia turcologica memoriae Alexii Bombaci dicata*, Napoli, 101–114
- Doerfer G., Tezcan S. 1980, *Wörterbuch des Chaladsch (Dialekt von Xarrab)*, Budapest
- Doerfer G., Weiers M. (ed.) 1985, *Beiträge zur Nordasiatischen Kulturgeschichte (= Tungusica 1)*, Wiesbaden
- Doniyorova S. 2008, *Parlons Koumyk*, Paris – Budapest – Torino
- DS = *Derleme Sözlüğü*
- DTS = Nadeljaev/Nasilov/Tenišev/Ščerbak 1969
- Dubeux L. 1856, *Éléments de la grammaire turque a [sic] l'usage des élèves de l'École Impériale et Spéciale des Langues Orientales Vivantes*, Paris
- DW = Stachowski M. 1993

- DWS = Stachowski M. 1998a
 DUzbF = Balcı et al. 2004
 Dyrenkova 1940 = Дыренкова Н.П., *Грамматика ойротского языка*, Москва – Ленинград
 Dyrenkova 1941 = Дыренкова Н.П., *Грамматика шорского языка*, Москва – Ленинград
 Dyrenkova 1948 = Дыренкова Н.П., *Грамматика хакасского языка. Фонетика и морфология*, Абакан
 EDAL = Starostin/Dybo/Mudrak et al. 2003
 Egorov 1964 = Егоров В.Г., *Этимологический словарь чувашского языка*, Чебоксары
 Erdal M. 1991, *Old Turkic Word Formation. A Functional Approach to the Lexicon*, Wiesbaden
 Erdal M. 1998, *Old Turkic – TkcLangs*: 138–157
 Erdal M. 2004, *A Grammar of Old Turkic*, Leiden
 Èrgil-ool 1993 = Эргил-оол И.Ч., Демдек ады. – Монгуш Д.А., Доржу Ч.М., Ойдан-оол А.К., Баир О.М. (ed.), *Тыва дыл*, Кызыл, 103–110
 e-sozlik = *qaraqalpaqsha so'zlikler*, www.e-sozlik.net
 ÈSTJa = Sevortjan et al. 1974–
 Evans V. 2007, *A Glossary of Cognitive Linguistics*, Edinburgh
 Fedotov 1996 = Федотов М.Р., *Этимологический словарь чувашского языка*, Чебоксары
 Ferraguto 1611 = Bombaci 1940
 Frankle E. 1948, Some Problems of Word Formation in the Turkic Languages. – *Journal of the American Oriental Society* 68/2, 114–120
 Friederich M., Yakup A. 2002, *Uyghurisch Lehrbuch*, Wiesbaden
 Fritsch S., Günther F. 2012, *neuralnet* 1.31, cran.r-project.org/web/packages/neuralnet/index.html
 Fu'ad-Efendi, Ğävdät-Efendi 1855, *Grammatik der osmanischen Sprache*, Helsingfors [transl. H. Kellgren]
 von Gabain A. 1945, *Özbekische Grammatik mit Bibliographie, Lesestücken und Wörterverzeichnis* (= *Porta Linguarum Orientalium* 25), Leipzig – Wien
 von Gabain A. ²1950, *Alltürkische Grammatik mit Bibliographie, Lesestücken und Wörterverzeichnis, auch neutürkisch*, Leipzig
 Gadžieva 1975 = Гаджиева Н.З., *Проблемы тюркской ареальной лингвистики*, Москва
 GagTS = Baskakov et al. 1991
 Garkavecъ 2000 = Гаркавецъ О., *Урумський словник*, Алма-Ата
 Girfanova A.H. 2002, *Udeghe*, München
 Glück H. (ed.) 1993, *Metzler Lexikon Sprache*, Stuttgart – Weimar
 Golovkina/Osmanov/Denisova et al. 1966 = Головкина О.В., Османов М.М., Денисова Н.Т. et al. (ed.), *Татарско-русский словарь*, Москва
 Gorelova L.M. 2002, *Manchu Grammar* (= *Handbook of Oriental Studies. Handbuch der Orientalistik* 8/7), Leiden – Boston – Köln
 Grønbech K. 1942, *Komanisches Wörterbuch. Türkischer Wortschatz zu Codex Cumanicus* (= *Monumenta Linguarum Asiæ Maioris. Subsidia* 1), København
 GTS = *Güncl Türkçe Sözlük*, www.tdk.org.tr/index.php?option=com_gts&arama=gts
 Güler S. 2003, *A Study on Turkish Reduplications: Some Observations on Form and Function of Reduplication*, MA thesis, Mersin Üniversitesi (advisor M. Aksan)

- Günther F., Fritsch S. 2010, neuralnet: Training of Neural Networks. – *The R Journal* 2/1, 30–38
- Gürsoy-Naskali E., Duranlı M. 1999, *Altayca – Türkçe Sözlük* (= *Türk Dil Kurumu Yayınları* 725), Ankara
- Guzev 1979 = Гузев В.Г., *Староосманский язык*, Москва
- Насеминоğlu N. 1996, *Karahanlı Türkçesi Grameri* (= *Türk Dil Kurumu Yayınları* 638), Ankara
- Hafuz 1995 = Хафуз М.Э., *Русско-караимский словарь. Крымский диалект*, Москва
- HakOS = Patačakova/Čankov et al. 1988
- HakRS = Baskakov/Inkižekova-Grekul 1953
- Hamzaev 1962 = Хамзаев М.Я. (ed.), *Словарь туркменского языка*, Ашхабад
- Hanser O. 1977, *Turkmen Manual. Descriptive Grammar of contemporary literary Turkmen. Texts. Glossary*, Wien
- Hanser O. 2003, *Türkmençe Elkitabı*, İstanbul [transl. Z. Kargı Ölmez]
