

ISSN 2500-2902  
ISBN 978-1-4632-0168-5

# Ural-Altaic Studies

*Scientific Journal*

**№ 1 (32) 2019**

Established in 2009  
Published four times a year

**Moscow**

© Institute of Linguistics, Russian Academy of Sciences, 2019  
© Tomsk State University, 2019

ISSN 2500-2902  
ISBN 978-1-4632-0168-5

# Урало-алтайские исследования

*научный журнал*

**№ 1 (32) 2019**

Основан в 2009 г.  
Выходит четыре раза в год

**Москва**

© Институт языкознания Российской академии наук, 2019  
© Томский государственный университет, 2019

## CONTENTS

No 1 (32) 2019

<b>Roman V. Gaidamashko, Larisa G. Ponomareva.</b> Materials on history of the study of the Komi-Permyak manuscript heritage of Archpriest Antony Popov (1784—1788) .....	7
<b>Polina I. Li.</b> An Experimental phonetic study of the first-syllable vocalism in the Vakh dialect of Khanty .....	28
<b>Yulia V. Normanskaya.</b> Who stood at the beginnings of the Crimean Tatar literary language? .....	41

### DISCUSSIONS

<b>Anna V. Dybo, Philip S. Krylov, Vera S. Maltseva, Aleksandra V. Sheimovich.</b> Segmental rules in the automatic parser for the Khakass corpus .....	48
<b>Ilya B. Itkin.</b> Khakass adjectives on <i>-кЫ ~ -зЫ</i> (an addition to the article “Segmental rules in the automatic parser for the Khakass corpus” by A. Dybo et al.) .....	70
<b>Anna V. Dybo.</b> Continuation of the discussion of the article “Khakass Adjectives on <i>-кЫ ~ -ГЫ</i> ” .....	72
Style sheet .....	74

## СОДЕРЖАНИЕ

Номер 1 (32) 2019

<b>Р. В. Гайдамашко, Л. Г. Пономарева.</b> Материалы к истории изучения коми-пермяцкого рукописного наследия протоиерея Антония Попова (1784—1788).....	7
<b>П. И. Ли.</b> Экспериментально-фонетическое исследование вокализма первого слога в ваховском диалекте хантыйского языка .....	28
<b>Ю. В. Норманская.</b> Кто стоял у истоков создания крымско-татарского литературного языка?.....	41

### ДИСКУССИИ

<b>А. В. Дыбо, Ф. С. Крылов, В. С. Мальцева, А. В. Шеймович.</b> Сегментные правила в автоматическом парсере Корпуса хакасского языка .....	48
<b>И. Б. Иткин.</b> Хакасские прилагательные на <i>-кы ~ -гы</i> (дополнение к статье А. В. Дыбо, Ф. С. Крылова, В. С. Мальцевой и А. В. Шеймович «Сегментные правила в автоматическом парсере Корпуса хакасского языка»).....	70
<b>А. В. Дыбо.</b> Продолжение дискуссии по заметке «Хакасские прилагательные на <i>-Кы ~ -Гы</i> » .....	72
Требования к оформлению статей.....	74

Главный редактор

**А. В. Дыбо**

(д.ф.н., чл.-корр. РАН, Институт языкознания РАН)

Заместитель главного редактора

**Ю. В. Норманская**

(д.ф.н., Институт языкознания РАН)

*Редакционная коллегия*

В. Ю. Гусев (к.ф.н., Институт языкознания РАН), П. П. Дамбуева (д.ф.н., проф., Институт языкознания РАН), Е. В. Кашкин (к.ф.н., Институт русского языка РАН), О. А. Мудрак (д.ф.н., проф., Российский государственный гуманитарный университет), С. А. Мызников (д.ф.н., проф., Институт лингвистических исследований РАН), И. Николаева (PhD, Школа восточных и африканских исследований, Великобритания), Ф. Ш. Нуриева (д.ф.н., проф., Казанский (Приволжский) федеральный университет), Г. Ц. Пюрбеев (д.ф.н., проф., Институт языкознания РАН), М. Роббеетс (PhD, Лейденский университет, Нидерланды), И. Я. Селютина (д.ф.н., проф., Институт филологии СО РАН), Р. А. Тадинова (д.ф.н., доц., Институт языкознания РАН), З. Н. Экба (к.ф.н., Институт языкознания РАН), А. Б. Шлуинский (к.ф.н., Институт языкознания РАН)

*Редакционный совет*

В. М. Алпатов (д.ф.н., чл.-корр. РАН, Институт языкознания РАН), А. Е. Аникин (д.ф.н., чл.-корр. РАН, Институт филологии СО РАН), Р. Г. Ахметьянов (д.ф.н., проф., Бирский государственный педагогический институт), М. Бакро-Надь (проф., Университет Сегеда, Венгрия), В. Блажек (проф., Масариков университет, Чехия), Т. М. Гарипов (д.ф.н., чл.-корр. НАН РБ, Башкирский государственный педагогический институт им. М. Акмуллы), Н. И. Егоров (д.ф.н., проф., Чувашский государственный институт гуманитарных наук), И. В. Кормушин (д.ф.н., проф., Институт языкознания РАН), И. Л. Кызласов (д.и.н., проф., Институт археологии РАН), Й. Лааксо (проф., Венский университет, Австрия), К. М. Мусаев (д.ф.н., чл.-корр. РАЕН, Институт языкознания РАН), И. А. Невская (проф., Франкфуртский университет им. Иоганна Вольфганга Гёте, Германия), Т. Ризе (проф., Венский университет, Австрия), Е. К. Скрибник (проф., Мюнхенский университет, Германия), П. А. Слепцов (д.ф.н., проф., акад. НАН РСЯ, Институт гуманитарных исследований и проблем малочисленных народов Севера СО РАН), М. Стаховски (проф., Краковский университет, Польша), Ф. Г. Хисамитдинова (д.ф.н., проф., Институт истории, языка и литературы Уфимского научного центра РАН), Л. Хонти (проф., Будапештский университет, Венгрия), Н. Н. Ширококова (д.ф.н., проф., Институт филологии СО РАН), К. Шониг (проф., Берлинский университет им. Гумбольдта, Германия), М. Эрдал (проф., Франкфуртский университет им. Иоганна Вольфганга Гёте, Германия)

Editor-in-Chief

**Anna Dybo**

(Institute of Linguistics, Russian Academy of Sciences)

Deputy Editor-in-Chief

**Yulia Normanskaya**

(Institute of Linguistics, Russian Academy of Sciences)

*Editorial Board*

Polina Dambueva (Institute of Linguistics, Russian Academy of Sciences), Zarema Ekba (Institute of Linguistics, Russian Academy of Sciences), Valentin Gusev (Institute of Linguistics, Russian Academy of Sciences), Egor Kashkin (Institute of the Russian language, Russian Academy of Sciences), Oleg Mudrak (Russian State University for the Humanities), Sergey Myznikov (Institute for Linguistic Studies, Russian Academy of Sciences), Irina Nikolaeva (School of Oriental and African Studies, Great Britain), Fanuza Nurieva (Kazan (Volga region) Federal University), Grigoriy Pyurbuev (Institute of Linguistics, Russian Academy of Sciences), Martine Robbeets (Leiden University, Center for Linguistics, Institute for Area Studies, the Netherlands), Irina Selyutina (Institute of Philology, the Siberian Branch of Russian Academy of Sciences), Andrey Shluinski (Institute of Linguistics, Russian Academy of Sciences), Roza Tadinova (Institute of Linguistics, Russian Academy of Sciences)

*Advisory Board*

Rinat Ahmet'yanov (Birk State Pedagogical Institute), Vladimir Alpatov (Institute of Linguistics, Russian Academy of Sciences), Alexandr Anikin (Institute of Philology, the Siberian Branch of Russian Academy of Sciences), Marianne Bakró-Nagy (University of Szeged, Hungary), Václav Blažek (Masaryk University, Czech Republic), Nikolay Egorov (Chuvash State Institute for the Humanities), Marcel Erdal (Goethe University Frankfurt, Germany), Talmas Garipov (M. Akmulla Bashkir State Pedagogical University), Firdaus Hisamitdinova (Institute of History, Language and Literature, Ufa Scientific Centre of Russian Academy of Sciences), László Honti (University of Budapest, Hungary), Igor Kormushin (Institute of Linguistics, Russian Academy of Sciences), Igor Kyzlasov (Institute of Archaeology, Russian Academy of Sciences), Johanna Laakso (University of Vienna, Austria), Kenesbay Musaev (Institute of Linguistics, Russian Academy of Sciences), Irina Nevskaya (Goethe University Frankfurt, Germany), Timothy Riese (University of Vienna, Austria), Klaus Schönig (Humboldt University of Berlin, Germany), Natalia Shirobokova (Institute of Philology, the Siberian Branch of Russian Academy of Sciences), Elena Skribnik (University of Munich, Germany), Piotr Slepsov (Institute of Humanities Research and Indigenous Studies of the North, the Siberian Branch of Russian Academy of Sciences), Marek Stachowski (Krakow University, Poland)

## Сегментные правила в автоматическом парсере Корпуса хакасского языка<sup>1</sup>

### 0. Введение

Уже несколько лет при поддержке различных фондов<sup>2</sup> ведется работа над проектом создания и разработки корпусов миноритарных тюркских языков России. В рамках этого проекта создаются параллельные корпуса (все тексты обеспечены русским переводом) с морфологической разметкой. В наиболее близком к рабочему состоянию сейчас находится корпус хакасского языка<sup>3</sup>.

В корпус хакасского языка сейчас введены хакаские литературно-художественные тексты; эпические тексты; газетные тексты; переводы на хакасский язык [Новый Завет 2011, Алиса в стране чудес 2017]. Переводы на русский язык, при отсутствии печатных версий таковых, делались участниками проекта — носителями хакасского языка. В распоряжении составителей корпуса есть оцифрованная версия Большого хакасско-русского словаря на 22 тыс. слов под ред. О. В. Субраковой [БХРС 2006]. Правка распознанного материала произведена участниками проекта, после чего текстовый файл преобразован в базу данных Starling, и словарная база дорабатывается участниками в этой системе. В качестве эталонного текстового материала мы используем также иллюстративный материал к словарю и оцифрованные текстовые примеры из [ГХЯ 1975].

Кроме того, постепенно создается диалектный подкорпус хакасского языка. Диалектный подкорпус содержит тексты звучащей речи на всех традиционно выделяемых и сохранившихся к настоящему времени хакасских диалектах; фонетическую транскрипцию — для некоторых текстов; запись в орфографии, приближенную к хакасской литературной норме; расчлененную морфонологическую запись, морфологическую разметку и русский перевод. К настоящему времени в рамках проекта обработаны полевые материалы по трем хакасским диалектам (качинский, сагайский и бельтырский). По сагайскому диалекту (с. Казановка) обработаны тексты, собранные экспедициями РГГУ (2001, 2002 гг.), РГГУ и ИЯз РАН (2007 г.), около 2,5 часов звучания, 12 тыс. словоформ (расшифровка, перевод, глоссирование, привязка к звуку). По бельтырскому диалекту (сс. Чиланы, Карагай, Бутрахты) полностью обработаны тексты, собранные экспедицией ИЯз РАН 2011 г., 9,5 часов звучания, 43 тыс. словоформ (расшифровка, перевод, глоссирование, привязка к звуку), планируется дополнить материалами, собранными в 2018 г. По качинскому диалекту (экспедиция ИЯз РАН 2015 г., сс. Трошкино, Черное озеро расшифровано и переведено около 5 часов записи текстов (около 30 тыс. словоформ), идет подготовка к разметке. Также собрано значительное количество текстов на шорском диалекте (сёла Таштыпского р-на, экспедиции ИЯз РАН и ТГУ 2016—2017 гг.), несколько часов из них расшифрованы. Звуковые файлы и расшифровка текстов, размеченная по времени, вывешены в Интернете<sup>4</sup>, по окончании глоссирования предполагается размещение глоссированных файлов в формате ELAN также в Интернете, в частности, на платформе LingvoDoc<sup>5</sup>. Кроме текстов, в диалектном подкорпусе вывешиваются различные анкеты-опросники, звуковые файлы в сопровождении расшифровки (100-словники, 200-словники, основанные на историческом принципе большие фонетические опросники (см. [Дыбо, Мудрак 2009]), морфологические и синтаксические опросники).

---

<sup>1</sup> Работа выполнена при поддержке грантов Правительства РФ № 14.Y26.31.0014, проект «Языковое и этнокультурное разнообразие Южной Сибири в синхронии и диахронии: взаимодействие языков и культур» (сбор и обработка материала); РНФ № 18-18-00501, проект «Создание электронного диалектологического атласа тюркских языков России» (написание статьи); проект Программы фундаментальных исследований Президиума РАН «Памятники материальной и духовной культуры в современной информационной среде», направление «Развитие корпусов миноритарных тюркских языков России» (программное обеспечение анализа).

<sup>2</sup> См. список [khakas.altaiica.ru/texts/](http://khakas.altaiica.ru/texts/)

<sup>3</sup> [khakas.altaiica.ru/corpus/](http://khakas.altaiica.ru/corpus/)

<sup>4</sup> <https://dialects.altaiica.ru/>

<sup>5</sup> <http://lingvodoc.ispras.ru>

Мы уже писали об особенностях идеологии и структуры автоматического морфологического анализатора, работающего на сайте корпуса ([Мальцева 2004; Дыбо, Шеймович 2014])<sup>6</sup>. Упомянем здесь кратко основные отличия от модели хакасской грамматики, принятой в существующих описаниях [ГХЯ 1975; Дыренкова 1948; Баскаков 1953; Карпов 2007].

А) Мы выделяем в хакасском языке три грамматических класса — имя, глагол и неизменяемое (неизменяемые делятся на два подкласса, один из которых может присоединять частицы, другой не может) — на основании чисто грамматических критериев: наборов словоизменяемых грамматических категорий (следуя в этом за [Зализняк 1967]). Те части речи, которые выделяются в хакасских грамматиках помимо этого набора грамматических классов (местоимения, числительные; прилагательные; отчасти наречия), мы считаем семантическими подклассами имени, особые синтаксические свойства которых могут проявляться как повышенная частотность каких-то типов употреблений, а не как системные запреты и разрешения, и потому должны выявляться в результате корпусного исследования, а не встраиваться в парсер, на основании работы которого впоследствии будут проводиться корпусные исследования. Имя отличается от глагола урезанным количеством словоизменяемых грамматических категорий.

Б) Мы принципиально считаем бессмысленным построение словоизменяемой грамматики тюркского языка как набора парадигм; эти агглютинативные языки, по-видимому, следует считать непарадигматическими (ср., например, [Алпатов 2018: 35—43]). Как для целей автоматического анализа, так и для целей описания морфологии тюркских языков мы полагаем наиболее адекватной грамматику порядка с циклами. Циклы в области глагольного набора категорий исторически объясняются как следствие синтезирования ранее аналитических форм. Циклы в области именных категорий появляются вследствие возможности продвижения словоформы, маркированной по падежу/числу/посессивности, в вершину именной группы (при помощи аффиксов транспозиции или при помощи безаффиксной конверсии).

В) Свойство обязательности выражения не характерно для грамматических категорий тюркских языков, каждый грамматический слот может быть заполнен или не заполнен, и морфологические нули в общем случае не вводятся в описание<sup>7</sup>.

