



*Anna Dybo  
Alexandra Sheymovich  
Institute of linguistics, RAS (Moscow)*



**An Attempt of Computer-Oriented Description  
of Tuvan Morphophonology  
within the Framework  
of the Automatic Morphological Analysis System**



# Selected bibliography

- Aelita Salchak, Aziyana Bayir-ool. The main results of the project on creation an electronic corpus of Tuvan language // Сборник трудов конференции «TurkLang-2015». Казань, 2015. С. 259–268.
- Oorzhak B., Khertek A. Development of semantyc markup the corpus of Tuvan language // Сборник трудов конференции «TurkLang-2015». Казань, 2015. С. 351–362.
- Исхаков Ф.Г., Пальмбаха А.А. Грамматика тувинского языка. М.,1961. 473 с.
- Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. 2014. № 2 (36). С. 20-26;
- Дыбо А.В., Шеймович А.В. Порядковая модель тувинской словоформы // Материалы региональной конференции «Языки народов Сибири и сопредельных регионов». Новосибирск, ИФ СО РАН, 6 – 9 октября 2015 г.
- Тувинско-русский словарь / Под ред. Э.Р. Тенишева. М., 1968.
- Толковый словарь тувинского языка. Т. 1-2. Новосибирск 2011.
- Хертек А.Б., Ооржак Б.Ч. О морфологической разметке электронного корпуса текстов тувинского языка // Филологические науки. Вопросы теории и практики. Тамбов. 2012 г. № 7 (18). Ч. II. С. 214–218.
- Шамина Л.А. Аналитические грамматические формы и конструкции в функции сказуемого в тувинском языке. Новосибирск, 2010 г., 240 с.
- Washington, Bayyr-ool 2016 – Washington J. N, Bayyr-ool A. et al. The development of a finite-state morphological analyser for Tuvan // Родной язык, № 1(4), М., 2016.

# Model of Tuvan wordform

Table 1

0	1	2	3	4	5	6	7		8	9
S	Distr	Perf	Neg	Tense/ Mood	Num(Pl)	Poss	Case		Person (1, 2)	Ptcl
							Simple declencion	Possessive declencion		
	Distr <i>КЫЛА</i>	Perf <i>(Ы)БЫМ</i>	Neg <i>БА</i>	Praes <i>ыр</i>	Pl <i>ЛАр</i>	1pos.sg <i>(Ы)м</i>	Gen <i>НЫң</i>	Gen <i>НЫң</i>	1sg <i>м</i>	Interr <i>(Ы)л</i>
			Neg.Conv <i>БАйн</i>	Past <i>ТЫ</i>		2pos.sg <i>(Ы)ң</i>	Dat <i>ГА</i>	Dat <i>нГА</i>	2sg <i>ң</i>	Add <i>-даа</i>
			Neg.Fut <i>БАС</i>	PastRes <i>ГАН</i>		3pos. (з) <i>Ы</i>	Acc <i>НЫ</i>	Acc <i>Н</i>	1pl <i>БЫс</i>	Indir <i>-ТЫр</i>
				PastRtr <i>ЧЫк</i>		1pl <i>(Ы)БЫс</i>	Loc <i>ТА</i>	Loc <i>нТА</i>	2pl <i>ңАр</i>	Emph <i>-Ла</i>
				Cont <i>БЫшААН</i>		2 pl <i>(Ы)ңАр</i>	Abl <i>ТАн</i>	Abl <i>нТАн</i>	3pl <i>ЛАр</i>	
				Fut <i>Ыр, Ар</i>		Gen.3Pos <i>НЫЫ</i>	Lat1 <i>Че</i>	Lat1 <i>нЧе</i>	Praes.3sg <i>у</i>	
				Cunc <i>ГАЛАк</i>			Lat2 <i>ТЫВА</i>	Lat2 <i>нТЫВА</i>	Imp.1sg <i>БЫн, Айн</i>	
				ConvLim <i>ГАЛА</i>			Lat3 <i>КЫДЫ</i>	Lat <i>КЫДЫ</i>	Imp.3sg <i>СЫн</i>	
				ConvPast <i>(Ы)н</i>					Imp.1.Dual <i>ААл(Ы), БЫл(Ы)</i>	
				ConvPraes <i>А, Ы, ү</i>					Imp.1.pl <i>ААлЫңАр, БЫлЫңАр,</i>	
				Prosp <i>ГАш</i>					Imp.2.pl <i>(Ы)ңАр</i>	
				ConvLim2 <i>ГЫже</i>					Imp.3.pl <i>СЫн.(нАр)</i>	
				Abstr <i>КЫ</i>					1.sg.Lim <i>м.че</i>	
				Opt. <i>ГАй</i>					2.sg.Lim <i>ң.че</i>	
									1.pl.Lim <i>вис.че</i>	
									2.pl.Lim <i>ңер.же</i>	
									1.sg.Cond <i>СЫ.м.зА</i>	
									2.sg.Cond <i>СЫ.ң.зА</i>	
									3sg,pl Cond <i>зА</i>	
									1.pl.Cond <i>СЫ.вЫс.сА</i>	
									2.pl.Cond <i>СЫ.ңАр.зА</i>	