- Harrison K.D. 2000, *Topics in the Phonology and Morphology of Tuva*, doctoral dissertation, Yale University (advisor S. Anderson) [www.swarthmore.edu/SocSci/dharris2/Harrison-Dissertation.pdf]
- Harrison K.D. 2004, South Siberian Sound Symbolism. – Vajda 2004, 199–214
- Harsány 1672 = Hazai 1973
- Hatiboğlu V. 1973, *Pekiştirme ve kuralları* (= *Türk Dil Kurumu Yayınları* 372), Ankara [after Müller 2004]
- Hazai G. 1973, *Das Osmanisch-Türkische im XVII. Jahrhundert. Untersuchungen an den Transkriptionstexten von Jakab Nagy de Harsány*, Budapest
- Hebert R.J., Poppe N. 1963, *Kirghiz Manual* (= *Uralic and Altaic Series* 33), Bloomington – The Hague
- Helimski 2001 = Хелимский Е.А., «Этническая рознь» на Таймырском Севере. Кто ищет, тот всегда найдет... – *Сибирская заимка: Профессиональный подход к изучению Сибири* 7: 173–180 [rev. of Андерсон Д.Дж. 1998, *Тундровики: Экология и самосознание таймырских эвенков и долган*, Новосибирск]
- Hızır = Önler 1999
- Holderman J.B.D. 1730, *Grammaire turque ou methode courte & facile pour apprendre la langue turque. Avec un Recueil des Noms, des Verbes, & des Manieres de parler les plus necessaires a sçavoir, avec plusieurs Dialogues familiers*, Constantinople
- Houtsma M.Th. 1894, *Ein türkisch-arabisches Glossar. Nach der Leidener Handschrift*, Leiden
- Hugjiltu W. 1998, One Shared Method to Express Reinforced Meaning of Adjectives in the Altaic Languages. – *Altai Hakpo. Journal of the Altaic Society of Korea* 8, 211–225
- Imart G. 1981, *Le Kirghiz (Turk d'Asie Centrale Soviétique). Description d'une langue de litterisation recente*, Aix-en-Provence
- Inkscape Team 2013, *Inkscape*, www.inkscape.org
- Ishakov/Palymbah 1961: 187 = Исхаков Ф.Т., Пальмбах А.А., *Грамматика тувинского языка. Фонетика и морфология*, Москва
- JakRS = Slepcev 1972
- Janhunen J. (ed.) 2003, *The Mongolic Languages*, London – New York
- Jankowski H. 1992, *Gramatyka języka krymskotatarskiego*, Poznań

- Jankowski H. 1997, A Bible Translation Into the Northern Crimean Dialect of Karaim. – *Studia Orientalia* 82, 1–84
- Jankowski H. 2003, Position of Karaim among the Turkic languages. – *Studia Orientalia* 95, 131–153
- Jarring G. 1933, *Studien zu einer osttürkischen Lautlehre*, Lund
- Jarring G. 1964, *An Eastern Tuki-English Dialect Dictionary* (= *Lunds Universitets Årsskrift. N.F.* 1/56/4), Lund
- Jastremskij 1938 = Ястремский С.В., *Грамматика якутского языка. Пособие для педагогов*, Москва
- Jaubert P.A. ²1833, *Éléments de la grammaire turke, à l'usage des élèves de l'école royale et spéciale des langues orientales vivantes*, Paris
- Jehliſschka H. 1895, *Türkische Konversations-Grammatik*, Heidelberg
- Johanson L. 1998, The Structure of Turkic. – *TkcLangs*, 30–66
- Johanson L., Csátó É.Á. (ed.) 1998, *The Turkic Languages*, London – New York
- Józefowicz G. 2008, *Słownik polsko-karaimski w dialekcie trockim*, Troki – Wilno – Warszawa – Wrocław – Gdańsk – Nashville
- Judahin 1985 = Юдахин К.К., *Киргизско-русский словарь. Кыргызча-орусча сөздүк*, Фрунзе
- Juldašev 1981 = Юлдашев А.А. (ed.), *Грамматика современного башкирского литературного языка*, Москва
- Kakuk Zs. 1991, Ein krimkaraimisches Wörterverzeichnis. – *Acta Orientalia Academiae Scientiarum Hungaricae* 45/2–3, 347–401
- Kałuzyński S. 1979 = Калужинский С., Этимологические исследования по якутскому языку. Двусложные основы (III). – *Rocznik Orientalistyczny* 40/2, 39–49
- Karamanlioğlu A.F., *Kırçak Türkçesi Grameri* (= *Türk Dil Kurumu Yayınları* 579), Ankara
- KarRPS = Baskakov/Zajončkovskij/Šapšal 1974
- Kary-Nijazov/Borovkov 1941 = Кары-Ниязов Т.Н., Боровков А.К., *Узбекско-русский словарь*, Ташкент
- Kasapoğlu Çengel 2005 = Kasapoğlu Çengel H., *Kırgız Türkçesi Grameri. Ses ve Şekil Bilgisi*, Ankara
- Kaşgarlı S.M. 1992, *Modern Uygur Türkçesi Grameri*, İstanbul
- Kelepir M. 2000, To be or not to be faithful. – Göksel A., Kerslake C., *Studies on Turkish and Turkic Languages* (= *Turcologica* 46), Wiesbaden, 11–18
- Khabtagaeva B. 2001, Colour names and their suffixes. A study on the history of Mongolian word formation. – *Acta Orientalia Academiae Scientiarum Hungaricae* 54/1, 85–165