В настоящей статье мы остановимся на сегментных преобразованиях в грамматике синтеза хакасской словоформы, обращение которых используется в программе автоматического анализа. Мы излагаем правила по уровням, в том порядке, в котором они следуют в грамматике синтеза словоформы, поскольку анализ осуществляется «через синтез», т. е. путем многократных прямых и обратных проходов с проверкой гипотез. Инвентарь словоизменяемых аффиксов, к которым применяются описываемые правила, см. в Приложении, табл. 1. Стрелкой → обозначаются одношаговые морфонологические преобразования, знаком «больше» > обозначаются преобразования, состоящие из нескольких одношаговых преобразований. По умолчанию аффикс выписывается в морфонологической форме, поэтому при его изолированном приведении фигурные скобки {} могут опускаться (но они выписываются вокруг набора морфов, составляющих словоформу, если он рассматривается на морфонологическом уровне).

## 1. Морфонологическая запись

Инвентарь сегментов (букв) для записи означающих хакасских словоформ.

### 1.0. Вокализм<sup>8</sup>

Хакасские гласные различаются по долготе/краткости, по ряду: передний/задний, по подъёму: широкие (нижнего и среднего подъёма) / узкие (верхнего подъёма).

<sup>6</sup> Ср. также модель сегментных преобразований тувинской словоформы, подготовленную для аналогично устроенного тувинского парсера, в [Dybo, Sheimovich 2017].

<sup>7</sup> Более частные отличия данного описания от традиционных грамматик изложены в статье [Шеймович 2017].

<sup>8</sup> Речь идет не о фонетических характеристиках описываемых звуковых сегментов, а о признаках, характеризующих абстрактные сегментные единицы морфонологического уровня языка — морфемы. В набор признаков попадают такие, от которых зависят действия правил сегментных преобразований. Морфема в нашем понимании — это единица инвентаря символов, с помощью которого записываются означающие морфем в словаре морфем и в лексиконе. Соответственно, инвентарь устроен так, чтобы по возможности одна морфема имела в словаре одну и только одну запись (но омонимичные морфемы бывают). Для облегчения использования человеком электронного грамматического словаря хакасского языка мы выписываем в морфонологической транскрипции, т. е. с использованием условных букв, только звуковую оболочку аффиксов, а не корней (морфонологическая транскрипция словообразовательных аффиксов дается внутри расчлененной морфонологической записи основы в поле

**1.0.0.** Сочетания согласной *й* с гласными могут альтернативно записываться в орфографии с помощью русских йотированных букв. Соответственно, *я=йа*; *ю=йу*. Употребление гласных букв *е, э* в хакасских заимствованных словах подчиняется правилу распределения: в абсолютном начале встречаются только *э, ээ*, не в абсолютном начале встречаются только *е, ее*. Буквы *е, ё, ю, я* используются также в заимствованиях из русского языка: *Таня, бюро, ликёр*; сегментные правила воспринимают их как соответствующие нейотированные буквы *е, о, у, а*.

**1.0.1.** *Краткие гласные:* *а, е, ы, і, и* (в первом слоге исконной основы, в любом слоге заимствований), *о* (в первом слоге исконной основы, в любом слоге заимствований), *ө* (в первом слоге), *у, ү*.

*Долгие гласные:* все двойные написания кратких гласных, независимо от позиции: *аа, ее, ыы, ии, оо, өө, уу, үү, ээ* (*іі* — запрещенное сочетание); гласные *и, о, ө* в непервом слоге в незаимствованной морфеме (случай типа *парц* ‘уходя’, *пызо* ‘теленок’, *көзө* ‘сучок’). Таким образом, правила выбора алломорфов и правила сегментных преобразований, учитывающие долготу предшествующей гласной, после *и, о, ө* непервого слога работают как после долгой гласной. Так, выпадение согласной после долгой гласной невозможно: *пар-и-ГА(н) > париган* ‘уходил’, *салаа-ГА > салаага* ‘пальцу’, *пызо-ГА > пызога* ‘телёнку’ — см. ниже.

**1.0.2.** Широкие гласные: *а (я), аа (яа), е (э), ее (ээ), о(ё), оо, ө, өө*;

Узкие гласные: *ы, ыы, и, ии, і, у(ю), уу (юу), ү, үү*.

**1.0.3.** Задние гласные: *а (я), аа (яа), ы, ыы, о, оо, у(ю), уу(юу)*;

Передние гласные: *е (э), ее (ээ), і, ө, өө, ү, үү*.

В литературной норме гласные *и, ии* ведут себя как передние, когда находятся в первом слоге, т. е. в последующих слогах гласные также должны быть передними: *пистер* ‘пятерки’, *тииннер* ‘белки’, *чистек* ‘ягода’. Гласные *и, ии* непервых слогов в литературном языке являются немаркированными по рядности (нейтральными), и выбор варианта аффикса зависит от гласных предыдущих слогов. Как правило, *и* в непервом слоге появляется в составе аффикса, либо на поверхностном уровне в результате стяжения на границе двух морфем, одна из которых заканчивается, а другая начинается на гласный (*{таста-ир-ГА}* ‘бросать-FUT-DAT’ > *тастирга* ‘бросать’, *{төре-ир-ГА}* ‘рождаться-FUT-DAT’ > *төрирге* ‘рождаться’)<sup>9</sup>, а *ии* в результате стяжения с выпадением интервокального согласного (*{сурыг-(з)Ы-ЛАР}* ‘вопрос-3ros-PL’ > *сурилар* ‘его вопросы’, *{пөлік-(з)Ы-ЛАР}* ‘комната-3ros-PL’ > *пөлилер* ‘его комнаты’). Гласная *и* непервого слога в неадаптированных русских заимствованиях (помета *rus* в поле словарной базы ЕТУМ) ведет себя как передняя, и следующие за ней аффиксы выбирают переднерядные варианты: *академигі* ‘академик чего-то’, *магазинзер* ‘в магазин’. В ряде говоров сагайского диалекта гласные *и, ии* также ведут себя в непервом слоге как передние: саг. *париген* ~ лит. *париган* ‘уходил’, саг. *улиге* ~ лит. *улига* ‘старшему’.

## 1.1. Консонантизм

Согласные хакасского литературного языка:

Глухие: *п, т, к, х, ч, с, ш, ф, ц, ц*.

Звонкие: *б, д, г, з, ж, в*; сонорные: *м, н, ң* (носовые), *л, р* (плавные), *й*.

Сегменты *ф, в, ш, ж, ц, ц* встречаются только в заимствованиях, в основном из русского языка [*ш/ж* — также в диалектных шорских и кызыльских словах, вошедших в литературный язык].

Заднеязычные взрывные согласные *к, г* согласно правилу аккомодации согласных литературного хакасского языка могут встречаться в исконно хакасских корнях и аффиксах только при наличии в той же словоформе переднерядных гласных (*е, и* первого слога, *і, ө, ү*), а увулярные щелевые *з, х* — только при

---

словарной базы DERIV: *{аал=ЛА=(Ы)с-}* (> *ааллас*) ‘гостить друг у друга’. Это решение является чисто формальным. Отсутствие условной морфонологической записи корней компенсируется информацией из полей словарной базы ALTERNAT, ALTERNATEN и FORM; поскольку соответствующая информация является релевантной для небольшого числа основ, такое решение представляется оправданным. Ниже в тексте статьи мы приводим некоторые примеры основ в условной морфонологической записи.

<sup>9</sup> Гласная *-и-* в аффиксах исторически восходит к стяжению на границе аффиксов сочетания *\*(V)jV*, ср. например, древнетюркский показатель отрицательного деепричастия *-тауXн* [Erdal 2004: 324] и хак. *-Пин*; древнетюркский показатель IMP.1SG *-(A)yIn* [Erdal 2004: 324] и хак. *-им*; древнетюркский показатель «аориста» после вокалических основ *-yUr* [Erdal 2004: 240] и хак. *-ир*.



наличии в той же словоформе заднерядных (*a, ы, o, y*). В неадаптированных заимствованиях *к, з* используются независимо от рядности окружающих гласных, в соответствии с русской орфографией: *анестезиолог, автокружок*.

Для сегментных правил при синтезе словоформ релевантно главным образом противопоставление согласных по глухости/звонкости. Звонкие несонантные *б, в, д, ж, з, ч* в позиции конца основы могут встречаться только перед гласной или звонкой согласной (*б, д, з*); либо в русских заимствованиях. При этом морфонологически они тождественны соответствующим глухим (поскольку фонетически были восприняты из начальной формы русского слова в препаузальной позиции, где по правилам русского произношения звонкие согласные оглушаются), а пишутся как звонкие они ради сохранения русской орфографической формы. Поэтому после этих звонких согласных действуют те же правила выбора алломорфов, что и после глухих: {завод} + {ЛАр} PL > заводтар 'заводы', {скандинав} + {ЛАр} PL > скандинавтар 'скандинавы' (ассимиляция: оглушающиеся *д > т, в > ф* требуют после себя алломорфа мн. ч. с глухим). То же самое касается *з* конца слова в неадаптированных русских заимствованиях: ср. *биологтар истезиглери* 'исследования биологов', где после заимствования *биолог* выбирается глухой вариант аффикса ЛАр, а после исконного *истезиг* — звонкий вариант (ср. ниже, распределение вариантов аффиксов в таблице 2). То есть, в морфонологической записи основ в словаре заимствованные основы должны были бы записываться с помощью специальных морфем; на практике выбрано более простое решение: соответствующие правила содержат ссылку на наличие пометы *rus* в поле словарной базы ЕГУМ.

Условная морфонологическая запись аффиксов включает, кроме строчных букв хакасского алфавита (в случае однозначного соответствия поверхностному графическому выражению в словоформах), прописные буквы, соответствующие дополнительным морфемам, которые в поверхностном выражении репрезентируются как фонемы, чередующиеся в зависимости от фонетических позиций (соответственно, в орфографическом выражении им могут соответствовать разные буквы). Аффиксы в морфологической таблице 1 (см. Приложение) записаны с помощью морфем, что позволяет не выписывать в каждом случае весь набор алломорфов того или иного аффикса:

Дополнительные согласные морфемы	Дополнительные гласные морфемы
П: б/п/м	А: е/а
К: з/з/х/к	Ы: и/ы
Г: з/з/х/к/∅	О: о/ö
Т: т/д	
Д: т/д/н	
С: с/з	
Л: л/н/т	
Л: л/н <sup>10</sup>	
Н: н/т	
Ч: ч/ч <sup>11</sup>	

О правилах выбора репрезентаций дополнительных морфем в разных контекстах см. пункт 2.2.1 ниже.

## 2. Правила сегментных преобразований морфов хакасского языка

Необходимо отметить, что здесь мы рассматриваем только те правила, которые проявляются орфографически, т. е. морфологический анализатор работает только с письменными текстами. Для парсинга диалектных текстов мы сначала производим нормализацию текста, приближая его запись к литератур-

<sup>10</sup> В словоизменительной грамматике хакасского эта морфема встречается только в показателе CONT (встраиваемая континуативная частица, позиция 3 в таблице 1). В CONT, в отличие от других словоизменительных показателей, начинающихся с *л-* в интервокальной позиции, отсутствует вариант с *-т-* после глухих (употребляется вариант с *-л-*): *Хайдар-да нимелер тартыпдлачалар, тартыпдлачалар* 'Все что-то возят и возят (без конца, только этим и занимаются)'; *Ўзигі чох пасдлачазың* 'Без конца только и пишешь и пишешь' (см. ниже, табл. 2).

<sup>11</sup> В современной литературной орфографической норме показатели PRES *чА*, DUR *чАТ* и PRES<sub>2</sub> *чАдЫ(р)* всегда пишутся с *ч-*, в отличие от НАВ *ЧА(н)* и PROL *ЧА*, которые пишутся с *ч-* или *ч-* в зависимости от позиции. Однако в современной произносительной норме все эти показатели при выборе глухого/звонкого начала ведут себя одинаково, т. е. начальная согласная озвончается после гласных и звонких согласных. Отсюда написания со звонким, которые иногда можно встретить в текстах: *пардаача* 'и он пишет'. Технически мы принимаем для анализа таких случаев (пока редких) вариативный морфемный состав для показателей PRES *чА*, *ЧА*, DUR *чАТ*, *ЧАТ* и PRES<sub>2</sub> *чАдЫ(р)*, *ЧАдЫ(р)*.

ной норме. Впрочем, некоторые частотные диалектные морфемы и дополнительные варианты литературных морфем введены в парсер с пометой *dial*. Словарь также дополнен некоторыми частотными диалектными лексемами, встреченными в текстах. Следует отметить, что литературная нормированность хакасского языка не особенно жесткая, и хакасские писатели — носители диалектов легко вносят в свои произведения привычные для них диалектные формы.

## 2.0. Сегментные правила выбора алломорфов (поверхностно-морфологические правила выбора одного из алломорфов морфемы, при заданном морфемном составе словоформы)

### 2.0.1. Выбор из алломорфов, стоящих в одной клетке таблицы 1 через запятую: *A* vs *u*.

Алломорфы, различающиеся первой морфемой *A* или *u* (FUT *A(p)*, *u(p)*; CONV<sub>a</sub> *A*, *u*; CONV<sub>пас</sub> *Abac*, *ибAc*; PROSP *AK*, *uK*; ITER *AдЫ(p)*, *идЫ(p)*), выбираются в зависимости от того, на что кончается предыдущий морф. Если он кончается на гласную, выбирается вариант с *u*, если на согласную — вариант с *A*. После гласной: {узу-*up*} (> *узup*) ‘он будет спать’, {сөле-*ибAc*} (> *сөлибес*) ‘говоря’, {төзе-*u*} (> *төзи*) ‘расстилая’, {сана-*идЫp*} (> *санидыр*) ‘считает’, {узу-*uK-чA*} (> *узихча*) ‘вот-вот заснет’. После согласной: {ат-*Ap*} (> *адар*) ‘он будет стрелять’, {нар-*AbAc*} (> *парабас*) ‘уходя’, {сал-*A*} (> *сала*) ‘кладя’, {пiл-*AдЫp*} (> *пiледip*) ‘знает’, {нар-*AK-чA*} (> *парахча*) ‘собирается уходить’.

### 2.0.2. Выбор из алломорфов, стоящих в одной клетке таблицы 1 через запятую: *(Ы)п* vs *θ*.

Показатель NF выбирает алломорф *(Ы)п*, если: 1) он непосредственно следует за основой, которая оканчивается на (выпадающие, см. ниже) *п*, *з*, *э* или *ң*; или 2) он непосредственно следует за основой или аффиксом, оканчивающимися на гласную; или 3) непосредственно за ним следует интраклитика: ASS *OK* или CONT *LA* (хотя в последнем случае может выбираться и алломорф *θ* при основах на согласную). В прочих случаях NF выбирает алломорф *θ*, если находится в середине словоформы. Непосредственно в конце словоформы для NF возможны оба алломорфа. Перед глаголом, начинающимся с гласной, выбирается обязательно алломорф *(Ы)п*: *алып алча* ‘берет себе’; перед глаголом, начинающимся на согласную, может выбираться алломорф *θ*: *ал салган* ‘взял’. Но, так как парсер не анализирует аналитические словоформы, то распределение для него остается свободным)<sup>12</sup>.

NF.NEG не опускается: {нар-*θ-чA*} ‘идти-NF-PRES’ (> *нарча*) ‘он идет’, {нар-*Пин-чA*} ‘идти-NF.NEG-PRES’ (> *нарбинча*) ‘он не идет’.

### 2.0.3. Выбор из алломорфов, стоящих в одной клетке таблицы 1 через запятую: *тЫп* vs *Даң*.

Алломорф ABL *тЫп* может стоять только после ALL *CAp*: *столзартын* ‘со стороны стола’. Алломорф ABL *Даң* употребляется во всех прочих случаях: *столдаң* ‘со стола’. (О соответствии набора падежных показателей в данном грамматическом описании традиционному см. [Шеймович 2017].)