Cumulatively expressed grammemes are separated by dots.



## Meaning for symbols in the morphophoneme system

Consonant morphophonemes

*Б: б/п/м/в*

*К: г/к*

*Г: г/к/0*

*Т: т/д*

*Н: д/т/н*

*С: с/з*

*Л: л/т/д/н*

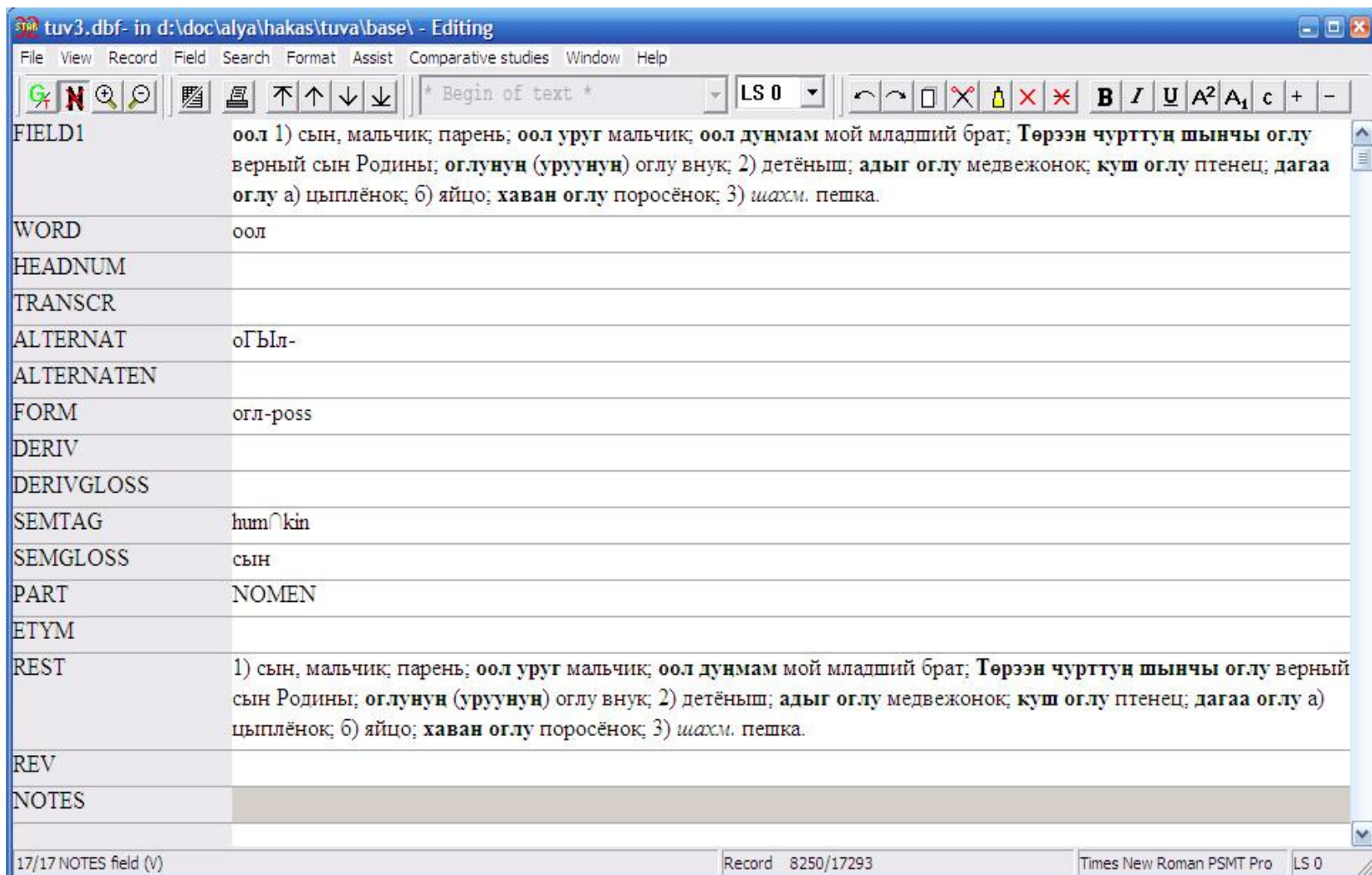
*Ч: ч/ж/ш*

Vowel morphophonemes

*А: е, а*

*Ы: и, ы, у, ʏ*





The screenshot shows a database editing window titled 'tuv3.dbf- in d:\doc\alya\hakas\tuva\base\ - Editing'. The window contains a table with the following fields and values:

FIELD1	оол 1) сын, мальчик; парень; оол уруг мальчик; оол дунмам мой младший брат; Төрээн чурттуң шынчы оглу верный сын Родины; оглунун (уруунун) оглу внук; 2) детёныш; адыг оглу медвежонок; куш оглу птенец; дагаа оглу а) цыплёнок; б) яйцо; хаван оглу поросёнок; 3) шахм. пешка.
WORD	оол
HEADNUM	
TRANSCR	
ALTERNAT	оГЫл-
ALTERNATEN	
FORM	огл-poss
DERIV	
DERIVGLOSS	
SEMTAG	hum <sup>o</sup> kin
SEMGLOSS	сын
PART	NOMEN
ETYM	