- Kim H.-S. 2003, A new look at partial reduplication in Korean. – *Studies in Phonetics, Phonology and Morphology* 9/2, 353–380
- Kim H.-S. 2009, The Full-to-Partial Reduction in Korean and Turkish Reduplication. – *Linguistic Research* 26/2, 121–148
- KirgRS = Judahin 1985
- Kirişcioğlu M.F. ²1999, *Saha (Yakut) Türkçesi Grameri*, Ankara
- Kissen 1975 = Киссен И.А. ⁷1975, *Краткий грамматический справочник по узбекскому языку для 8–10 классов русской школы*, Ташкент
- KklpRS = Baskakov 1958

- KmkRFS = Daibova 1981
 KmkRS = Vammatov 1969
 Кононов 1956 = Кононов А.Н., *Грамматика современного турецкого языка*, Москва – Ленинград
 Кононов 1960 = Кононов А.Н., *Грамматика современного узбекского литературного языка*, Москва – Ленинград
 Konstantinova 1964 = Константинова О.А., *Эвенкийский язык. Фонетика. Морфология*, Москва – Ленинград
 Korkina/Ubrjatova et al. 1982 = Коркина Е.И., Убрятова Е.И. (ed.), Харитонов Л.Н., Петров Н.Е. 1982, *Грамматика современного якутского литературного языка*, Москва
 Köroğlu = Mahir et al. 1973
 Kotvič V.L. ²1929 = Котвич В.Л., *Опыт грамматики калмыцкого разговорного языка*, Ржев-нище у Праги
 Kowalski T. 1929, *Karaimische Texte im Dialekt von Troki. Eingeleitet, erläutert und mit einem karaimisch-polnisch-deutschen Glossar versehen* (= *Prace Komisji Orientalistycznej Polskiej Akademji Umiejętności* 11), Kraków
 Krueger J.R. 1961, *Chuvash Manual. Introduction, Grammar, Reader, and Vocabulary*, Bloomington – The Hague
 Krueger J.R. 1997, *Tuvan Manual* (= *Indiana University Uralic and Altaic Series* 126, Bloomington)
 Krzywinski M. et al. 2009, *Circos: an information Aesthetic for Comparative Genomic.* – *Genome Research* 19, 1639–1645
 Kulikovskaja/Musaeva 2006 = Куликовская Л.К., Мусаева Э.Н., *Грамматика казахского языка в таблицах и схемах в сопоставлении с грамматикой русского языка (простое предложение)*, Алматы
 Kurpeško Tannagaşeva N.N., Akalın Ş.H. 1995, *Şor sözlüğü* (= *Türkoloji Araştırmaları* 2, *Sözlük Dizisi* 1), Adana
 Kuryžšanov 1970 = Курьшжанов А.К., *Исследование по лексике старокыпчакского письменного памятника XIII в. — «Тюрско-арабского словаря»*, Алма-Ата
 KzkED = Shnitnikov 1966
 KzkRS = Bektaev 2001
 Ladstätter O., Tietze A. 1994, *Die Abdal (Äynu) in Xinjiang*, Wien
 Laturov 1969 = Латыпов А.М., *Имя прилагательное* – Курбатов Х.Р., Махмутова Л.Т., Смолякова Л.П., Тенишев Э.Р. (ed.), *Современный татарский литературный язык. Лексикология, фонетика, морфология*, Москва, 173f.
 Laude-Cirtautas I. 1961, *Der Gebrauch der Farbbezeichnungen in den Türkdialekten* (= *Uralaltaische Bibliothek* 10), Wiesbaden
 Lessing F.D. (ed.) et al. 1960, *Mongolian-English dictionary*, Berkeley – Los Angeles
 Levi 1996 = Леви Б.З., *Русско-караимский словарь. Крымский диалект. Урус-Қарай сөзлүк*, Одесса
 Li F., Whaley L.J. 2000, *Emphatic reduplication in Oroqen and its Altaic context.* – *Linguistics* 38/2, 355–372

- Li P., MacWhinney B. 2002, PatPho: A phonological pattern generator for neural networks. – *Behavior Research Methods, Instruments & Computers* 34/3, 408–415
- Li Y.-S. 2011, *A study of Dolgan*, Seoul
- Li Y.-S., Lee H.-Y., Choi H.-W., Kim G.-S., Lee D.-E., Ölmez M. 2008, *A study of Middle Chulyum Dialect of the Chulyum Language*, Seoul
- Li Y.-S., Ölmez M., Juwon K. 2007, Some Newly Identified Words in “Fuyu Kirghiz” (Part 1). – *Ural-Altische Jahrbücher, neue Folge* 21: 141–69
- Magrufov et al. 1981 = Мағруфов З.М. (ed.), *Толковый словарь узбекского языка. Ўзбек тилининг изоҳли луғати*, Москва
- Mahir B., Kaplan M., Akalın M., Bali M. 1973, *Köroğlu Destanı* (= *Atatürk Üniversitesi Yayınları* 314 = *Edebiyat Fakültesi Yayınları* 63 = *Araştırma Serisi* 52), Ankara
- Malchukov A.L. 1995, *Even* (= *Languages of the World/Materials* 12), München – Newcastle
- Mallouf N. 1862, *Grammaire élémentaire de la langue turque suivie de dialogues familiers avec la prononciation figurée et d'un petit secrétaire ou modèle de lettres avec la traduction française en regard*, Paris