2.0.4. Показатель DUR<sub>1</sub> *u(p)* перед PAST *ГA(н)* и перед COND *CA* выступает как *u*: {нар-*u(p)-ГAн*} ‘идти-DUR<sub>1</sub>-PAST’ → {нар-*u-ГAн*} (> *париган*) ‘уходил’, {нар-*u(p)-CA*} ‘идти-DUR<sub>1</sub>-COND’ → {нар-*u-CA*} (> *париза*) ‘когда идет’, перед деепричастием CONV<sub>п</sub> — как *up*: {нар-*u(p)-(Ы)п*} ‘идти-DUR<sub>1</sub>-CONV<sub>п</sub>’ → {нар-*up-Ып*} (> *парирып*: *ирт парирып* ‘проходя мимо’). Перед Person, PredPL и концом словоформы варианты *u* и *up* находятся в свободном (точнее, диалектном) варьировании: {нар-*u(p)*} > *пару/парир* ‘он идет’; {нар-*u(p)-ЛАp*} > *парилар/парирлар* ‘они идут’. В литературном языке *p* обычно не опускается. Как и в случае FUT, описанном ниже в 2.0.7, наблюдается тенденция к распределению показателей 1SG: непосредственно перед 1SG.mix показатель принимает форму *u*: {нар-*u(p)-(Ы)м*} → {нар-*u-(Ы)м*} (> *парим* ‘я иду’), перед 1SG.full — форму *up*: {нар-*u(p)-ПЫп*} → {нар-*up-ПЫп*} (> *парирбын* ‘я иду’).

2.0.5. Показатель PAST *ГA(н)* принимает форму *ГA* непосредственно перед всеми показателями позиции 19 Person: {нар-*ГA(н)-(Ы)м*} ‘идти-PAST-1SG.mix’ → {нар-*ГA-м*} (> *паргам*) ‘я шел’, {мик-*ГA(н)-CAp*} ‘шить-PAST-1SG.mix’ → {мик-*ГA-CAp*} (> *миккесер*) ‘вы шили’. В остальных случаях принимается форма *ГAн*.

<sup>12</sup> Таким образом, если у нас имеется форма, заканчивающаяся на *-(Ы)п*, то она имеет два разбора: через CONV<sub>п</sub> — в случае свободной конструкции, и через NF — в случае грамматикализованной аналитической словоформы, например, со вспомогательным глаголом *ал-* ‘брать’ (бенефактив). Показатель NF — это единственный грамматический ноль, введенный нами в систему; в принципе его можно было бы отождествить с показателем деепричастия CONV<sub>п</sub>, распределив его по двум разным слотам (что соответствует исторической правде и некоторым особенностям постановки отрицания в глагольных формах), они совершенно омонимичны. Главное, чем вызвано принятое решение — омонимия глоссировки формы деепричастия с нулевым показателем и формы императива 2SG при ином решении.

**2.0.6.** Показатель НАВ  $ЧА(н)$  может принимать как форму  $ЧА$ , так и форму  $ЧАн$  непосредственно перед всеми показателями позиции 19 Person. В 1SG, если НАВ принял форму  $ЧА$ , то выбирается морфема личного аффикса  $(БІ)м$ , а если НАВ принял форму  $ЧАн$ , выбирается морфема личного аффикса  $ПЫн$ :  $\{ырла-ЧА(н)-ПЫн\}$  ‘петь-НАВ-1SG.full’  $\rightarrow \{ырла-ЧАн-ПЫн\}$  ( $>$  *ырлачаңмын*) //  $\{ырла-ЧА(н)-(БІ)м\}$  ‘петь-НАВ-1SG.mix’  $\rightarrow \{ырла-ЧА-м\}$  ( $>$  *ырлачам*) ‘я пел (обычно)’, но:  $\{кил-ПА-ЧА(н)-ПЫс\}$  ‘прийти-НАВ-1PL’  $\rightarrow \{кил-ПА-ЧАн-ПЫс\}$  ( $>$  *килбечеңміс*) //  $\{кил-ПА-ЧА-ПЫс\}$  ( $>$  *килбечебіс*) ‘мы не приходили (обычно)’.

**2.0.7.** Показатель FUT  $A(p)$ ,  $u(p)$  в литературном языке может принимать форму  $A$ ,  $u$  непосредственно перед 1SG.mix:  $\{нар-А(p)-(БІ)м\}$  ‘идти-FUT-1SG’  $\rightarrow \{нар-А-м\}$  ( $>$  *парам*) ‘я пойду’,  $\{ада-и(p)-(БІ)м\}$  ‘называть-FUT-1SG’  $\rightarrow \{ада-и-м\}$  ( $>$  *адим*) ‘я буду называть’,  $\{кил-А(p)-(БІ)м\}$  ‘прийти-FUT-1SG’  $\rightarrow \{кил-А-м\}$  ( $>$  *килем*) ‘я приду’. Во всех диалектах FUT часто принимает форму  $A$ ,  $u$  непосредственно перед показателем DAT  $ГА$ :  $\{нар-А(p)-ГА\}$  ‘идти-FUT-DAT’  $\rightarrow \{нар-А-ГА\}$  ( $>$  *парага*) ‘идти’, перед AFFIRM  $ЧЫК$ :  $\{нар-А(p)-ЧЫК\}$  ‘идти-FUT-AFFIRM’  $\rightarrow \{нар-А-ЧЫК\}$  ( $>$  *парачык*) ‘пошёл бы’, а также перед другими личными показателями:  $\{нар-А(p)-СЫң\}$  ‘идти-FUT-2SG’  $\rightarrow \{нар-А-СЫң\}$  ( $>$  *паразың*) ‘ты пойдешь’,  $\{нар-А(p)-ПЫс\}$  ‘идти-FUT-1PL’  $\rightarrow \{нар-А-ПЫс\}$  ( $>$  *парабыс*) ‘мы пойдём’. В кызыльском диалекте форма  $A$ ,  $u$  встречается также и в конце словоформы: *чари* ‘ладно’ ( $<$   $\{чара-А\}$  ‘быть годным-FUT’), лит. *чарир*. В прочих случаях принимается форма  $Ap$ ,  $up$ . В частности, перед 1SG.full в лит. яз.:  $\{нар-А(p)-ПЫн\}$  ‘идти-FUT-1SG’  $\rightarrow \{нар-Аp-ПЫн\}$  ( $>$  *парарбын*) ‘я пойду’,  $\{ада-и(p)-ПЫн\}$  ‘называть-FUT-1SG’  $\rightarrow \{ада-ип-ПЫн\}$  ( $>$  *адирбын*) ‘я буду называть’,  $\{кил-А(p)-ПЫн\}$  ‘прийти-FUT-1SG’  $\rightarrow \{кил-Аp-ПЫн\}$  ( $>$  *килербин*) ‘я приду’ и 1SG.mix в диалектах:  $\{нар-А(p)-(БІ)м\}$  ‘идти-FUT-1SG’  $\rightarrow \{нар-Аp-БІм\}$  ( $>$  *парарым*) ‘я пойду’,  $\{кил-А(p)-(БІ)м\}$  ‘прийти-FUT-1SG’  $\rightarrow \{кил-Аp-БІм\}$  ( $>$  *килерім*) ‘я приду’.

**2.0.8.** Показатели ITER  $Ады(p)$ ,  $иды(p)$ , PRES<sub>2</sub>  $чАды(p)$  в литературном языке не опускают  $p$ . В качинском и кызыльском диалектах, где эти формы употребляются в устной речи,  $p$  часто опускается. Аналогично случаям с показателями DUR<sub>1</sub> и FUT, 1SG.full сочетается только с полной формой аффикса:  $\{чурта-Ады(p)-ПЫн\}$  ‘жить-ITER-1SG’  $\rightarrow \{чурта-Адыр-ПЫн\}$  ( $>$  *чуртидырбын* ‘живу’),  $\{ніл-Ады(p)-(БІ)м\}$  ‘знать-ITER-1SG’  $\rightarrow \{ніл-Ады-(БІ)м\}$  ( $>$  *ніледім* ‘знаю’),  $\{одыр-чАды(p)-(БІ)м\}$  ‘сидеть-PRES<sub>2</sub>-1SG’  $\rightarrow \{одыр-чАды-(БІ)м\}$  ( $>$  *одырчадым* ‘сизу’). В прочих случаях распределение свободное, точнее, диалектное.

**2.0.9.** Аффиксы посессивного склонения выбираются после посессивных аффиксов множественного числа.

**2.0.10.** Гласная в скобках, стоящая в начале морфа, обозначает, что если предыдущий морф кончается на согласную, будет выбран вариант данного морфа, начинающийся на эту гласную морфону; а если предыдущий морф кончается на гласную, будет выбран вариант данного морфа, начинающийся на согласную морфону (т. е. гласная в скобках опускается).

Согласная в скобках, стоящая в начале морфа, обозначает, что если предыдущий морф кончается на гласную, будет выбран вариант данного морфа, начинающийся на эту согласную морфону; а если предыдущий морф кончается на согласную, будет выбран вариант данного морфа, начинающийся на следующую после согласной в скобках морфону (т. е. согласная в скобках опускается).

- $-(V)\{Mrphn+1\} \rightarrow -V\{Mrphn+1\}/\{Mrphn\}C\_ \{am\}$  ‘имя’ +  $\{(БІ)м\}$  1pos.SG  $\rightarrow \{am-БІм\}$  ( $>$  *адым*)
- $-(V)\{Mrphn+1\} \rightarrow -0\{Mrphn+1\}/\{Mrphn\}V\_ \{ада\}$  ‘отец’ +  $\{(БІ)м\}$  1pos.SG  $\rightarrow \{ада-м\}$  ( $>$  *адам*)
- $-(C)\{Mrphn+1\} \rightarrow -C\{Mrphn+1\}/\{Mrphn\}V\_ \{ада\}$  ‘отец’ +  $\{(з)БІ\}$  3pos  $\rightarrow \{ада-зБІ\}$  ( $>$  *адазы*)
- $-(C)\{Mrphn+1\} \rightarrow -0\{Mrphn+1\}/\{Mrphn\}C\_ \{am\}$  ‘имя’ +  $\{(з)БІ\}$  3pos  $\rightarrow \{am-БІ\}$  ( $>$  *ады*)

## 2.1. Правила морфонологических чередований, применение которых требует информации из словарей морфем: из грамматического словаря основ (например, сведения о лексико-грамматическом классе основы или словарные данные этой основы) и из таблицы 1 словоизменяемых аффиксов (например, граммема данной морфемы)

**2.1.1.** (Обязательно действующее). Преобразование конечных (стоящих непосредственно перед морфологической словоизменяемой границей)  $n$ ,  $z$ ,  $г$ ,  $н$  односложных глагольных основ на границе с аффиксами CONV<sub>n</sub> и NF на  $(БІ)n$  в  $Г$ :  $\{тап\}$  ‘находить’ +  $\{БІn\}$   $\rightarrow \{таГ-БІn\}$  ( $>$  *таан*) ‘найдя’,  $\{теп\}$  ‘толкать’ +  $\{БІn\}$   $\rightarrow \{теГ-БІn\}$  ( $>$  *теен*) ‘толкая’,  $\{суз\}$  +  $\{БІn\}$   $\rightarrow \{суГ-БІn\}$  ( $>$  *суин*) ‘чертя’,  $\{чыг\}$  +  $\{БІn\}$   $\rightarrow \{чыГ-БІn\}$  ( $>$  *чыын*) ‘собирая’,  $\{чиң\}$  +  $\{БІn\}$   $\rightarrow \{чиГ-БІn\}$  ( $>$  *чиин*) ‘побеждая’,  $\{тоң\}$  +  $\{БІn\}$   $\rightarrow \{тоГ-БІn\}$  ( $>$  *тоон*) ‘замерзая’. Информация о наличии и обязательности преобразования извлекается из

характеристики первой морфемы как основы, второй морфемы как аффикса CONV<sub>n</sub> или NF и их сегментного состава.

**2.1.2.** (Факультативное). Преобразование конечных (стоящих непосредственно перед морфологической словоизменительной границей) *з, э, ң* в основах и аффиксах на границе со следующими аффиксами, начинающимися с гласной, в Г<sup>13</sup>. Правило действует факультативно, т. е. преобразование может происходить, а может и не происходить.

Примеры.

#### Многосложные основы имен

По имеющимся в настоящее время данным *з, э* в этом случае в литературном языке выпадают<sup>14</sup> обязательно, но для части диалектов это не так; в нормативном словаре зафиксированы разные варианты, иногда форма принадлежности в грамматической зоне словаря БХРС выписана одним образом, а в примере попадает форма принадлежности от того же имени, образованная иначе; в грамматике этот вопрос подробно не рассматривается. В набор диалектологических анкет для хакасско-шорского ареала мы включили полный список слов с такими финалами.

Примеры на *э* и *з*: {улуз} ‘большой’ + {БЛ} Зрос → {улуГ-БЛ} (> улии) ‘старший’, {марыз} ‘пари’ + {БЛ} Зрос → {марыГ-БЛ} (> мариин) ‘его пари’, {абылаз} ‘нападение’ + {БЛ} Зрос → {абылаГ-БЛ} (> абылаа) ‘его нападение’, {чөгиз} ‘зной’ + {БЛ} Зрос → {чөгизГ-БЛ} (> чөгизи: күн чөгизи ‘дневной зной’), {чөлег} ‘опора’ + {БЛМ} 1рос.SG → {чөлеГ-БЛМ} (> чөлеем) ‘моя опора’, {сүрдез} ‘тревога’ + {БЛ} Зрос → {сүрдеГ-БЛ} (> сүрдее) ‘его тревога’;

НО: {ойыз} ‘пятно’ + {БЛ} Зрос → {ойызГ-БЛ} (> ойизи ~ ойызы) ‘его пятно’, {орыз} ‘место обитания’ + {БЛ} Зрос → {орызГ-БЛ} (> орызы: пулан орызы ‘место обитания лосей’; турна орызы ‘гнездовье журавлей’); {игез} имя действия от игирге ‘пилка’ + {БЛ} Зрос → {игизГ-БЛ} (> игее ~ игеги: одың игеги ‘пилка дров’); {көлиг} ‘запряжка; упряжка’ + {БЛ} Зрос → {көлиГ-БЛ} (> көлизи ~ көлиги); {төзег} ‘поле (фон чего-л.)’ + {БЛ} Зрос → {төзегГ-БЛ} (> төзее ~ төзеге: сатиннің төзее көк ‘поле у сатина синее’; төзеге хызыл товар ‘ткань с красным фоном’);

Примеры на *ң* (все зафиксированные случаи): {пулуң} ‘угол’ + {БЛ} Зрос → {пулуГ-БЛ} (> пулиин) ‘его угол’; {чилиң} ‘мозг’ + {БЛ} Зрос → {чилиГ-БЛ} (> чилиин) ‘его мозг’; {хордың} ‘солод, закваска’ + {БЛ} Зрос → {хордыГ-БЛ} (> хордлин): ас хордиин ‘ячменный солод’; айран хордиин ‘закваска для айрана’; НО {үрдүң} ‘наковальня’ + {БЛ} Зрос → {үрдүңГ-БЛ}: сапхы үрдүңи ‘наковальня для отбивания литовки’; {чүчүң} ‘древесный сок’ + {БЛ} Зрос → {чүчүңГ-БЛ}: хазың чүчүңи тадылыг ‘берёзовый сок сладкий’; ср. кызыльское диалектное {утхуң} ‘верхний слой сети’ + {БЛ} Зрос → {утхуңГ-БЛ}: сөзирбениң утхуңи ‘верхняя часть невода’.