REST	1) сын, мальчик; парень; оол уруг мальчик; оол дунмам мой младший брат; Төрээн чурттуң шынчы оглу верный сын Родины; оглунун (уруунун) оглу внук; 2) детёныш; адыг оглу медвежонок; куш оглу птенец; дагаа оглу а) цыплёнок; б) яйцо; хаван оглу поросёнок; 3) шахм. пешка.
REV	
NOTES	

The status bar at the bottom indicates '17/17 NOTES field (v)', 'Record 8250/17293', 'Times New Roman PSMT Pro', and 'LS 0'.

Fig. 1. An example of a nominal article from the Russian-Tuvan digital database

The screenshot shows a software window titled 'tuv3.dbf- in d:\doc\alya\hakas\tuva\base\ - Editing'. The window contains a table with the following fields and values:

FIELD1	чынныр /чылын*/ <i>возвр. от чылы*</i> (см. чылыыр) греться (у огня).
WORD	чынныр
HEADNUM	
TRANSCR	
ALTERNAT	чылын
ALTERNATEN	Ы
FORM	
DERIV	чыл=Ын-
DERIVGLOSS	греть=Refl
SEMTAG	changest(temper) Subj(hum/ animal)
SEMGLOSS	греться
PART	VERBUM
ETYM	
REST	<i>возвр. от чылы*</i> (см. чылыыр) греться (у огня).
REV	

The status bar at the bottom indicates '17/17 NOTES field (v)', 'Record 15162/17293', and 'Times New Roman PSMT Pro LS 0'.