- Malov 1951 = Малов С.Е., *Памятники древнетюркской письменности*, Москва – Ленинград
- Malov 1954 = Малов С.Е., *Уйгурский язык. Хамийское наречие. Тексты, переводы и словарь*, Москва – Ленинград
- Marantz A. 1982, Re Reduplication. – *Linguistic Inquiry* 13/3, 435–482
- Marchand H. 1952, Alliteration, Ablaut und Reim in den türkischen Zwillingsformen. – *Oriens* 5/1, 60–69
- Mardkowicz A. 1935, *Karaj sez-bitigi. Słownik karaimski. Karaimisches Wörterbuch*, Łuck
- Mascis A. 1677, *Vocabolario toscano, e turchesco*, Firenze
- McCarthy J., Prince A. 1988, Quantitative Transfer in Reduplicative and Templatic Morphology. – *Linguistics in the Morning Calm* 2, Seoul, 3–35
- Menges K.H. [1936], *Volkskundliche Texte aus Ost-Türkistan. Aus dem Nachlaß von N. Th. Katanov* (= *Abhandlungen der Preussischen Akademie der Wissenschaften. Philosophisch-historische Klasse*), Leipzig 1976
- Menges K.H. 1959, Die aralo-kaspische Gruppe (Kasakisch, Karakalpakisch, Nogaisch, Kipçak-Özbekisch, Kirgisisch). – *PhTF*, 434–488
- Menges K.H. ²1995, *The Turkic Languages and Peoples. An Introduction to Turkic Studies* (= *Veröffentlichungen der Societas Uralo-Altaica* 42), Wiesbaden [after Müller 2004]
- Meninski à Mesgnien F. 1680, *Thesaurus linguarum Orientalium Turcicæ Arabicæ Persicæ. Lexicon Turcico-Arabico-Persicum*, Vienna [2nd ed. by M. Ölmez, S. Stachowski, İstanbul 2000]
- Mes‘ūd = Dilçin 1991
- MifyNM = Tokarev et al. 1991–92
- MoLangs = Janhunen 2003
- Molino 1641 = Siemieniec-Gołaś 2005
- Müller H.-G. 2004, *Reduplikationen im Türkischen. Morphophonologische Untersuchungen* (= *Turcologica* 26), Wiesbaden

- Musaev 1964 = Мусаев К.М., *Грамматика караимского языка. Фонетика и морфология*, Москва
- Musaev 1977 = Мусаев К.М., *Краткий грамматический очерк караимского языка*, Москва
- Mustafaev/Starostov et al. 1977 = Мустафаев Э.М.-Э., Старостов Л.Н. et al. 1977, *Туреукорусский словарь*, Москва
- Nadeljaev/Nasilov/Tenišev/Ščerbak 1969 = Наделяев В.М., Насилов Д.М., Тенишев Э.Р., Щербак А.М., *Древнетюркский словарь*, Ленинград 1969
- Nedjalkov I. 1997, *Evenki*, London – New York
- Németh J. 1916, *Türkische Grammatik*, Berlin – Leipzig
- Nerbonne J., Kleiweg P., Manni F., Heeringa W. 2008, Projecting Dialect Distances to Geography: Bootstrap Clustering vs. Noisy Clustering. – Preisach Ch. et al. (ed.), *Data Analysis, Machine Learning and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikationen e.V., Albert-Ludwigs-Universität Freiburg, March 7–9, 2007* (= *Studies in Classification, Data Analysis, and Knowledge Organization* 2008), Berlin–Heidelberg, 647–654
- Nişanyan ÇTES = S., *Sözlerin Soyağacı. Çağdaş Türkçenin Etimolojik Sözlüğü*, www.nisanyansozluk.com
- Ölmez M. 1991, Ana Altayca Sözbaşı *d-. – *Türk Dilleri Araştırmaları*, 167–190
- Ölmez M. 2007, *Tuwinischer Wortschatz mit alttürkischen und mongolischen Parallelen. Tuwacanın Sözcükleri Eski Türkçe ve Moğolca Denklemleri* (= *Veröffentlichungen der Societas Uralo-Altaica*, Wiesbaden
- Ölmez M. 2011, Kutadgu Bilig'in sözcükleri ve adakşu “birlikte, beraber, yanyana” üzerine. – *Doğumunun 990. yılında Yusuf Has Hacib ve eseri Kutadgu Bilig bildirileri. 26–27 Ekim 2009*, Ankara, 399–406
- Önler Z. 1999, *Celâlüddin Hızır (Hacı Paşa), Müntahab-ı Şifa, 2: Sözlük*, İstanbul
- Özer Ş. 2008, *Die nominale Wortbildung im Altosmanischen am Beispiel der Übersetzung von Ta'labîs »Qışaş al-Anbiyâ« aus dem 14. Jahrhundert*, Wiesbaden
- Özkan N. 1996, *Gagavuz Türkçesi Grameri. Giriş – Ses Bilgisi – Şekil Bilgisi – Cümle – Sözlük – Metin Örnekleri* (= *Türk Dil Kurumu Yayınları* 657 = *Grammer Bilim ve Uygulama Yayınları* 18), Ankara
- Öztürk R. 1994, *Yeni Uygur Türkçesi Grameri* (= *Türk Dil Kurumu Yayınları* 593), Ankara
- Pallas 1786–89 = P.S. Pallas 1786–89 / П.С. Палласъ 1787–89, *Linguarum totius orbis vocabularia comparativa / Сравнительные словари всехъ языковъ и наречій*, Petropolis / Санктпетербургъ
- Palmbach 1955 = Пальмбах А.А., *Тувинско-русский словарь с приложением грамматического очерка тувинского языка*, Москва