Хабитуальное причастие НАВ {ЧАң} (при субстантивации) — правило обязательно: {сыда} ‘выносить’ + {ЧАң} НАВ + {БЛ} Зрос → {сыда-ЧАГ-БЛ} (> сыдачаа: сыдачаа чох ‘невыносимый’); {чөр} ‘ходить’ + {ЧАң} НАВ + {БЛ} Зрос → {чөр-ЧАГ-БЛ} (> чөрчее: Тасхар чөрчее чох соох ‘Такой мороз, что на лицу не выйти’).

#### Односложные основы имен

Правило факультативно, есть частичные распределения по лексемам и дублеты как в диалектных системах, так и в литературном языке.

Примеры на *э* и *з*: {суз} ‘вода’ + {БЛ} Зрос → {суГ-БЛ} (> суу: саңыс суу ‘сок калины’), {нуз} ‘каприз, прихоть’ + {БЛ} Зрос → {нуГ-БЛ} (> нуу) ‘его прихоть’, {таз} ‘гора’ + {БЛ} Зрос → {таГ-БЛ} (> таа) ‘его гора’<sup>15</sup>, {саз} ‘сила’ + {БЛ} Зрос → {саГ-БЛ} (> саа) ‘его сила’, {наз} ‘веревка’ + {БЛ} Зрос → {наГ-БЛ}

<sup>13</sup> Формально такое и предыдущее преобразование, представляющие собой трансформации на морфонологическом уровне, обеспечивают приведение к морфеме, которая на фонологическом уровне имеет реализацию Ø; эта реализация обеспечивается фонологическим правилом, см. ниже. Некоторые соображения по единообразию соответствующих процессов в разных тюркских языках (например, реализация этой морфемы в виде фарингального призвука в тубаларском и, возможно, в некоторых хакасских диалектах) не позволяют нам уже на этом этапе говорить о выпадении согласного.

<sup>14</sup> Выпадение является следствием данного преобразования по следующему ниже правилу 2.3.2.

<sup>15</sup> Так в литературной грамматике, и так же выглядит пример в словаре. Однако ср. в корпусе: *Хара мрамор таста пазыл парган: «Трошкиннер родының тагы»* ‘На чёрном мраморном камне написано: «Гора рода Трошкиных»’. [ЭКХЯ: Воспоминания Н.Ф. Трошкина]. Н. Ф. Трошкин по происхождению — носитель качинского диалекта. Аналогичные формы дают и носители шорского диалекта.

(> *naa*) ‘его веревка’, {*kög*} ‘песня’ + {*Bl*} 3pos → {*köG-Bl*} (> *köö*) ‘его песня’, {*ciɣ*} ‘черта’ + {*Bl*} 3pos → {*ciɣ-Bl*} (> *ciui*) ‘его черта’;

НО {*хан-ниг*} ‘правитель’ + {*Bl*} 3pos → {*хан-ниг-Bl*} (> *хан-ниги*) ‘его правитель’, {*aɣ*} ‘сеть’ + {*Bl*} 3pos → {*aɣ-Bl*} (> *aɣы: палых ағы* ‘рыболовная сеть’), {*soɣ*} ‘жар’ + {*Bl*} 3pos → *soɣ-Bl* (> *soɣы: көс соғы* ‘жар углей’), {*чүг*} ‘перо’ + {*Bl*} 3pos → {*чүг-Bl*} (> *чүги*) ‘его перо’.

Примеры на *н*: {*öң*} ‘цвет’ + {*Bl*} 3pos → {*öG-Bl*} (> *öö: суғ өө чох* ‘вода бесцветна’), {*ман*} ‘досуг’ + {*Bl*} 3pos → {*маG-Bl*} (> *маа*) ‘его досуг’; {*соң*} ‘конец’ + {*Bl*} 3pos → {*соG/ң-Bl*} (> *соо ~ соңы: ай соо* ‘конец месяца’; *нымах соо* ‘конец сказки’).

НО: {*аң*} ‘зверь’ + {*Bl*} 3pos → {*аң-Bl*} (> *аңы*) ‘его зверь’.

Ср. также наличие в словаре таких двух входов (обе леммы ни разу не встретились в нашем корпусе, и, в частности, мы не можем в данном случае судить о правильности выделения в словаре-источнике начальной формы), как {*нүг*} ‘пух’ + {*Bl*} 3pos → {*нүг-Bl*} (> *нүги: хус нүги* ‘птичий пух’) и {*нүң*} ‘пух’ + {*Bl*} 3pos → {*нүң-Bl*} (> *нүңи: хас нүңи* ‘гусиный пух’).

#### Однословные основы глаголов (многословных основ глаголов с конечным *з, э, ң* не бывает)

При сочетании с показателем FUT *Ap*, (от которого, в частности, образуется инфинитив — из-за этого у нас много примеров, т. к. инфинитив служит словарным входом) правило факультативно в литературном языке. В диалектах встречаются разные варианты действия (в вопросник диалектологической анкеты мы включили полный список таких основ), но, кажется, преобладают регулярные случаи — либо на всех основах действует, либо на всех не действует.

{*ciɣ*} ‘чертить’ + {*Ap*} FUT → {*ciɣ-Ap*} (> *ciup*) ‘начертит’, {*чыг*} ‘собирать’ + {*Ap*} FUT → {*чыG-Ap*} (> *чыыр*) ‘соберет’, {*тоң*} ‘мерзнуть’ + {*Ap*} FUT → {*тоG-Ap*} (> лит. *тоор*)

или:

{*ciɣ*} + {*Ap*} FUT → {*ciɣ-Ap*} (> *ciɣer*) ‘начертит’, {*чыг*} ‘собирать’ + {*Ap*} FUT → {*чыг-Ap*} (> *чыгар*) ‘соберет’, {*тоң*} ‘мерзнуть’ + {*Ap*} FUT → {*тог-Ap*} (> диал. шор. *тогар*) ‘будет мерзнуть’.

Перед прочими показателями, начинающимися с гласной, — CONV<sub>a</sub>, ITER и императивными — это преобразование не происходит: {*чыг*} ‘собирать’ + {*A*} CONV<sub>a</sub> → {*чыг-А*} (> *чыга*) ‘собирая’, {*чыг*} ‘собирать’ + {*АдЫр*} ITER → {*чыг-АдЫр*} (> *чыгадыр*) ‘обычно собирает’. Для PROSP *AK* нет данных, но более вероятно, что преобразования (и последующего стяжения) нет. Перед показателем DAT посессивного склонения *-А* конечная согласная показателя 2pos.SG *-ң* не переводится в *Г* ({*ат*} ‘лошадь’ + {*Blң*} 2pos.SG + {*A*} DAT → {*ат-Blң-А*} (> *адыңа*) ‘твоей лошади’).

К настоящему времени преобразование 2.1.2. отмечено для некоторых слов в словаре в поле FORM, но опыт работы с корпусом демонстрирует, что оно относится к ненормированным в литературном языке диалектным по происхождению явлениям, так что следует считать его не лексически связанным, а морфонологически факультативным, возможным в принципе для любой основы такой структуры.

**2.1.3.** Преобразование конечных (стоящих непосредственно перед морфологической словоизменительной границей<sup>16</sup>) согласных *-к, -х, -К* неодносложной последовательности морфем на границе с аффиксом, начинающимся на гласную, в *Г*: {*харах*} ‘глаз’ + {*Blм*} 1pos.SG → {*хараG-Blм*} (> *хараам*) ‘мой глаз’, {*хузух*} ‘орех’ + {*Blм*} 1pos.SG → {*хузуг-Blм*} (> *хузим*) ‘мой орех’; {*изик*} ‘дверь’ + {*Blм*} 1pos.SG → {*изиG-Blм*} (> *изиим*) ‘моя дверь’, {*пöрик*} ‘шапка’ + {*Blм*} 1pos.SG → {*пöриG-Blм*} (> *пöриим*) ‘моя шапка’. Ср. указание на дублетность в словаре-источнике: {*көгенек*} ‘рубашка’ + {*Bl*} 3pos > *көгенее, көгенегі* ‘его рубашка’ и нек. др. Правило действует факультативно, по тому же принципу, что правило 2.1.2.

На данный момент нам встретились только случаи действия правила на границе с аффиксом принадлежности. С глагольными основами аналогичной структуры, типа *хорых-* ‘бояться’, в формах причастий FUT *Ap*, императива *им, BlңAp, АңAp*, деепричастий CONV<sub>n</sub> (*Bl*)*n* и т. п. такого не происходит. Таким образом, информация о наличии чередования извлекается из сегментного состава морфем, стоящих перед морфологической границей, сегментного состава и имени аффикса принадлежности.

#### **2.1.4.** «Беглые» гласные.

В д в у с л о ж н ы х нечленимых основах узкие краткие гласные (*y, j, ы, i*) закрытого конечного слова, находящиеся между согл. *р-н, л-н, й-н*), выпадают при присоединении аффиксов принадлежности, начинающихся на гласную. См. [Баскаков 1953: 373; ГХЯ 1975: 22]. Так сформулировано правило в нормативных грамматиках. Однако, при обращении к материалам БХРС и корпуса оказывается, что од-

<sup>16</sup> В условной морфонологической расчлененной записи граница между словообразовательными и словоизменительными показателями различается (у нас это «⇒» для словообразовательных и «→» для словоизменительных). Формулируемое правило учитывает только «→».

на часть основ не демонстрирует отклонений от него, другая часть основ такой структуры не показывают редукции, а еще одна часть дает дубликаты с редукцией и без редукции.

**Основы с выпадением:** {алын} ‘перед’ + {БІ} Зрос → {алн-БІ} (> алны ‘передняя часть чего-л.’ — встречено в грамматиках; примерах в БХРС; 663 примера в литературном корпусе); {ирин} ‘губа’ + {БІ} Зрос → {ирн-БІ} (> ирні ‘его губа’; встречено в грамматиках; примерах в БХРС; 2 случая в корпусе); {килин} ‘невестка’ + {БІ} Зрос → {килн-БІ} (> килні ‘ее невестка’; грамматика; примеры в БХРС; 15 случаев в корпусе); {мойын} ‘шея’ + {БІ} Зрос → {мойн-БІ} (> мойны ‘его шея’; грамматика; примеры в БХРС; 47 случаев в корпусе); {пурун} ‘нос’ + {БІ} Зрос → {пурн-БІ} (> пурны ‘его нос’; грамматика; пример в БХРС; 13 случаев в корпусе); {сүрүн} ‘дух, душа; жизненная сила’ + {БІ} Зрос → {сүрн-БІ} (> сүрні ‘его душа’; 3 примера в БХРС); {харын} ‘живот’ + {БІ} Зрос → {харн-БІ} (> харны ‘его живот’ — грамматика; БХРС; 25 случаев в корпусе); {хойын} ‘пазуха; объятия; ложбина на склоне’ + {БІ} Зрос → {хойн-БІ} (> хойны ‘его пазуха’; грамматика; 5 примеров в БХРС; 9 случаев в корпусе) {чилін} ‘грива’ + {БІ} Зрос → {чилн-БІ} (> чилні ‘его грива’; 3 примера в БХРС); {чилін} ‘вымя’ + {БІ} Зрос → {чилн-БІ} (> чилні ‘ее вымя’; пример в БХРС).

**Основы без выпадения:** {ойын} ‘игра’ + {БІ} Зрос → {ойын-БІ} (> ойыны (грамматика; 5 примеров в БХРС; 1 пример в корпусе: ойынында ЗРос-ЛОС ‘сделались) позорищем’ 1 Кор 4:9); {сайын} ‘скорлупа’ + {БІ} Зрос → {сайын-БІ} (> хузух сайыны ‘скорлупа кедрового орешка’ (пример в БХРС); {сарын} ‘песня’ + {БІ} Зрос → {сарын-БІ} (> сарыны (12 примеров в корпусе); {сүрүн} ‘облик, лицо’ (этимологически тождественно {сүрүн} ‘душа’) + {БІ} Зрос → {сүрүн-БІ} (> сүрүні (1 пример в БХРС); {хулун} ‘жеребёнок-сосунок’ + {БІ} Зрос → {хулун-БІ} (> хулуны: тибе хулуны ‘верблюжонок’ БХРС; 11 примеров в корпусе); {хурун} ‘копоть, сажа в дымоходе’ + {БІ} Зрос → {хурун-БІ} (> хуруны 1 пример в БХРС); {чалын} ‘пламя’ + {БІ} Зрос → {чалын-БІ} (> чалыны (1 пример в БХРС, 3 примера в корпусе).

**Дублетные основы:** {орын} ‘место’ + {БІ} Зрос > → {орн-БІ} (орны (10 примеров в БХРС, 166 примеров в корпусе; НО также 1 пример *Илханның чир түбі орыны* ‘Подземное царство Илхана’ (ЭКХЯ: Топоев Илья Прокопьевич. Хара торааттың Хароол 1.890); {чалын} ‘подгривный жир, холка’ + {БІ} Зрос > → {чалн-БІ} (> чалны (2 примера в БХРС), → {чалын-БІ} (> чалыны (1 пример в БХРС); {чарын} ‘лопатка’ + {БІ} Зрос → {чарн-БІ} (> чарны (ГХЯ 22; 3 примера в БХРС; 5 примеров в корпусе) и → {чарын-БІ} (> хой чарыны ‘баранья лопатка’, 1 пример в БХРС).

Нет данных о выпадении/невыпадении для слов *пырын* ‘сажа’, *сойын* ‘чугун’, *турун* ‘головня’, *чайын* ‘разлив’, *чүлүн* ‘физиономия человека’.

Относительно происхождения и бытования этого и сходных явлений в тюркских языках ср. [Дубо 2005]. Как можно видеть, в хакасском распределение по классам относительно выпадения/невыпадения узкого гласного не связано прямо с этимологической структурой основы (старые конечные кластеры с *j*, например, в хакасском упростились — см. [Дыбо 2007: 44—46]; а слова с реконструируемым нестянувшимся сочетанием \**V-jin* (‘шея’, ‘пазуха’, ‘игра’) попали в разные классы). Нынешняя картина может, как нам кажется, указывать на постепенную утрату чередования, сохраняющегося у наиболее частотных слов. В парсере вид основы с выпавшей гласной представлен в словарной статье словаря в поле ALTERNATEN.

**2.1.5.** Выбор варианта основы у слов, оканчивающихся на *-ст*, *-ц*. В основах, оканчивающихся на допустимое сочетание согласных *ст* и не имеющих пометы о заимствовании (таких основ в словаре три: *ист-* ‘слушать’, *сист-* ‘развязывать’, *үст-* ‘верх’), вариант со *ст-* выступает перед аффиксом на гласную, вариант с *с-* — в абсолютном конце или при следующем аффиксе, начинающемся на согласную. Формы глагола *ист-* ‘слушать’: *ис!* ‘слушай!’, *иснес* < {*ис-ПА.с*} ‘слушать-NEG.FUT’ ‘непослушный’, *иссең* < {*ис-СА-ң*} ‘слушать-COND-2SG.br’ ‘если ты слышишь’, *истерге* < {*ист-Ар-ГА*} ‘слушать-FUT-DAT’ ‘слушать’; формы глагола *сист-* ‘развязывать’: *сисне* < {*сис-ПА*} ‘развязывать-NEG’ ‘не развязывай’, *систин* < {*сист-Ын*} ‘развязывать-CONV<sub>п</sub>’ ‘развязав’, *үст-* ‘верх’: *үстү* < {*үст*} + {БІ} Зрос (это слово в хакасском языке всегда употребляется с аффиксом принадлежности). В словарной базе в поле ALTERNAT в качестве словоизменительной основы выписаны варианты *ис-* и *сис-*.