Fig. 2. An example of a verbal article from the Russian-Tuvan digital database



## 0. Segmental rules for allomorph choosing

(surface morphological rules for choosing one of the morpheme's allomorphs, with established morphemic composition of the word form):

**0.1.**

$-(V)\{Mrph_{n+1}\} \rightarrow -V\{Mrph_{n+1}\}/\{Mrph_n\}C\_$	$am-(BI)_M \rightarrow am-BIM$
$-(V)\{Mrph_{n+1}\} \rightarrow -\emptyset\{Mrph_{n+1}\}/\{Mrph_n\}V\_$	$ada-(BI)_M \rightarrow ada-M$
$-(C)\{Mrph_{n+1}\} \rightarrow -C\{Mrph_{n+1}\}/\{Mrph_n\}V\_$	$ada-(z)BI \rightarrow ada-zBI$
$-(C)\{Mrph_{n+1}\} \rightarrow -\emptyset\{Mrph_{n+1}\}/\{Mrph_n\}C\_$	$ad-(z)BI \rightarrow ad-BI$

**0.2.**  $A, BI \{Mrph_{n+1}\} \rightarrow V=V_{ALTERNATEN(S)}\{Mrph_{n+1}\}$

*ак-* ‘to flow’ – Fut. *аг-ар*, ConvPraes *аг-а*, Imp.1Pl *аг-айн* (A)

*ал-* ‘to take’ Fut. *ал-ыр*, ConvPraes *ал-ы*, Imp.1Pl *ал-ыйн* (BI)

**0.3.**  $\{Mrph_n\}V\text{-ConvPraes} \rightarrow \{Mrph_n\}V\text{-й}$

*бодa-* ‘to think’ – *бодa-й* ‘thinking’

*тaры-* ‘to sow, to plant’ – *тaры-й* ‘sowing, planting’



## 1. Morphological alternations

use information from morpheme dictionaries (from the grammatical dictionary of stems – e.g. data on lexico-grammatical class of a stem or dictionary data of this stem, – and from the table of inflectionary affixes – e.g. the grammeme of this morpheme).

**1.1.**  $(N(V)>1)\{Mrph_n\}K-,к- \rightarrow (N(V)>1)\{Mrph_n\}\Gamma-/_\{(Poss)Mrph_{n+1}\}$

*тавак-Ыл* → *таваГ-Ыл* ‘my plate’;

*белек-Ыл* → *белеГ-Ыл* ‘my present’,

*кезек-Ы* → *кезеГ-Ы* (> *кезээ*) ‘his part’

But : *аак-Ы* → *аак-Ы* ‘his consequence’

**1.2.**  $(N(V)=1)\{Mrph_n\}к-,п-,м-,л-,[н-] \rightarrow (N(V)=1)\{Mrph_n\}\Gamma-/_\{(ConvPast)Mrph_{n+1}\}$

*бол-Ын* → *боГ-Ын/бол-Ын* (→ *боон/болун*) ‘having been’,

*кел-Ын* → *кеГ-Ын* (→ *кээн*) ‘having come’,

but: *бил-Ын* → *билип* ‘having known’.

? *доң-Ын* → *доон/доңун* ‘having frozen’

NB: *ал-Ын* → *ап* ‘having taken’,

*кып-Ын* → *хып* ‘having caught fire’

*сок-Ын* → *сон* ‘having punched’

**1.3.**  $(N(V)=1)\{Mrph_n\}р-,л- \rightarrow (N(V)=1)\{Mrph_n\}\Gamma-/_\{(Fut)Mrph_{n+1}\}$

*кел-* + *Ыр* → *кеГ-Ыр/кел-Ыр* (> *кээр/келур*) ‘to come’,

*бер-* + *Ыр* → *беГ-Ыр* (> *бээр*) ‘to give’



1.4. (C)VСЫС {Mrph<sub>n</sub>}- → (C)VСС {Mrph<sub>n</sub>}-/\_V {Mrph<sub>n+1</sub>}

*ойЫн-Ы* → *ойн-Ы* → *ойну* ‘his game’

*эрин-Ы* → *эРН-Ы* → *эрни* ‘his lip’

*оГЫл-Ы* → *огл-Ы* → *оглу* ‘his son’

*бурЫн-Ы* → *мурну* ‘in front of, before’ (3Poss from *бурун* ‘former’)

But cf.: *бүрЫн-Ы* → *бүрүн-ү*: *бүрүнү биле* ‘fully’ (3Poss from *бүрүн* ‘everything’)

*дайын-ы* ‘war’ → *дайын-ы*: *Ада-чурттуң Улуг дайыны* ‘Great Patriotic War’

*хайын-Ыр* → *хайн-Ыр* → *хайныр* ‘to boil’