- Patačakova 1962 = Патачакова Д.Ф., *Хакас тілі. Пастагы чардыгы. Лексика, фонетика, морфология. Педагогической училищее учебник*, Абакан
- Patačakova/Čankov 1988 = Патачакова Д.Ф., Чанков Д.И. et al. (ed.), *Хакас тілінің орфографиясы. Орфография сөстігі*, Абакан
- Peckarskij 1907–30 = Пекарский Э.К. 1907–30, *Словарь якутского языка*, Санкт-Петербург – Петроград – Ленинград
- Petrova 1936 = Петрова Т.И., *Ульчский диалект нанайского языка*, Ленинград

- Petrova 1967 = Петрова Т.И., *Язык ороков (ульта)*, Ленинград
- PhTF = Deny et al. 1959
- Podolak B., Nykiel P. 2008, *Kieszonkowy słownik turecko-polski, polsko-turecki*, Kraków
- Pokrovskaja 1964 = Покровская Л.А., *Грамматика гагаузского языка. Фонетика и морфология*, Москва
- Pomorska M. 2004, *Middle Chulym noun formation* (= *Studia Turcologica Cracoviensia* 9), Kraków
- Pomorska M. 2010, Kilka uwag o nowych źródłach słownictwa czułymskiego. – Siemieniec Gołaś E., Georgiewa-Okon J. (ed.), *Od Anatolii po Syberię*, Kraków, 171–77
- Poppe N. 1951, *Khalkha-mongolische Grammatik mit Bibliographie, Sprachproben und Glossar*, Wiesbaden
- de Preindl J. 1790, *Grammaire turque d'une toute nouvelle methode d'apprendre cette langue en peu de semaines avec un vocabulaire enrichi d'anecdotes utiles et agréables*, Berlin
- Prik 1976 = Прик О.Я., *Очерк грамматики караимского языка (крымский диалект)*, Махачкала
- Qorqut = Rossi 1952
- Quantum GIS Development Team 2013, *Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project*, qgis.osgeo.org
- R Core Team 2013, *R: A language and Environment for Statistical Computing*, Vienna, r-project.org
- Radloff W. 1883, *Phonetik der nördlichen Türksprachen, mit einer ethnographischen Übersicht der Türkstämme*, Leipzig
- Radloff W. 1893–1911 = Радлов В.В., *Опытъ словаря тюркскихъ нарѣчій* = Radloff W., *Versuch eines Wörterbuches der Türk-Dialecte*, Санктъ-Петербургъ
- Rahimov/Nadžip 1968 = Рахимов Т.Р. (ed.), Наджип Э.Н., *Уйгурско-русский словарь*, Москва
- Raimy E. 2000, *The Phonology and Morphology of Reduplication* (= *Studies in Generative Grammar* 52), Berlin – New York
- Ramstedt G.J. 1952, *Einführung in die altaische Sprachwissenschaft, 2: Formenlehre*, Helsinki
- Räsänen M. 1949, *Materialien zur Lautgeschichte der türkischen Sprachen*, Helsinki
- Räsänen M. 1955 = Рясänen М., *Матерялы по исторической фонетике тюркских языков*, Москва [transl. А.А. Юдашев, ed. by Н.А. Баскаков]
- Räsänen M. 1957, *Materialien zur Morphologie der türkischen Sprachen*, Helsinki
- Räsänen M. 1969, *Versuch eines etymologischen Wörterbuchs der Türksprachen* (= *Lexica Societatis Fenno-Ugricae* XVII,1), Helsinki
- Rassadin 1978 = Рассадин В.И., *Морфология тофаларского языка в сравнительном освещении*, Москва
- Rebi 2004 = Реби Д., *Крымчакский язык. Крымчакско-русский словарь*, Симферополь
- Redhouse 92001 = *İngilizce-Türkçe / Türkçe-İngilizce Redhouse Yeni Elsözlüğü*, İstanbul
- Redhouse Sir J.W. 1884, *A Simplified Grammar of the Ottoman-Turkish Language*, London
- RKarS-Haf = Hafuz 1995
- RKarS-Lev = Levi 1996
- RKmkS = Bammatov 1960

- Röhrborn K. 1977–, *Uigurisches Wörterbuch. Sprachmaterial der vorislamischen türkischen Texte aus Zentralasien*, Wiesbaden
- Romero J.A. [18th c.], *Gramatica Turca ô Methodo breve, y facil para aprehender la lengua turca con una coleccion de nombres, verbos, y phrases mas comunes, y varios dialogos familiares* [manuscript dated 18th c. by Biblioteca Nacional de España, bdh.bne.es/bnsearch/detalle/3427019]
- Roos M.E. 2000, *The Western Yugur (Yellow Uygur) Language, Grammar, Texts, Vocabulary*, doctoral thesis, Universiteit Leiden (advisor F.H.H. Kortlandt)
- Rossi E. 1952, *Il “Kitāb-i Dede Qorquṭ”: racconti epico-cavallereschi dei Turchi Oğuz, tradotti e annotati con “facsimile” del MS. Vat. Turco 102 (= Studi e testi 159)*, Città del Vaticano
- R Development Core Team 2011, *R: A Language and Environment for Statistical Computing*, Vienna, R Foundation for Statistical Computing