У именных основ, имеющих варианты, оканчивающиеся на *-ст*, *-ц* (только заимствования), вариант на *-с* (представлен в словарной статье словаря в поле ALTERNATEN) выбирается при следующем аффиксе, начинающемся на гласную, вариант на *-ст* — в абсолютном конце или при следующем аффиксе, начинающемся на согласную. Примеры: *артист* — {*артис*} + {БІ} Зрос (> *артизи*) ‘его артист’, {*артист*} + {ЛАр} PL (> *артисттер*) ‘артисты’, *полководец* — {*полководес*} + {БІ} Зрос (> *полководези*) ‘его полководец’, {*полководец*} + {ДАң} ABL (> *полководецтең*) ‘от полководца’.

**2.1.6.** COND *CA* и RPAST *ТЫ*, непосредственно после которых идет личный аффикс кратко набора 2PL *ңАр*, претерпевают следующие преобразования: конечный гласный первого аффикса (*A* или *Ы* соответственно) и начальный *ң* второго аффикса выпадают. Примеры: {*нас-CA-ңАр*} ‘писать-COND-2PL.br’ → {*нас-С.Ар*} (> *нассар*) ‘если будем писать’, {*нас-ПА-CA-ңАр*} ‘писать-NEG-COND-2PL.br’ → {*нас-ПА-С.Ар*} (> *наспаза*) ‘если не будем писать’, {*нас-ТЫ-ңАр*} ‘писать-RPAST-2PL.br’ → {*нас-Т.Ар*} (> *насттар*) ‘вы недавно написали’. Правило действует факультативно, в литературном языке допускаются как стяженный, так и полный (без выпадений) варианты данных форм: {*нас-CA-ңАр*} > *нассаңар*, {*нас-ПА-CA-ңАр*} > *наспазаңар*, {*нас-ТЫ-ңАр*} > *настыңар*.

## 2.2. Морфонологические правила, применение которых определяется сегментным составом сочетающихся морфем, включающим информацию о морфемных границах

### 2.2.1. Преобразование согласных после словоизменяющей морфемной границы (см. таблицу 2).

Морфы (аффиксы), начинающиеся с определенной морфемы, преобразуют эту морфему в другую морфему в зависимости от качества конечной морфемы предшествующего морфа (инвентарь словоизменяющих морфем в морфонологической условной записи см. в таблице 1; условная запись прочих морфем извлекается из полей лексической базы ALTERNAT, ALTERNATEN, или DERIV; либо, если эти поля в словарной статье не заполнены, то совпадает с начальной формой слова).

Морфонема *П* выбирает репрезентацию *б* после звонкой носовой согласной, *п* после гласной или глухой согласной, *м* после носовой согласной.

{*нар*} ‘идти’ + {*ПА*} NEG → {*нар-бА*} (> *нарба*) ‘не иди’  
 {*сана*} ‘считать’ + {*ПА*} NEG → {*сана-пА*} (> *санаба*) ‘не считай’  
 {*нас*} ‘писать’ + {*ПА*} NEG → {*нас-пА*} (> *наспа*) ‘не пиши’  
 {*сом*} ‘купаться’ + {*ПА*} NEG → {*сом-мА*} (> *сомма*) ‘не купайся’

Морфонема *Т* выбирает репрезентацию *д* после звонкой согласной, *т* после гласной или глухой согласной.

{*им*} ‘лекарство’ + {*ТА*} LOC → {*им-дА*} (> *имде*) ‘в лекарстве’  
 {*халас*} ‘хлеб’ + {*ТА*} LOC → {*халас-тА*} (> *халаста*) ‘в хлебе’  
 {*ада-ТА*} → {*ада-тА*} (> *адада*) ‘у отца’

Морфонема *Д* выбирает репрезентацию *д* после звонкой носовой согласной, *н* после носовой согласной, *т* после гласной или глухой согласной.

{*аал*} ‘деревня’ + {*ДАң*} AVL → {*аал-дАң*} (> *аалдаң*) ‘из деревни’  
 {*кем*} ‘кто’ + {*ДАң*} AVL → {*кем-нАң*} (> *кемнең*) ‘от кого’  
 {*агас*} ‘дерево’ + {*ДАң*} AVL → {*агас-тАң*} (> *агастаң*) ‘из дерева’  
 {*тура*} ‘дом’ + {*ДАң*} AVL → {*тура-тАң*} (> *турадаң*) ‘из дома’

Морфонема *С* выбирает репрезентацию *з* после звонкой согласной, *с* после глухой согласной или гласной.

{*сом*} ‘купаться’ + {*СА*} COND → {*сом-зА*} (> *сомза*) ‘если купается’  
 {*кис*} ‘надевать’ + {*СА*} COND → {*кис-сА*} (> *киссе*) ‘если надевает’  
 {*сана*} ‘считать’ + {*СА*} COND → {*сана-сА*} (> *саназа*) ‘если считает’

Морфонема *Л* выбирает репрезентацию *л* после гласной или звонкой носовой согласной, *т* после глухой, *н* после носовой.

{*күске*} ‘мышь’ + {*ЛАр*} PL → {*күске-лар*} (> *күскелер*) ‘мыши’  
 {*адай*} ‘собака’ + {*ЛАр*} PL → {*адай-лар*} (> *адайлар*) ‘собаки’  
 {*хас*} ‘гусь’ + {*ЛАр*} PL → {*хас-тар*} (> *хастар*) ‘гуси’  
 {*хозан*} ‘заяц’ + {*ЛАр*} PL → {*хозан-нар*} (> *хозаннар*) ‘зайцы’

Морфонема *Н* выбирает репрезентацию *н* после звонкой согласной или гласной, *т* после глухой согласной.

{*суз*} ‘вода’ + {*НЫ*} → {*суз-ны*} (> *сузны*) ‘воду’  
 {*хоосха*} ‘кошка’ + {*НЫ*} → {*хоосха-ны*} (> *хоосханы*) ‘кошку’  
 {*сеек*} ‘муха’ + {*НЫ*} → {*сеек-ты*} (> *сеекти*) ‘муху’

Морфонема *Ч* выбирает репрезентацию *ч* после звонкой согласной или гласной, *ч* после глухой согласной.

{*чол*} ‘дорога’ + {*ЧА*} PROL → {*чол-ЧА*} (> *чолча*) ‘по дороге’  
 {*сом*} ‘купаться’ + {*ЧАң*} НАВ → {*сом-ЧАң*} (> *сомчаң*) ‘он обычно купался’  
 {*тура*} ‘дом’ + {*ЧА*} PROL → {*тура-ЧА*} (> *турача*) ‘по дому’  
 {*пас*} ‘писать’ + {*ЧАң*} НАВ → {*пас-ЧАң*} (> *пасчаң*) ‘он обычно писал’

Морфонема *L* выбирает репрезентацию *л* после глухой или звонкой неносовой согласной, *н* после носовой<sup>17</sup>.

{*пас*} ‘писать’ + {0} NF + {*ЛА*} CONT + {*ЧА*} PRES → {*пас-ла-ЧА*} (> *паслача*) ‘постоянно пишет’  
 {*пар*} ‘идти’ + {0} NF + {*ЛА*} CONT + {*ЧА*} PRES → {*пар-ла-ЧА*} (> *парлача*) ‘все идет и идет’  
 {*тын*} ‘дышать’ + {0} NF + {*ЛА*} CONT + {*ЧА*} PRES → {*тын-на-ЧА*} (> *тыннача*) ‘все время дышит’  
 {*сана*} ‘считать’ + {*н*} NF + {*ЛА*} CONT + {*ЧА*} PRES → {*сана-н-ла-ЧА*} (> *саналача*) ‘все время считает’

Морфонема *K* выбирает репрезентацию *К* после гласной и глухой согласной, *Г* после звонкой, в том числе носовой согласной<sup>18</sup>.

{*ырах=КЫ*} ‘далеко=АТТР’ → {*ырах=КЫ*} (> *ыраххы*) ‘дальний’  
 {*ам=КЫ*} ‘сейчас=АТТР’ → {*ам=ГЫ*} (> *амгы*) ‘теперешний’  
 {*иур=КЫ*} ‘вечер=АТТР’ → {*иур=ГЫ*} (> *иурги*) ‘вечерний’  
 {*иб-ТА-КЫ*} ‘дом-ЛОС-АТТР’ → {*иб-да-КЫ*} (> *ибдеги*) ‘тот, что в доме’

Морфонема *Г* выбирает репрезентацию *К* после глухой согласной, *Г* в прочих случаях.

{*хыс*} ‘девушка’ + {*ГА*} → {*хыс-КА*} (> *хысха*) ‘девушке’  
 {*пала*} ‘ребенок’ + {*ГА*} → {*пала-ГА*} (> *палаа*) ‘ребенку’  
 {*адай*} ‘собака’ + {*ГА*} → {*адай-ГА*} (> *адайга*) ‘собаке’  
 {*кем*} ‘кто’ + {*ГА*} → {*кем-ГА*} (> *кемге*) ‘кому’

Исключение: после показателя FUT *A* морфонема *Г* трансформируется в *К* (в диалектах, см. выше): {*пар-А-ГА*} → {*пар-А-КА*} (> *парага*) ‘идти’. (Это, разумеется, связано с историческим наличием на конце этого аффикса согласной {*p*}. Ср. также диалектное *поган* < *полган* {*пол-ГАН*} ‘был’ и др. аналогичные основы на *p*, *л*: здесь также наличие глубинного сонанта препятствует обычному развитию в поверхностно интервокальной позиции).

<sup>17</sup> В словоизменении эта морфонема представлена только в аффиксе континуатива и не встречается после гласных (по правилу оформления NF 2.0.2). В словообразовании, по-видимому, можно считать содержащим эту морфону аффикс глаголообразования от звукоподражаний (в отличие от стандартного аффикса глаголообразования от имен ЛА), ср.: *майтлат-* ‘шлепать по воде’ {*майт=ЛА=(Ы)Т-*} ‘звукоподражание звуку шлепания=OPER=CAUS’; *мыхлас-* ‘пищать вместе’ {*мых=ЛА=(Ы)С-*} ‘звукоподражание писку=OPER=REC-’ (так после глухих на всем объеме БХРС); *марлас* ‘рёва’ {*мар=ЛА=ГАС-*} ‘звукоподражание реву=OPER=ADJ’; *соолат-* ‘вызывать рокот’ {*соо=ЛА=(Ы)Т-*} ‘звукоподражание рокоту=OPER=CAUS-’.

<sup>18</sup> Внутри основы (в словообразовательных аффиксах) после сонорных морфонема *К* может принимать и глухой вариант: *андархы* ‘тамошний’, *үстүңзархы* ‘южный/верхний’. Проблема нуждается в дальнейшем изучении. По БХРС: А) глухой вариант выступает в подавляющем большинстве случаев после *p*, единично после *n* [16 отыменных производных: *аархы* ‘дальний’, *алтынзархы* ‘северный’, *андархы* ‘находящийся в тех краях’, *іскеркі* ‘восточный’, *кидеркі* ‘западный’, *мындархы* ‘местный’, *озархы* (тот+сторона=АТТР) ‘находящийся на той стороне’, *оңархы* (правый=DIR=АТТР) ‘правый’, *өөркі* (вверх=АТТР) ‘верхний’, *пееркі* (сюда=АТТР) ‘ближайший’, *тасхархы* (наружная сторона=DIR=АТТР) ‘наружный’, *тогырхы* (поперек=АТТР) ‘поперечный’, *төрзеркі* (почетное место в юрте=ALL=АТТР) ‘передний (о месте в комнате)’, *үстүңзархы* (верх+сторона=АТТР) ‘южный/верхний’, *хайдархы* (какой=LOC=DIR=АТТР) ‘находящийся в какой-л. стороне’, *чогархы* (вверх=АТТР) ‘верхний’; *төбінкі* ‘нижний’; 2 отглагольных: *ибірки* (кружить=АТТР) ‘близлежащий’, *индірки* (спускаться под гору=CAUS=АТТР) ‘нижний’]; Б) звонкий вариант выступает редко после *p* [4 случая, из них 1 отглагольный: *иурги* ‘вечерний’, *ирги* (старый=АТТР) ‘старый (о вещах)’, *пылтыргы* ‘прошлогодный’, *нүдіргі* (возникать=CAUS=АТТР) ‘творительный (падеж)’, обычно после других сонорных [14 случаев: *амгы* ‘теперешний’, *иртенгі* ‘утренний’, *орынгы* ‘местный падеж’, *пурунгы* (раньше=АТТР) ‘прежний’, *нүүлгі* (< *пу=ЧЫл=КЫ* этот+год=АТТР) ‘нынешний’, *нүүңгі* (< *пу=күн=КЫ* этот+день=АТТР) ‘сегодняшний’, *сагамгы* (время+сейчас=АТТР) ‘теперешний’, *хачангы* (когда=АТТР) ‘когдашний’, *чагынгы* ‘ближний’; отглагольные: *адалгы* (называть=Pass=АТТР) ‘основной (падеж)’, *көрімгі* (смотреть=Quant=АТТР) ‘винительный (падеж)’, *пирілгі* (давать=Pass=АТТР) ‘дательный (падеж)’, *тартылгы* (тянуть=Pass=АТТР) ‘притяжательный’, *ызылгы* (отправлять=Pass=АТТР) ‘направительный (падеж)’. Ср. также звонкий после *й* в названии времени года: *чайгы* ‘лето’ (в литературном хакасском оно, как и остальные, утратило производящую основу: *хысхы* ‘зима’, *часхы* ‘весна’, *күскү* ‘осень’).



Таблица 2. Реализации согласных морфем после морфемной границы

Конечная морфонема предшествующего морфа ↓	Начальная морфонема → последующего морфа									
	П	Т	Д	С	Л	Л	Н	Ч	К	Г
C <sub>v</sub>	б	д	д	з	л	л	н	ч	Г	Г
C <sub>n</sub>	м	д	н	з	н	н	н	ч	Г	Г
C <sub>u</sub>	п	т	т	с	т	л	т	ч	К	К
V	п	т	т	с	л	л	н	ч	К	Г

### 2.2.2. Сингармонизм.

Правило преобразования вокалической морфемы аффикса в зависимости от предшествующей охарактеризованной по ряду гласной в последовательности морфем. Вокалические морфемы аффикса выбирают реализацию того ряда, который характеризует ближайшую из предшествующих аффиксу гласных, охарактеризованную по ряду. Напомним, что гласная *и* (в литературной норме) переднерядная, если это первая гласная в словоформе (гласная первого слога), и неохарактеризована по ряду в прочих случаях.

{*кис-чАт-Ар-ГА*} ‘резать-DUR-FUT-DAT’ → {*кис-чет-ер-Ге*} (> *кисчедерге*) ‘резать некоторое время’

{*нар-ПА-ЧАң-ПЫс*} ‘идти-NEG-HAB-1PL’ → {*нар-ба-чаң-мыс*} (> *нарбачаңмыс*) ‘мы обычно не уходили’

{*кис-Пин-чАт-ТЫр*} ‘резать-NF.NEG-DUR-INDIR’ → {*кис-пин-чет-тір*} (> *киспинчеттір*) ‘он, оказывается, не режет’

{*нар-Пин-ча-СЫң*} ‘идти-NF.NEG-PRES-2SG’ → {*нар-бин-ча-сың*} (> *нарбинчазың*) ‘ты не уходишь’

NB: Для неадаптированных русских заимствований (они имеют в словаре помету *rus* в поле ЕГУМ) выбор огласовки первого после основы аффикса происходит в соответствии с последней гласной основы; если это *и*, то выбирается переднерядный вариант: *гастрит* (-ды) ‘гастрит’ — *гастриттең ирееленерге* ‘страдать гастритом’. Похожее правило (переднерядные аффиксы после *и* в любом слоге предшествующей последовательности морфов) действует в сагайском и бельтырском диалектах для всех слов: *париген* < {*нар-и-ГАН*} ‘он шел’ в м. лит. *париган*. Только это примечание о поведении заимствований не позволяет переформулировать для литературного языка правило о рядном сингармонизме на фонологическом уровне, как учитывающее счет слогов и не учитывающее морфемных границ (как в таблице 3).

Морфонема *А* выбирает репрезентацию *a* после заднерядной гласной; *e* после переднерядной гласной.

Морфонема *Ы* выбирает репрезентацию *ы* после заднерядной гласной; *i* после переднерядной гласной.

Морфонема *О* выбирает репрезентацию *o* после заднерядной гласной; *ö* после переднерядной гласной.

Гласные *у*, *ү* не встречаются в аффиксах в литературном языке.

Таблица 3. Реализации гласных морфем

V <sub>n-1</sub>	V <sub>n</sub>		
	А	Ы	О
V <sub>back</sub>	<i>a</i>	<i>ы</i>	<i>o</i>
V <sub>front</sub>	<i>e</i>	<i>i</i>	<i>ö</i>

**2.3. Фонологические правила**, включающие исключительно фонологические условия (не зависящие от морфологического членения словоформы). Это фактически правила перехода с морфемного на фонемный уровень записи, при котором осуществляется пересчет «морфем» предыдущего уровня в «фонемы». Для парсера, работающего с текстом в орфографической форме, это пересчет с записи, содержащей условные буквы, в просто буквенную запись.

#### 2.3.0. Все границы морфем ликвидируются.

**2.3.1.** Если с обеих сторон от морфемной границы стояли гласные (уже без скобок — скобки были убраны на этапах 2.0.1—2.0.5), то эти гласные стягиваются по следующим правилам: первая гласная всегда поглощается второй, а вторая становится долгой. NB: Как было отмечено выше, гласные первого слога *и*, *о*, *ö* автоматически долгие. Соответственно, если вторая гласная сочетания — *и* или *о*, *ö*, то первая гласная устраняется и остается только *и* или *о*, *ö*:

*ырла* ‘петь’ + *им* IMP.1SG > *ырлим* ‘пусть я спою’; *сине* ‘мерять’ + *им* IMP.1SG > *синим* ‘пусть я измерю’; *ырла* ‘петь’ + *-ибыс* IMP.1PL > *ырлибыс* ‘давайте споём’; *сине-* ‘мерять’ + *ибыс* IMP.1PL > *синибыс* ‘давайте измерим’; *ойна* ‘играть’ + *ир* FUT > *ойнир* ‘будет играть’, *узу* ‘спать’ + *ир* FUT > *узир* ‘он будет спать’, *сана* ‘считать’ + *ибас* CONV<sub>нас</sub> > *санибас* ‘считаю’, *сөле* ‘говорить’ + *ибес* CONV<sub>нас</sub> > *сөлибес* ‘говоря’, *сана* ‘считать’ + *и* CONV<sub>а</sub> > *сани* ‘считаю’, *төзе* ‘расстилать’ + *и* CONV<sub>а</sub> > *төзи* ‘расстилая’, *сана* ‘считать’ + *идыр* ITER > *санидыр* ‘считает’, *төзе* ‘расстилать’ + *идир* ITER > *төзидир* ‘расстиляет’, *узу* ‘спать’ + *иК* PROSP+ *ча* PRES → *узиКча* (> *узихча*) ‘вот-вот заснет’; *чи* ‘есть’ + *им* IMP.1SG > *чим/чим* ‘поем-ка я’; *тура* ‘дом’ + *оК* ASS → *туроК* (> *турох*) ‘дом же’; *анда* ‘там’ + *оК* ASS → *андоК* (> *андох*) ‘там же’<sup>19</sup>.

Если вторая гласная сочетания — *А* (здесь уже реализованная как *а* или *е*), то образуется долгая *aa/ee*: *узу* ‘спать’ + *аң* IMP.1.INCL > *узааң* ‘давай спать’, *сине* ‘мерять’ + *еңер/елер* IMP.1.INCL.PL > *синееңер/синеелер* ‘давайте измерять’. Для первого слога так же: *чи* ‘есть’ + *ең* IMP.1.INCL > *чеең* ‘давай поедим’.

Все исключения касаются сочетаний с частицей *ОК* (что, разумеется, связано с тем, что это клитика, формы с ней не до конца синтезированы):

1) если первая гласная сочетания *ы* или *і*, вторая *о* или *ө* — в таком сочетании обе гласные могут сохраняться. Подавляющее большинство таких случаев — сочетание ассертивной частицы *ОК* с показателем *Зрос* (в корпусе больше 20 примеров, все без усечения), но не только:

*чахсы* ‘хороший’ + *оК* ASS → *чахсыоК* (> *чахсыох*): *чахсаа чахсыох урунча* ‘будешь сам добрым — и к тебе будут относиться с добром’ (букв. ‘хорошему хорошее встречается’; пословица); *аны* ‘он.ACC’ + *ох* ASS → *аныоК* (> *аныох*) ‘и его/её’.

*пирсі* ‘один из множества’ + *өК* ASS → *пирсіөК* (> *пирсіөк*): *Аның оң холы, пирсіөк осхас, хазых пол парган*. ‘И правая рука его, как другая, стала здорова.’ (Лк. 6:10)

*хыс* ‘дочь’ + *ы* *Зрос* + *оК* ASS → *хызыоК* (> *хызыох*) ‘его/ее дочь же’.

(В корпусе правило об усечении первой гласной вообще часто нарушается, обнаруживая и другие частичные морфологические распределения. Так, среди более чем 90 форм локатива на *-ТА* с частицей *ОК* нет ни одного случая без усечения, а среди более чем 200 форм датива на *-ГА* с частицей *ОК* нет ни одного случая с усечением; с пролативом *ЧА* в корпусе встретилось 3 формы, все с усечением. Вероятно, тут вмешиваются диалектные системы правил; ср. бельтырский говор, где может усекается, наоборот, гласная частицы: *чаа-га* ‘война-ДАТ’ + *ОК* > *чаагах* ‘на войну же’);

2) если первая гласная сочетания является гласной первого слога словоформы, а вторая — *о* или *ө*, то первая гласная никогда не устраняется, остается сочетание гласных: *пу* ‘этот’ + *оК* ASS → *пуоК* (> *пуох*) ‘этот же’;

3) если первая гласная сочетания является долгой, она никогда не устраняется, остается сочетание гласных: *магаа* ‘я.ДАТ’ + *оК* ASS → *магааоК* (> *магааох*) ‘мне же’.

**2.3.2.** Все *Г*, имеющиеся в условной записи словоформы до сих пор и оказавшиеся в интервокальной позиции, выпадают, а окружающие их гласные стягиваются в одну долгую: в какую именно, зависит от позиции слога и состава гласных, см. таблицу 4.

Исключение: если предшествующая или следующая за *Г* гласная долгая (т. е. двойная или *и*, *о*, *ө* первого слога исконного слова), выпадения и стяжения не происходит:

*салаа* ‘палец’ + *Га* DAT > *салаага* ‘пальцу’, *көзө* ‘сучок’ + *Ге* DAT > *көзөгө* ‘сучку’, *пызо* + *Га* DAT ‘теленки’ > *пызога* ‘теленку’, *пар* ‘идти’ + *и* DUR + *Ган* PAST > *париган* ‘уходил’.

*улуГ*<sup>20</sup> ‘большой’ + *оК* ASS → *улугоК* (> *улугох*) ‘тоже большой’, *чөлеГ* ‘опора’ + *өК* ASS → *чөлегөК* (> *чөлег-өк*) ‘тоже опора’, *саГ* ‘дойти’ + *им* IMP.1SG > *сагим* ‘пусть я подою’.

Но: *государство* + *Га* DAT > *государстваа/?государствога* ‘государству’.

<sup>19</sup> Здесь описано поведение частицы *-ОК* так, как следует из изложения и примеров в [ГХЯ 1975: 249]. Так, по-видимому, устроена разговорная норма литературного хакасского языка (информанты оценивают как правильные формы: *кiзi* + *ОК* > *кiзöк* ‘человек же’, *түлгү* + *өк* > *түлгөк* ‘лиса же’, *школа-да-гы* ‘школа-ЛОС-АТТР’ + *ОК* > *школадагах* ‘школьный же’, *паба-зы* ‘отец-Зрос’ + *ох* > *пабазох* ‘его отец же’, *пар-ды* ‘уйти-RPAST’ + *ОК* > *пардох* ‘он ушел же’, *пар-гла* ‘уйти-DISTR’ + *ОК* > *парглох* ‘расходитесь же!’, *чол-ча* ‘путь-PROL’ + *ОК* > *чолчох* ‘по дороге же’. По-видимому, такое правило действует в качинском диалекте, ср. примеры из с. Трошкино (экспедиция 2016 г.) *тогын-а-га* ‘работать-FUT-DAT’ + *ОК* > *тогынагах* ‘работать же’, *чаа-га* ‘война-ДАТ’ + *ОК* > *чаагах* ‘на войну же’. Реально в парсере принято факультативное опущение не-двойной гласной перед частицей *-ОК*, что соответствует ситуации в корпусе, см. ниже.

<sup>20</sup> Напомним, конечное *г* основы было преобразовано в *Г* по правилу 2.1.2, а в случае с *саГ-* — по правилу 2.1.1.

Таблица 4. Правила стяжения гласных при выпадении интервокальной согласной.

	<i>А</i>	<i>Ы</i>	<i>У</i>	<i>О</i>	<i>Е</i>	<i>И</i>	<i>І</i>	<i>Ӗ</i>	<i>ӱ</i>
1 слог + <i>А</i>	<i>аГа</i> > <i>аа</i> <i>саг+ар</i> > <i>саар</i> 'будет до- ить'	<i>ыГа</i> > <i>ыы</i> <i>чыг+ар</i> > <i>чыыр</i> 'будет собирать'	<i>уГа</i> > <i>уу</i> <i>нуг+ар</i> > <i>нуур</i> 'будет жмурить'	<i>оГа</i> > <i>оо</i> <i>сог+ар</i> > <i>соор</i> 'будет остывать'	<i>еГе</i> > <i>ее</i> <i>эг+ер</i> > <i>ээр</i> 'будет гнуть'	<i>иГе</i> > <i>ии</i> <i>сиг+ер</i> > <i>сиир</i> 'будет чертить'	<i>іГе</i> > <i>ее</i> <i>ти+ген</i> > <i>теен</i> 'он сказал'	<i>ӓГе</i> <i>сӓг+ер</i> > <i>сӓӓр</i> 'будет вы- тягивать'	<i>ӱГе</i> > <i>ӱӱ</i> <i>ӱг+ер</i> > <i>ӱӱр</i> 'будет окучивать'
1 слог + <i>Ы</i>	<i>аГы</i> > <i>аа</i> <i>саг+ып</i> > <i>саап</i> 'доя'	<i>ыГы</i> > <i>ыы</i> <i>чыг+ып</i> > <i>чыып</i> 'собирая'	<i>уГы</i> > <i>уу</i> <i>суг+ы</i> > <i>суу</i> 'его вода'	<i>оГы</i> > <i>оо</i> <sup>21</sup> <i>пог+ып</i> > <i>пооп</i> 'душив'	<i>еГі</i> > <i>ее</i> <i>тег+ип</i> > <i>тееп</i> 'касясь'	<i>иГі</i> > <i>ии</i> <i>сиг+ип</i> > <i>сиип</i> 'чертя'	нет примеров	<i>ӓГі</i> > <i>ӓӓ</i> <i>кӓг+і</i> > <i>кӓӓ</i> 'мелодия'	<i>ӱгі</i> > <i>ӱӱ</i> <i>ӱг+ип</i> > <i>ӱӱп</i> 'окучивая'
Непервый слог + <i>А</i>	<i>аГа</i> > <i>аа</i> <i>тура+га</i> > <i>тураа</i> 'дому' <i>сана + ган</i> 'РАСТ' > <i>санаан</i> 'он считал'	<i>ыГа</i> > <i>аа</i> <i>казы+га</i> > <i>казаа</i> 'полю' <i>таны+ган</i> > <i>танаан</i> 'он узнал'	<i>уГа</i> > <i>аа</i> <i>уйгу+га</i> > <i>уйгаа</i> 'сну'; <i>узу + гай</i> 'ОРГ' > <i>узаай</i> 'пусть спит'	не стяги- вается из-за долготы	<i>еГе</i> > <i>ее</i> <i>киме+ге</i> > <i>кимее</i> 'лодке'	не стяги- вается из-за долготы	<i>іГе</i> > <i>ее</i> <i>ізі + гелек</i> 'CUNC' > <i>ізеелек</i> 'еще не согрелся'	не стяги- вается из-за долготы	<i>ӱГе</i> > <i>ее</i> <i>тӱлӱг+ге</i> > <i>тӱлӱгее</i> 'лисе'
Непервый слог + <i>Ы</i>	<i>аГы</i> > <i>аа</i> <i>чайлаг+ы</i> > <i>чайлаа</i> 'его летняя стоянка'	<i>ыГы</i> > <i>ии</i> <i>чылыг+ы</i> > <i>чылии</i> 'тепло'	<i>уГы</i> > <i>ии</i> <i>улуг+ы</i> > <i>улии</i> '(его) старший'	не стяги- вается из-за долготы	<i>еГі</i> > <i>ее</i> <i>чӓлег+і</i> > <i>чӓлее</i> 'его опора'	не стяги- вается из-за долготы	<i>іГі</i> > <i>ии</i> <i>сіліг+і</i> > <i>сілии</i> '(ее) кра- сота'	не стяги- вается из-за долготы	<i>ӱГі</i> > <i>ии</i> <i>чӱзӱг+і</i> > <i>чӱзӱии</i> 'его чехол для топора'

**2.3.3.** После правила о выпадении в интервокальной позиции действует правило озвончения в интервокальной позиции для согласных *n, m, c, k, x, K*:  $VnV \rightarrow VbV$ ,  $VmV \rightarrow VdV$ ,  $VcV \rightarrow VzV$ ,  $VkV \rightarrow VzV$ ,  $VxV \rightarrow VzV$ ,  $VKV \rightarrow VTV$ .

Глухой вариант, соответственно, остается перед следующей согласной или концом слова.