- Rybatzki V. 1994, Bemerkungen zur türkischen und mongolischen Metallterminologie. – *Studia Orientalia* 73, 193–251
- Sablin I., Savelyeva M. 2011, Mapping Indigenous Siberia: Spatial Changes and Ethnic Realities, 1900–2010. – *settler colonial studies* 1, 77–110
- Savdan uulu A., Vatmanov I.A. 1933, *Qırqız tilinin elementar dırq qramatikesi. Элементарные основы грамматики киргизского языка*, Prunza – Taškent
- Sadvakasov 1976 = Садвакасов Г., *Язык Уйгуров ферганской долины, 2: Лексика, морфология и языковая интерференция*, Алма-Ата
- Salminen T. 2006, Genuine and confused information about Central Siberian languages. – *Finnisch-Ugrische Forschungen* 59/1–3: 142–49 [rev. of Vajda 2004]
- Sami Ş. 1900–01 = *سامی ش. قاموس ترکی، در سعادت* [ed. İstanbul 1978]
- Ščerbak 1977 = Щербак А.М., *Очерки по сравнительной морфологии тюркских языков (имя)*, Ленинград
- Schönig C. 1988, Wortecho und Echowort im Tatarischen. – *Studia Uralo-Altica* 30: *Tatarische etymologische Studien* II, 89–139
- Schönig C. 1997–98, A new attempt to classify the Turkic languages. – *Turkic Languages* 1/1–2/1, 117–33, 262–77, 130–51
- Schönig C. 1998a, Azerbaijani – *TkcLangs*, 248–60
- Schönig C. 1998b, South Siberian Turkic – *TkcLangs*, 403–16
- Schönig C. 1998c, Turkmen – *TkcLangs*, 261–72
- Schönig C. 2004, On some unclear, doubtful and contradictory passages in Maḥmud al-Kāšğari’s “Dīwān Luḡāt at-Turk”. – *Türk Dilleri Araştırmaları* 14, 35–56
- Schroeder G. ²1835, *Grammaire turque à l’usage des Français et Anglais contenant en outre une collection d’idiotismes, de discours familiers et un vocabulaire en français*, Leipsic
- Seaman G. 1670, *Grammatica Linguæ Turcicæ, In Quinque partes diftributa*, Oxonia
- Sem 1976 = Сем Л.И., *Очерки диалектов нанайского языка. Бикинский (уссурийский) диалект*, Ленинград
- Serebrennikov/Gadžieva 1986 = Серебренников Б.А., Гаджиева Н.З. 1986, *Сравнительно-историческая грамматика тюркских языков*, Москва
- Sevortjan 1966 = Севортян Э.В., *Аффиксы именного словообразования в азербайджанском языке*, Москва

- Sevortjan et al. 1974– = Севортян Э.В. (ed. vol. 1–4) / Левитская Л.С., Дыбо А.В., Рассадин В.И. (vol. 5–6) / Дыбо А.В. (ed. vol. 7) 1974–, *Этимологический словарь тюркских языков*, Moskva
- Sezer T., *TS Corpus* [draft; tscorpus.com]
- Shnitnikov B.N. 1966, *Kazakh-English Dictionary*, London – The Hague – Paris
- ShorTS = Kurpeško Tannagaşeva / Akalın 1995
- Siemieniec-Gołaś E. 2005, *Turkish Lexical Content in Dittionario della lingua Italiana, Turchesca by Giovanni Molino (1641)*, Kraków
- Simpson C.G. 1957, *The Turkish Language of Soviet Azerbaijan. Selections from the Press and from modern authors with a grammatical introduction*, London
- Şiraliyev/Sevortjan 1971 = Ширалиев М.Ш., Севортян Э.В. (ed.), *Грамматика азербайджанского языка (Фонетика, морфология и синтаксис)*, Баку
- Sjoberg A.F. 1963, *Uzbek structural grammar* (= *Uralic and Altaic series* 18), Bloomington – The Hague
- Skorik et al. 1968 = Скорик П.Я. et al. (ed.), *Монгольские, тунгусо-маньчжурские и палеоазиатские языки* (= *Языки народов СССР* 5), Ленинград
- SKzkP = Aqtaу/Jankowski 2011
- Slepcev 1972 = Слепцов П.А. (ed.), *Якутско-русский словарь*, Москва
- Sofu H. 2005, Acquisition of reduplication in Turkish. – Hurch B. (ed.), *Studies on Reduplication*, Berlin, 491–507
- Sofu H., Altan A. 2009, Partial reduplication: revisited. – Ay S., Aydın Ö., Ergenç İ., Gökmen S., İşsever S., Peçenek D. (ed.), *Essays on Turkish Linguistics*, Wiesbaden, 63–72
- Somfai Kara D. 2002, *Kazak* (= *Languages of the World/Materials* 417), München
- Stachowski K. 2010, On First Syllable Reduplications in Karaim. – Mańczak-Wohlfeld E., Podolak B. (ed.), *Studies on the Turkic World. A Festschrift for Professor St. Stachowski on the Occasion of His 80th Birthday*, Kraków, 139–56
- Stachowski K. 2011, Remarks on the usefulness of different types of transcription, with a particular regard to Turkic comparative studies. – *Suomalais-Ugrilaisen Seuran Aikakauskirja / Journal de la Société Finno-Ougrienne* 93, 303–338
- Stachowski K. 2012a, W. Kotwicz's unpublished study *Les voyelles longues dans les langues altaïques* (1938). – *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 129, 245–315
- Stachowski K. 2012b, A phonological encoding of Turkish for neural networks. – *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 129 supplementum, 47–56
- Stachowski K. 2013, The influx rate of Turkic glosses in Hungarian and Polish post-mediaeval texts. – Köhler R., Altmann G. (ed.), *Issues in Quantitative Linguistics 3 dedicated to Karl-Heinz Best on the occasion of his 70th birthday* (= *Studies in Quantitative Linguistics* 13), Lüdenscheid, 100–116
- Stachowski M. 1993, *Dolganischer Wortschatz*, Kraków
- Stachowski M. 1996, Über das Alter der dolganischen Sprache. – *Türk Dilleri Araştırmaları* 6, 123–130
- Stachowski M. 1997, *Dolganische Wortbildung*, Kraków
- Stachowski M. 1998a, *Dolganischer Wortschatz. Supplementband*, Kraków

- Stachowski M. 1998b, Notizien zur schorischen und tschulymischen Etymologie. – *Studia Etymologica Cracoviensia* 3, 107–123
- Stachowski M. 2001, Uralistik und Turkologie – geschieden und doch verliebt. – Pusztay J. (ed.), *Vade tecum! A huszonötödik óra. A Berzsenyi Dániel Főiskola Uralisztikai Tanszékének (1991–2001) jubileumi konferenciája (2001. április 26–27.)* (= *Az Uralisztikai Tanszék kiadványai* 9.), Szombathely, 209–225
- Stachowski M. 2009, *Gramatyka języka tureckiego w zarysie*, Kraków
- Stachowski M. 2010, Ist das slavische Schaf (**овьса*, **овьсь*) bis nach Transkaukasien (*hōgač*) und Anatolien (*öveč*, *ögeč*) gewandert?. – *Studia Etymologica Cracoviensia* 15, 127–136
- Stachowski M. [in print], *Kurzgefaßtes etymologisches Wörterbuch der türkischen Sprache*
- Stachowski S. 1998, *Osmanlı Türkçesinde Yeni Farsça Alıntılar Sözlüğü. Wörterbuch der neupersischen Lehnwörter im Osmanisch-Türkischen*, İstanbul
- Stachowski S. 2002, *Lexique turc dans le Vocabulaire de P.F. Viguier (1790)*, Kraków
- Starostin S., Dybo A., Mudrak O. et al. 2003, *Etymological Dictionary of the Altaic Languages*, Leiden – Boston
- Steuerwald K. 1972, *Türkçe Almanca Sözlük. Türkisch-Deutsches Wörterbuch*, Wiesbaden [after Müller 2004]
- Steuerwald K. 1974, *Almanca Türkçe Sözlük. Deutsch-Türkisches Wörterbuch*, Wiesbaden [after Müller 2004]
- STrkmJa = Hamzaev 1962
- Subrakova 1999 = Субракова О.В., *Хакасско-русский словарь. Хакас-орыс сөстүк*, Новосибирск
- Sulimowicz J. 1973, Materiał leksykalny krymskokaraimskiego zabytku językowego (druk z 1734 r.). II: Słownik – Dictionnaire. – *Rocznik Orientalistyczny* 36/1, 47–107
- Šupa/Aleksandrovič-Miškinene 1995 = Шупа С. (ed.), Александрович-Мишкінене Г., *Турэцка-беларускі размоўнік 1836 году з збораў Нацыянальнага Музею Літоўскай Рэспублікі ў Вільні*, New York
- Sunik 1958 = Суник О.П., *Кур-урмийский диалект. Исследования и материалы по нанайскому языку*, Ленинград
- Sunik 1985 = Суник О.П., *Ульчский язык. Исследования и материалы*, Ленинград
- Svantesson J.-O., Tsendina A., Karlsson A., Franzén V. 2005, *The Phonology of Mongolian*, Oxford
- Tağıyev M.T. et al. 2006, *Azərbaycanca–rusca lügət*, Bakı
- Takashima H. 高島尚生 2008, *Kiso Tuva-go bumpō* [‘Basic grammar of the Tuvian language’], Tōkyō
- Tantau T. et al. 2012, *TikZ & PGF*, sourceforge.net/projects/pgf
- TaS = *XIII. yüzyıldan beri Türkiye Türkçesiyle yazılmış kitaplardan toplanan tanıklariyle Tarama Sözlüğü*, Ankara 1965–77 [www.tdkterim.gov.tr/tarama]
- Tatarincev 2000–08 = Татаринцев Б.И., *Этимологический словарь тувинского языка*, Новосибирск
- TatGramm = *Татарская грамматика. Татар грамматикасы, 2: Морфология*, Казань 1993
- TatRS = Golovkina/Osmanov/Denisova et al. 1966
- Tekin T. 1990, A New Classification of the Turkic Languages, *Türk Dilleri Araştırmaları*, 5–18

- Tekin T., Ölmez M., Ceylan E., Ölmez Z., Eker S. (ed.) 1995, *Türkmençe-Türkçe Sözlük*, Ankara
- Tenišev 1963 = Тенишев Э.Р., *Саларский язык*, Москва
- Tenišev 1968 = Тенишев Э.Р. (ed.), *Тувинско-русский словарь*, Москва
- Tenišev 1976a = Тенишев Э.Р., *Строй саларского языка*, Москва