$n \rightarrow n$ : {*сан*}#  $\rightarrow$  *сан* 'рукоятка', {*сана*} + {*n*} CONV<sub>n</sub> + {*чА*} PRES  $\rightarrow$  {*сана-н-ча*}  $\rightarrow$  *санапча* 'он считает'

$n \rightarrow b$ : {*сан*} + {*Ы*} Зрос  $\rightarrow$  {*саб-ы*} > *сабы* 'его рукоятка', {*сана*} + {*n*} CONV<sub>n</sub> + {*ОК*} ASS + {*чА*} PRES {*сана-п-оК-ча*}  $\rightarrow$  *санабоКча* (> *санабохча*) 'он тоже считает',

$c \rightarrow c$ : {*тус*}#  $\rightarrow$  *тус* 'время',

$c \rightarrow z$ : {*тус*} + {*Ы*} Зрос  $\rightarrow$  {*тус-ы*} > *тузы* 'его время', {*сана*} + {*Пас*} CONV<sub>пас</sub> + {*ОК*} ASS > {*сана-пас-оК*}  $\rightarrow$  *санабазоК* (> *санабазох*) 'он тоже не посчитает';

$m \rightarrow m$ : {*тут*}#  $\rightarrow$  *тут* 'держи', {*кис*} 'резать' + {*чАТ*} DUR + {*ГАН*} PAST  $\rightarrow$  {*кис-чет-Кен*} > *кис-четКен* ( $\rightarrow$  *кисчеткен*) 'он порой резал',

$m \rightarrow d$ : {*тут*} + {*Ар*} FUT  $\rightarrow$  {*тут-ар*} > *тудар* 'будет держать', {*кис*} 'резать' + {*чАТ*} DUR + {*Ар*} FUT + {*ГА*} DAT  $\rightarrow$  {*кис-чет-ер-Ге*} > *кисчедерГе* ( $\rightarrow$  *кисчедерге*) 'резать время от времени', {*тура*} 'дом' + {*ДАң*} AVL  $\rightarrow$  {*тура-таң*} > *турадаң* 'из дома'.

Исключение: в нескольких корневых морфемах, исключительно, если *m* следует за первой гласной от начала словоформы, оно может сохраняться в любой позиции ({*ат-Ар-ГА*} > *атарга* /ат-/ I 'стрелять'; II 'рассветать'; {*пат-Ар-ГА*} > *патарга* /пат-/ 'вязнуть'; {*хат-Ар-ГА*} > *хатарга* /хат-/ I 'засыхать'; II 'свивать'; {*ут-Ар-ГА*} > *утарга* /ут-/ 'выигрывать'; {*хат-Ы*} > *хаты* 'его баба/жена'). Эта словарная информация отмечена в поле ALTERNAT словарной базы<sup>22</sup>.

<sup>21</sup> Единственное односложное существительное на *-ог*, как будто, не демонстрирует стяжения: *кӓс согы* 'жар углей'.

<sup>22</sup> Возможным решением было бы ввести морфемему «неозвончающееся *m*» для обозначения сегмента на конце этих корней, но, как уже было сказано, мы стараемся не усложнять запись основ в лексическом словаре. Статус этого «неозвончающегося *m*» к настоящему моменту совершенно неясен, по диалектам наблюдается варьирование форм с озвончением и без озвончения, так что в принципе здесь возможно заимствование в литературный язык из «неозвончающих» диалектов. Но вероятно и историческое решение: почти все эти корни (кроме *хат-* 'свивать', ко-

$x \rightarrow x$ : {сых} 'выходить' + {ГАН} PAST > сыхГан 'выходил', {хах}# 'дупло' > хах;  
 $x \rightarrow z$ : {сых} 'выходить' + {АР} FUT > сыгар 'будет выходить', {хах} 'дупло' + {Ы} Зрос > хагы (тыт хагы 'дупло лиственницы'), {чух} 'смола' + {Ы} Зрос > чугы (харагай чугы 'сосновая смола')  
 $k \rightarrow k$ : {тик} 'шить' + {ГАН} PAST > тикГен 'шил'; {чүк}# 'котомка' > чүк;  
 $k \rightarrow z$ : {тик} 'шить' + {АР} FUT > тигер 'он будет шить'; {чүк} 'котомка' + {Ы} Зрос > чүгі 'его котомка', {түк} 'овечья шерсть' + {Ы} Зрос > түгі (хой түгі 'овечья шерсть');  
 $K \rightarrow K$ : {нар} 'идти' + {АК} PROSP + {ЧА} PRES > нараКча 'вот-вот уйдет';  
 $K \rightarrow G$ : {иб-де-Ки} 'дом-LOC-ATTR' > ибдеГі 'тот, что в доме'.

В многосложных основах в интервокальной позиции  $k$ ,  $x$ ,  $K$  могли, перейдя по правилу 2.1.2. в  $G$ , выпасть затем по правилу 2.3.2. — тогда окружавшие их гласные стянулись по общим правилам: харага 'глаз' + ым 1рос.SG > хараам 'мой глаз', хузуГ 'орех' + ым 1рос.SG > хузиим 'мой орех'; изиГ 'дверь' + им 1рос.SG > изиим 'моя дверь', нөриГ 'шапка' + им 1рос.SG > нөриим 'моя шапка'. Формально это те случаи, когда в словаре в поле ALTERNATEN выписана альтернативная запись основы, оканчивающаяся на  $-z$  или  $-z$ , а в поле FORM — словоформа на  $-ии$ . Но правило 2.1.2. действует не обязательно, и если оно не подействовало и заднеязычный сохранился, то теперь он озвончается: харах 'глаз' + ым 1рос.SG > харагым 'мой глаз', хузух 'орех' + ым 1рос.SG > хузугым 'мой орех'; изик 'дверь' + им 1рос.SG > изигим 'моя дверь', нөрик 'шапка' + им 1рос.SG > нөригим 'моя шапка'.

В неадаптированных заимствованиях правило о выпадении не действует: автокружок, автокружогы; крановщик, крановщиגי.

Правило об озвончении в интервокале факультативно действует также на последние согласные  $ш > ж$  и  $ф > в$  в заимствованиях и диалектных словах: параграф — парграфы, шериф — шериві, карандаш — карандажы, чүгеш (кыз.) 'котомка' — чүгежі. Но: тариф — тарифы.

#### 2.3.4. Выпадение $G$ после звонких заднеязычных.

При стечении двух согласных, первой  $z$ ,  $z$ ,  $ң$ ,  $G$ , и второй  $G$ , вторая согласная выпадает, первая остается: суз 'вода' + Га DAT > суга 'воде', эг 'гнуть' + Гей OPT > эгей 'пусть гнет', тоң 'мерзнуть' + Ган PAST > тоңан 'мерз', чохса= $G$  (дремать= $S_{act}$  'дремота') + Га DAT > чохсага (чихсага түс парган 'погрузился в дремоту'). В шорском диалекте это правило обеспечивает реализацию формы 2рос.SG-DAT адыңа 'твоей лошади' (ат 'лошадь' + ың 2рос.SG + Га DAT)<sup>23</sup>.

2.3.5. Все сохранившиеся на этом этапе условные буквы (в морфонологической транскрипции применены как условные кириллические буквы верхнего регистра) преобразуются в соответствующие буквы нижнего регистра, таким образом возвращая орфографическую форму хакасской словоформы (до обработки графическими правилами).

Уточнение для заднеязычных: увулярные согласные  $x$ ,  $z$  в лексике, не имеющей в словаре пометы *rus*, употребляются только с гласными заднего ряда, а велярные  $k$ ,  $z$  — только с переднерядными: харах-тар-ыбыс-та 'в наших глазах', кирек-тен-деңер 'по делам'. Соответственно, условная буква (морфонема)  $G$  при переднерядном окружении выбирает реализацию  $z$ , при заднерядном —  $z$ ; условная буква (морфонема)  $K$  при переднерядном окружении выбирает реализацию  $k$ , при заднерядном —  $x$ .

2.4. «Графические сандхи», обусловленные способом употребления кириллической графики (запись сочетаний звуков вида  $й+V$  с помощью кириллических йотированных букв), производится факультативно. В соответствии с современной орфографией БХРС  $я$  пишется в первом слоге,  $йа$  — в непервом:

торый колеблется по озвончению/неозвончению в словообразовании, и самодийского заимствования 'баба') восходят к праалтайским корням с придыхательным  $*t'$  (см. EDAL), ср. {аТ-} < ПА  $*p'ət'ə$  'ударять', {наТ-} < ПА  $*pət'ə$  'попадать', {хаТ-} < ПА  $*k'ət'ə$  'твердый', {уТ-} < ПА  $*ūt'a$  'встречать', но {тум-} < ПА  $*t' [u]t'V$  'хватать'. Вопрос может проясниться после дополнительных корпусных и диалектологических исследований. В таких словах, как *чиміг* 'острый', *хатыг* 'твердый', мы имеем дело с орфографическим приемом, не выписывающим двойной согласный на месте сильного глухого (см. [ГХЯ 1975: 39]) в случае отсутствия синхронной морфологической границы (обе формы исторически являются дериватами с аффиксом  $-ЛЫГ$ , утерявшими производящую основу) и выписывающим в случае наличия такой границы: *сүттиг* 'молочный' < *сүт* 'молоко' +  $ЛЫГ$ . В *ікі* 'два', *чимі* 'семь', *отыс* 'тридцать' основа исторически содержала геминату в интервокальной позиции.

<sup>23</sup> Таким образом, анализ поверхностных словоформ должен действовать в этом случае так: если видишь в тексте  $ГА$  после гласного, то, если там есть граница морфемы, это либо  $VG-ГА$ , либо  $VK-A$ . Если видишь  $ңА$  после гласного, это либо  $Vң-ГА$ , либо  $Vң-A$ .

я ‘да’, *чахайах* ‘цветок’. С исконным сочетанием *й+у* в словаре есть только 3 основы, не имеющих общего правила: *туюх* ‘глухой (о местности)’, *хуюх* ‘гарь’, *уйуң-айың* ‘шум-гам’. Ср. также в ГХЯ: *сыгыбысхан*, *чуйу́х-чайы́х полыт* ‘он вышел, шатаясь из стороны в сторону’ (В. Кобяков). Для текстов, изданных в более ранней орфографии, характерно использование *я* в любых слогах. Некоторые авторы предпочитают такое написание и теперь.

**2.4.0.** *йу > ю, йа > я, йе > е* ([ГХЯ 1975: 9—10]).

## Заключение

Можно видеть, что сегментные процессы, обуславливающие синхронный облик хакасской словоформы, представляют довольно сложную, многоуровневую систему, которая может быть проинтерпретирована как конгломерат разновременных фонетических процессов, постепенно вытеснявшихся на более глубокие уровни языка, а также вымывавшихся аналогическими процессами. В том числе, процессы озвончения и выпадения интервокальных согласных<sup>24</sup> явно происходили за историю хакасского языка не один раз, и результаты их неравномерно сглаживаются процессами парадигматической унификации по разным диалектам и говорам. Автоматический морфологический анализатор может обойти часть этих явлений с помощью «лексических исключений», поместив соответствующую информацию в словарные статьи лексической базы, но значительное их число все-таки должно обрабатываться алгоритмом. Отметим, что существующие словари и грамматики хакасского языка содержат значительные лакуны в описании сегментного поведения ряда морфем и типов морфем, которые могут быть надежно закрыты только с помощью корпусного исследования и последующей направленной этим исследованием элицитации (в том числе в отношении границ диалектного варьирования). Часть этих лагун закрывается настоящей статьёй.

## Приложение

### Необходимые пояснения к таблице

По горизонтали в таблице расположен перечень позиций в словоформе, занимаемых аффиксами, выражающими граммы — значения, которые принимают грамматические категории<sup>25</sup>. Следует отметить, что позиции не обязательно соответствуют конкретным грамматическим категориям (это обычное явление для непарадигматических по сути агглютинативных языков, см., например, [Martin 1992: 244—246] о модели глагольной словоформы в корейском языке), ср., например, три разных позиции для показателей характера протекания действия — перфектива, дуратива и итератива; или две разных позиции для показателей типов эпистемической модальности — ассумптива и ирреалиса; или одна позиция (8) для аффиксов времени и наклонения (т. е. грамматической категории, выражающей модальность). Взаимное расположение позиций обусловлено лишь порядком, в котором могут следовать в словоформе рассматриваемые аффиксы. По вертикали в таблице расположены аффиксы, выражающие граммы (показатели значений грамматических категорий), записанные в морфонологической форме (т. е. в глубокой форме, из которой, с помощью применения правил реализации морфоном может быть выведено поверхностное выражение их в орфографически правильно записанной словоформе).

Каждая позиция может быть занята только одним из возможных для нее аффиксов, а может быть не занята вообще, и, следовательно, соответствующая грамма или категория в целом (в том случае, если позиция соответствует определенной категории) могут быть не выражены. Например, в словоформе может быть не выражено число (такая словоформа традиционно характеризуется как форма единственного числа), падеж (такая именная словоформа традиционно характеризуется как форма именительного/неопределенного падежа), лицо (такая финитная словоформа традиционно характеризуется как форма 3-его лица).

<sup>24</sup> Проблеме, в частности, посвящена специальная статья Ц. Номинханова [Номинханов 1955]. Основываясь на неполном материале и с частично неточным подходом, ученый пришел примерно к тому же выводу, что и мы.

<sup>25</sup> О соответствии данного грамматического описания традиционному описанию хакасской грамматики см. [Шеймович 2017].

Таблица 1. Модель изменяемой хакасской словоформы

№№ п/п	0	1	2	3	4	5	6	7	8	9	10	11
	S	DISTR	CONV <sub>1</sub>	PTCL <sub>1</sub>	PERF, PROSP	DUR	NEG/ ITER	Tense, Mood, CONV <sub>2</sub>	INDIR	TRANSP	AFFIRM	PL <sub>1</sub>
1.		КлА	NF (Ы)п, 0	ADD ТАА	PERF (Ы)б.Ыс	DUR чАт (?ЧАт)	NEG ПА	<b>причастия</b>	Тыр	COMIT ЛЫГ	ЧЫК	ЛАр
2.			NF.NE G Пин	CONT ЛА	[PERF <sub>0</sub> Ыс]	DUR <sub>1</sub> у(р)	ITER А.дЫ(р), и.дЫ(р)	PAST ГА(н)		ADV Ли		
3.				ASS ОК		PROSP (диал.) АК, иК		FUT А(р), и(р)				
								NEG.FUT ПА.с				
4.								НАВ ЧА(н)				
5.								ASSUM ГАдАГ				
								SIMUL АчЫК				
6.								CUNC ГАЛАК				
7.								<b>финитные</b>				
								PRES чА				
8.								PRES <sub>2</sub> чАдЫ(р) (?ЧАдЫ(р))				
9.								COND СА				
10.								RPAST ТЫ				
11.								OPT ГАй				
12.								<b>дееприча- стия</b>				
13.								CONV <sub>n</sub> (Ы)п				
14.								NEG.CONV Пин				
15.								NEG.Conv. ABL Пин.Аң				
16.								CONV <sub>a</sub> А, и NF (Ы)п, 0				
17.								CONV <sub>нас</sub> А.бАс, и.бАс (диал.)				
18.								LIM ГАли				

и набор хакасских словоизменительных аффиксов (см. подробнее [ЭКХЯ])

12	13		14	15	16	17		18	19	20	21
Poss <sub>1</sub>	Case <sub>1</sub>		ATTR	PL <sub>2</sub>	Poss <sub>2</sub>	Case <sub>2</sub>		PTCL <sub>2</sub>	Person	PredPL	PTCL <sub>3</sub>
	Simple declension	Possessive declension				Simple declension	Possessive declension				
1pos.SG (Б)М	GEN НЫҢ	GEN нЫҢ	КЫ	ЛАР	1pos.SG (Б)М	GEN НЫҢ	GEN нЫҢ	ASS ОК	<b>полные</b>	ЛАР	ASS ОК
					1.SG ПЫН						
2pos.SG (Б)Ң	LOC ТА	LOC (н)ДА			2pos.SG (Б)Ң	DAT ГА	DAT (н)А	CONT ЛА	1.SG СЫМ (диал.)		FOC ТЫР
									2.SG СЫҢ		
3pos (з)Ы	ALL САР	ALL (н)ЗАР			3pos (з)Ы	ACC НЫ	ACC н(Ы)		1.PL ПЫС		PREC ТАК
									1.PL СЫБЫС (диал.) 2.PL САР		
1pos.PL (Б)БЫС					1pos.PL (Б)БЫС	LOC ТА	LOC (н)ДА		2.PL СЫҢ.АР (диал.)		
2pos.PL (Б)Ң.АР					2pos.PL (Б)Ң.АР	ABL ДАҢ, ТЫН	ABL НАҢ		<b>смешан- ные</b> (mix)		
									1.SG.mix М		
					GEN.3pos НИ	ALL САР	ALL (н)ЗАР		2.SG.mix СЫҢ		
						INSTR НАҢ	INSTR НАҢ		1.PL.mix ПЫС		
									2.PL.mix САР		
						PROL ЧА	PROL (н)ЧА		<b>краткие</b> (br)		
						DELIB ДАҢАР	DELIB НАҢАР		1.SG.br М		
									2.SG.br Ң		
						COMP ТАГ	COMP (н)ДАГ		1.PL.br ПЫС		
									2.PL.br ҢАР		
									<b>импера- тивные</b>		
									IMP.1.SG ИМ		
									IMP.1.PL И.БЫС		
									IMP.1.INCL АҢ		
									IMP.1.PL.INCL АҢ.АР		
									IMP.1.INCL.PL АЛАР (диал.)		
									IMP.2.PL (Б)ҢАР		
									IMP.3 СЫН		

## Условные обозначения в таблице

1, 2, 3 — лица

1, 2, 3 — нумерация одинаковых показателей в разных слотах, а также слоты со сходными наборами показателей.