- Tenišev 1976b = Тенишев Э.Р., *Строй сарыг-югурского языка*, Москва
- Tenišev 1988 = Тенишев Э.Р. (ed.), *Сравнительно-историческая грамматика тюркских языков. Морфология*, Москва
- Tenišev/Todaeva 1966 = Тенишев Э.Р., Тодаева Б.Х., *Язык желтых уйгуров*, Москва
- TkcLangs = Johanson/Csató 1998
- TMEN = Doerfer 1963–75
- Tokarev et al. ²1991–92 = Токарев С.А. et al. (ed.), *Мифы народов мира. Энциклопедия*, Москва
- Tömür H. 2003, *Modern Uyghur Grammar (Morphology)*, İstanbul [transl. A. Lee]
- Toparlı R. 1993, *Kıpçak Türkçesi Sözlüğü*, Ankara
- Toparlı R. (ed.) 2000, *Ahmet Vefik Paşa. Lehce-i Osmânî* (= *Türk Dil Kurumu Yayınları* 743 = *Türkiye Türkçesi Sözlükleri Projesi* 1 = *Eski Sözlükleri Dizisi* 3), Ankara
- Trask L.R. 1996, *A dictionary of phonetics and phonology*, London – New York
- TrkmRS = Baskakov/Karryev/Hamzaev 1968
- TrkmTS = Tekin/Ölmez/Ceylan/Ölmez/Eker 1995
- Tryjarski E. 1968–72, *Dictionnaire arméno-kiptchak d'après trois manuscrits des collections viennoises*, 1, Warszawa
- TS = *Türkçe Sözlük* (= *Türk Dil Kurumu Yayınları* 505/1–2), Ankara ⁷1983
- Tsumagari T. 津曲敏郎 1997, Linguistic Diversity and National Borders of Tungusic. – Shōji H., Janhunen J. (ed.), *Northern Minority Languages: Problems of Survival* (= *Senri Ethnological Studies* 44), Osaka, 175–186
- Tsumagari T. 津曲敏郎 2003, Dagur. – *MoLangs*, 129–152
- Tsumagari T. 津曲敏郎 2009, A sketch of Solon grammar. – 北方人文研究 *Northern Humanities Research* 2, 1–22
- TTAS = *Türkiye Türkçesi Ağızları Sözlüğü*, tdkterim.gov.tr/ttas
- Tuḥfe (1514–15) = Verburg 1997
- TurRS = Mustafaev/Starostov et al. 1977
- Tursun N. 2007, *Uyghur Reader*, Hyattsville
- TuvRS = Tenišev 1968
- Ubrjatova 1985a = Убрятова Е.И., *Историческая грамматика якутского языка*, Якутск
- Ubrjatova 1985b = Убрятова Е.И., *Язык норильских долган*, Новосибирск
- UjgRS = Rahimov/Nadžip 1968
- UjgRS39 = Baskakov/Nasilov 1939
- Uraksin 1996 = Ураксин З.Г. (ed.), *Башкортса-русса һүзлек. Башкирско-русский словарь*, Москва
- UzbRS41 = Kary-Nijazov/Borovkov 1941
- UzbRS59 = Akobirov/Magrufov/Hodžahanov 1959
- Vajda E.J (ed.) 2004, *Languages and Prehistory of Central Siberia*, Amsterdam – Philadelphia

- Vasilevič 1940 = Василевич Г.М., *Очерк грамматики эвенкийского (тунгусского) языка*, Ленинград
- Vaughan Th. 1709, *A Grammar of the Turkish Language*, London
- Vefik Paşa 1890 = Toparlı 2000
- Verburg A.C. 1997, The *Tuhfe-i Şāhidī*: A Sixteenth Century Persian-Ottoman Dictionary in Rhyme. – *Archivum Ottomanicum* 17, 5–87
- VEWT = Räsänen 1969
- Viguiier M. 1790, *Éléments de la langue turque, ou tables analytiques de la langue turque usuelle, avec leur développement*, Constantinople [see also Stachowski S. 2002]
- Wedel A. 2000, Perceptual Distinctiveness in Turkish Emphatic Reduplication. – Billerey R., Lillehaugen B.D. (ed.), *Proceedings of the 19th West Coast Conference on Formal Linguistics*, Sommersville: 546–559 [after Müller 2004]
- Wedel A. 2003?, *Preservation of Contrast in Turkish Reduplicative Paradigms*, Santa Cruz [after Müller 2004]
- Weidmann N.B., Rød J.K., Cederman L.-E. 2010, Representing Ethnic Groups in Space: a New Dataset. – *Journal of Peace Research* 47/4, 491–499
- Wickham H. 2009, *ggplot2: elegant graphics for data analysis*, New York
- Wurm S. 1945, *Das özbekische Dialekt von Andidschan. Phonetische und morphologische Studien. Texte. Teil I*, Brünn – München – Wien
- Wurm S. 1949, The (Kara-)Kirghiz Language. – *Bulletin of the School of Oriental and African Studies, University of London* 13/1, 97–120
- Zajączkowski A. 1932, *Sufiksy imienne i czasownikowe w języku zachodniokaraimskim* (= *Prace Komisji Orientalistycznej Polskiej Akademji Umiejętności* 15), Kraków
- Zeynalov F. 1993, *Türk Lehçelerin Karşılaştırmalı Dilbilgisi*, İstanbul [trans. by Y. Gedikli]
- Zieme P. 1985, *Buddhistische Stabreimdichtungen der Uiguren* (= *Schriften zur Geschichte und Kultur des Alten Orients. Berliner Turfantexte* 13), Berlin [after Müller 2004]
- Žäfärov 1984 = Чәфәров Г., *Азәрбајҹан дилинин лексик-семантик системи*, Бақы