Счет ведется от начала словоформы. Например,  $PTCL_3$  обозначает третий (с начала) слот для частиц, а  $PL_2$  — показатель множественного числа, занимающий вторую из возможных (с начала) позицию в словоформе.

ABL — аблатив (исходный)

ACC — аккузатив (винительный)

ADD — аддитивная частица (и, же, ведь...)

ADV — адвербиальный показатель

AFFIRM — аффирматив, утвердительная форма

ALL — аллатив (направительный)

ASS — ассертивная (утвердительная) частица (также, тоже...)

ASSUM — ассумптив, предположительное наклонение, вводится оборотом «похоже, что...»

ATTR — атрибутивный показатель<sup>26</sup>

br — краткий (тип личных окончаний)

Case — падеж

COMIT — комитатив, показатель совместности («с...», «вместе с...»)

COMP — компаратив (сравнительный)

COND — условное наклонение

CONT — континуативная частица (все пишет и пишет...)

CONV.NEG — отрицательная форма деепричастия

CONV<sub>a</sub> — деепричастие на -а («параллельное»)

CONV<sub>n</sub> — деепричастие на -ып («последовательное»)

CONV<sub>нас</sub> — качинская форма «последовательного» деепричастия

CUNC — кунктатив, еще не совершившееся действие

DAT — датив

DELIB — делибератив (выражает объект речи или мысли, также причинно-следственные отношения)

DISTR — дистрибутив, обозначает множественность субъекта или объекта действия;

DUR, DUR<sub>1</sub> — дуратив (длительность действия)

FOC — фокусная частица

Full — полный (тип личных окончаний)

FUT — будущее время

GEN — генетив (родительный)

HAB — хабитуалис (привычное действие в настоящем и прошедшем времени)

IMP — императив (повелительное наклонение)

INCL — инклюзив

INDIR — индиректив, косвенная эвиденциальность (неочевидность либо заглазность) действия

INSTR — инструментальный (творительный)

ITER — итератив (периодичность действия)

LIM — деепричастие предела в прошлом

LOC — локатив (местный)

Mix — смешанный (тип личных окончаний)

Mood — наклонение

NEG — отрицание

NEG.CONV — отрицательная форма деепричастия

NEG.CONV.ABL — деепричастие мгновенного следования (аблатив отрицательного деепричастия)

NEG.FUT — отрицательная форма будущего времени

NF — (not finished | non finitum) грамматикализованный показатель, исторически восходящий к CONV<sub>n</sub>, выступающий в ряде синтетических и аналитических словоформ. (Т. к. производные формы не составляют группу, имеющую общий признак, считаем, что значение данного аффикса — словоформа еще не закончилась.)

<sup>26</sup> ATTR и ADV обычно функционируют как словообразовательные показатели, присоединяясь к чистой основе; тогда соответствующее слово как правило есть в словаре, и аффиксы учитываются только при словообразовательном анализе: *ырах=хы* (далеко=ATTR) 'далекий', *сiлiг=ли* (красивый=ADV) 'красиво'. Но они могут также присоединяться к концу словоформы, содержащей словоизменительные показатели — тогда парсер подвергает эти аффиксы анализу: *чол-лар-да-гы* (дорога-PL-LOC-ATTR) *оймах-осхылар* 'ухабы на дорогах', *чирге нас от-кен-ни* (пролезать-PAST-ADV) 'в место, куда голова бы пролезла'.



[NOM — номинатив, или нулевой падеж, отсутствует в таблице, т. к. не имеет поверхностного выражения]

Num (SG, PL) — число имени (единственное, множественное)

OPT — оптатив (желательное наклонение)

PAST — прошедшее время

PERF — перфектив (завершенность действия)

Person — лицо предиката

Possessive declension — набор падежных аффиксов притяжательного склонения (после показателя принадлежности)

Poss (pos) — принадлежность (лицо и число посессора)

PREC — прекативная частица (частица с просительным оттенком, смягчающая приказание)

PredPL — множественное число предиката

PRES — настоящее время

PRES<sub>2</sub> — качинская форма настоящего времени

PROL — пролатив (продольный)

PROSP — проспектив (состояние, предшествующее действию), исключительно сагайский диалект

PTCL — частица (пишущаяся слитно со словоформой, а также вставная).

RPAST — недавно прошедшее время

S — основа. Основа включает корень (R) со словообразовательными показателями, она присутствует в словаре в качестве заголовка словарной статьи. Регулярное наличие показателя в заглавных словарных формах служило критерием невключения того или иного показателя (например, аффиксов деятеля) в разряд словоизменяемых

Simple declension — набор падежных аффиксов простого склонения

SIMUL — симулятив («как если бы...»)

Tense — время

TRANSP — транспонирующий аффикс: показатель, меняющий синтаксическую функцию слова (напр., с субстантивной на адъективную, как *ЛЫГ*)

Кумулятивно выраженные грамемы разделяются точками.

#### Прочие сокращения

ADJ — атрибутивное производное имя

C — согласная

CAUS — каузатив

C<sub>n</sub> — носовая согласная

C<sub>u</sub> — глухая согласная

C<sub>v</sub> — звонкая согласная

DIR — директив

Mrphn — морфонема

OPER — переходный отыменной глагол

REC — реципрок

S<sub>act</sub> — имя действия

V — гласная

V<sub>back</sub> — заднерядная гласная

V<sub>front</sub> — переднерядная гласная

вм. — вместо

диал. — диалектный

кыз. — кызыльский

лит. — литературный

ПА — праалтайский

шор. — шорский

#### Литература

Алиса в стране чудес 2017 — Алисаның хайхастар чиринзер чорыгы (Alice's Adventures in Wonderland in Khakas). Хакас тіліне Мария Чертыкова тібестеен. Portlaoise, 2017.

Алпатов 2018 — *Алпатов В. М.* Слово и части речи. М., 2018. {*Alpatov V. M.* Word and Parts of Speech. М., 2018.}

Баскаков 1953 — Хакасско-русский словарь / Под ред. *Баскакова Н. А.* М., 1953. {*Khakass-Russian Dictionary / Ed. Baskakov N. A.* М., 1953.}

БХРС 2006 — Хакасско-русский словарь / Под ред. *О. В. Субраковой.* Новосибирск. {*Khakass-Russian Dictionary/Ed. O. V. Subrakova.* Novosibirsk.}

ГХЯ 1975 — Грамматика хакасского языка / Под ред. *Н. А. Баскакова.* М., 1975. {*Khakass Grammar./Ed. N. A. Baskakov.* М., 1975.}

ДМТЯ — Диалектологические материалы по тюркским языкам <https://dialects.altai.ru/>, 2019. {*Dialectological Materials on Turkic Languages https://dialects.altai.ru/*, 2019.}

Дыбо 2007 — *Дыбо А. В.* Лингвистические контакты ранних тюрков. Лексический фонд. Пратюркский период. М., 2007. {*Dybo A. V.* Language CONTACTS of early Turks. Lexical Stock. Proto-Turkic Era. М., 2007.}

Дыбо, Мудрак 2009 — *Дыбо А. В., Мудрак О. А.* Об историческом принципе при составлении фонетической анкеты для полевого обследования диалекта // III Международная конференция полевой лингвистике. Тезисы и

материалы. М., 2009. {*Dybo A. V., Mudrak O. A. On Using the Historical Principle while Compiling a Phonetic Survey for a Field Study of a Dialect // III International Conference on Field Linguistics. Abstracts and Materials. M., 2009.*}

Дыбо, Шеймович 2014 — *Дыбо А. В., Шеймович А. В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков // «Научное обозрение Саяно-Алтая». Вып. 1, 2014. С. 9—30. {Dybo A. V., Sheymovich A. V. Automatic Morphological Analysis for Khakass and Old Turkic Corpora // Sayan-Altai Scientific Review. Vol. 1, 2014. P. 9—30.}*

Дыренкова 1948 — *Дыренкова Н. П. Грамматика хакасского языка. Фонетика и морфология. Абакан, 1948. {Dyrenkova N. P. Khakass Grammar. Phonetics and Morphology. Abakan, 1948.}*

Зализняк 1967 — *Зализняк А. А. Русское именное словоизменение. М., 1967. {Zaliznyak A. A. Russian Nominal Inflection. M., 1967.}*

Карпов 2007 — *Карпов В. Г. Система глагола в современном хакасском языке (структурный и функционально-семантический аспекты) // Хакасский язык: проблемы и перспективы развития. Абакан, 2007. С. 105—161. {Karpov V. G. Verbal system in Modern Khakass (Structural and Functional Semantic aspects) // Khakass: Problems and Ways of Development. Abakan, 2007.}*

Мальцева 2004 — *Мальцева В. С. Структура глагольной словоформы в сагайском диалекте хакасского языка (говор с. Казановка). М., 2004. {Maltseva V. S. Structure of Verbal Wordforms in the Sagay Dialect of Khakass (Kazanovka Sub-Dialect). M., 2004.}*

Новый Завет 2011 — *Новый Завет на хакасском языке с параллельным русским Синодальным переводом. М., 2011. {Gospel in Khakass with Parallel Russian Synodal Translation. M., 2011.}*

Номинханов 1955 — *Номинханов Ц. Д. Интервокальное положение звуков *z* и *z* в хакасском языке // Вопросы хакасского языка и литературы. Абакан, 1955. С. 55—64. {Nominkhanov Ts. D. Intervocalic Position of *z* and *z* Sounds in Khakass // Issues in the Khakass language and Literature. Abakan, 1955. P. 55—64.}*

Шеймович 2017 — *Шеймович А. В. Терминологические соответствия между «Моделью изменяемой хакасской словоформы» и традиционной хакасской грамматикой // Народы Южной Сибири и сопредельных территорий. Материалы Четвертой Международной научной конференции, посвященной 155-летию со дня рождения Н.Ф. Катанова (27—29 сентября 2017 г.). Абакан, 2017. С. 203—212. {Sheymovich A. V. Terminological Correspondences between the “Model of Changeable Khakass Word-form” and Traditional Khakass Grammar // Peoples of Southern Siberia and Neighboring Territories. Materials of the 4<sup>th</sup> International Scientific Conference in Honor of 155<sup>th</sup> Anniversary of N. Katanov (27—29 September 2017). Abakan, 2017. P. 203—212.}*

ЭКХЯ — *Электронный корпус хакасского языка // <http://khakas.altai.ru/>, 2019. {Electronic Corpus of Khakass // <http://khakas.altai.ru/>, 2019.}*

Dybo 2005 — *Dybo A. On the Problems of Oghuz Morphophonology // Turkic Languages, 9. 199—2004.*

Dybo, Sheimovich 2017 — *Dybo A., Sheimovich A. Experience of computer-oriented description of the Tuvan morphophonology within the framework of the system of automatic morphological analysis // V Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017» Труды конференции. Том 2. Казань 2017. С. 27—49.*

EDAL — *Starostin S. A., Dybo A. V., Mudrak O. A. An Etymological Dictionary of Altaic Languages. Leiden, 2003.*

Erdal 2004 — *Erdal M. A Grammar of Old Turkic. Leiden: Brill, 2004.*

Martin 1992 — *Martin G. A Reference Grammar of Korean. Chicago, 1992.*

## РЕЗЮМЕ

В статье приводится описание фрагмента алгоритма работы автоматического парсера для литературного хакасского языка. Парсер осуществляет анализ «через синтез», т. е. путем многократных прямых и обратных проходов с проверкой гипотез. В статье излагается, как происходит перевод морфонологической записи словоформы в орфографическую. Так как хакасский литературный язык слабо нормирован, и не все особенности его морфонологии достаточно полно отражены в существующих грамматиках, для некоторых шагов алгоритма мы выводим правила на основе данных корпуса текстов и собранных нами материалов по хакасским диалектам.

## SUMMARY

The article describes a fragment of the work process of an automatic parser for Literary Khakass. It analyzes data using “synthesis”, i. e., through multiple straight and reverse passes that check various hypotheses. The article describes the transformation of a morphophonological transcription of a word-form into an orthographic one. Literary Khakass is not very normative as a language, and existing grammars do not reflect every feature of its morphophonology well enough. Thus, for some stages of the process we use rules derived from our analysis of the corpus data and of the dialectal materials we have collected.

*Ключевые слова:* хакасский язык, морфонология, агглютинация, автоматический морфологический анализ, лингвистический корпус

*Keywords:* Khakas language, morphophonology, agglutination, automatic parsing, language corpora

Дыбо Анна Владимировна, Институт языкознания РАН (Москва); Томский государственный университет (Томск); [adybo@mail.ru](mailto:adybo@mail.ru)

Anna V. Dybo, Institute of Linguistics of the Russian Academy of Sciences (Moscow); Tomsk State University (Tomsk); [adybo@mail.ru](mailto:adybo@mail.ru)

Крылов Филипп Сергеевич, АНО «Информационная культура» (Москва); [phil.krylov@gmail.com](mailto:phil.krylov@gmail.com)

Philip S. Krylov, ANPO “Information culture” (Moscow); [phil.krylov@gmail.com](mailto:phil.krylov@gmail.com)

Мальцева Вера Сергеевна, Институт языкознания РАН (Москва); Томский государственный университет (Томск); [malt.wh@gmail.com](mailto:malt.wh@gmail.com)

Vera S. Maltseva, Institute of Linguistics of the Russian Academy of Sciences (Moscow); Tomsk State University (Tomsk); [malt.wh@gmail.com](mailto:malt.wh@gmail.com)

Шеймович Александра Валерьевна, Институт языкознания РАН (Москва); Томский государственный университет (Томск); [asheimovich@yandex.ru](mailto:asheimovich@yandex.ru)

Aleksandra V. Sheimovich, Institute of Linguistics of the Russian Academy of Sciences (Moscow); Tomsk State University (Tomsk); [asheimovich@yandex.ru](mailto:asheimovich@yandex.ru)