*дир=Ыл-Ы* → *дирл-Ы* → *дирли* ‘having been revived’

*ажы=н=Ыш-Ыр* → *ажыниш-Ыр* → *ажынчыр* ‘to be angry at each other’

Cf.: *ажын-ар* /*аШы=н- 1*/ ‘to be angry, to sulk’

*ашт-ыр* /*аШы=н- 2*/ ‘to exonerate oneself’

### 1.5. Unique shifts:

*бөрт-Ы* → *бөрг-ү* ‘his headdress’

*аас-Ы* → *аксы* ‘his mouth’

## 2. Morphophonological rules

use is determined by the segmental composition of morphemes combining information on morphemic boundaries.

**2.1.** Assimilation in combinations of consonant morphonemes not separated by the inflectional boundary -

**2.1.1.** лV<sub>high</sub>Н- → НН-

*келин-ЫI → келн-ЫI → кенн-ЫI → кенни* ‘his daughter-in-law’;

*кыл=ын-ЫIр → кылн-ЫIр → кынн-ЫIр → кынныр* ‘to become’;

**2.1.2.** C<sub>fortis</sub>V<sub>H</sub>-, C<sub>fortis</sub>V<sub>L</sub>- → C<sub>fortis</sub>T-:

*иШ=ин-ЫI → иШн-ЫI → ишт-ЫI → ишти* ‘his stomach’

*эКин-ЫI → эКн-ЫI → экт-ЫI → экти* ‘his shoulder’

*саКын-ЫIn → сактын* ‘having remembered’

*оТун-ЫIр → оттур* ‘to wake up’

*тыП=ыл-ЫIр → тыптыр* ‘to be found’

**2.1.3.** C<sub>fortis</sub>V<sub>Ш</sub>-, лV<sub>Ш</sub>- → Сч-; рV<sub>Ш</sub>- → рж-:

*диК=иш-ЫIр → диКчир* ‘to help build’,

*кыП=ыс-ЫI → кыпсы* ‘igniting’

*таПар=ыш-ЫIр → таваржыр* ‘to meet’

**2.2.** Alternating initial consonants in a stem during inflection

**2.2.1.** #б{S}...н-{Mrph<sub>n</sub>} → #м{S}...н-{Mrph<sub>n</sub>}

*бил=ин-ЫIр → билн-ЫIр → бинн-ЫIр → миннир*;

*бурЫн-ЫI → мурну* ‘in front of, before’ (3<sup>pos</sup> from *бурун* ‘former’)

**2.2.2.** #C<sub>fortis</sub>{SVERBUM}...ЪC<sub>fortis</sub>-V{Mrph<sub>n</sub>} → #C<sub>lenisM</sub>{S}...ЪC<sub>fortis</sub>-{Mrph<sub>n</sub>}

*хыън-Ар → хыън-Ар → кыъвар* ‘to burn’

*төък-Ар → төък-Ар → дөъгер* ‘to pour out’

All these alternations are written in the dictionary in the ALTERNATEN field if they are reflected in orthography.



## 2.3. Assimilation of consonants separated by the inflectional morphemic boundary

Affixes starting with a certain morphoneme receive an implementation of this morphoneme, depending on the quality of the final morphoneme of the preceding morph.

Таблица 2

Initial morphonemes of affixes / Final morphonemes of the stem	<b>Б</b>	<b>К</b>	<b>Т</b>	<b>Н</b>	<b>С</b>	<b>Л</b>	<b>Ч</b>
<b>В</b>	В	Г	Д	Н	З	Л	Ж
<b>т, Т</b>	П	К	Т	Т	С	Т	Ч
<b>п, П</b>	П	К	Т	Т	С	Т	Ч
<b>М</b>	М	Г	Д	Н	З	Н	Ч
<b>н</b>	М	Г	Д	Н	З	Н	Ч
<b>ң</b>	М	Г	Д	Н	З	Н	Ч
<b>л</b>	б	Г	Д	Д	З	Д	Ч
<b>р</b>	б	Г	Д	Н	З	Л	Ж
<b>й</b>	б	Г	Д	Д	З	Л	Ж
<b>с, С</b>	П	К	Т	Т	С	Т	Ч
<b>к, К</b>	П	К	Т	Т	С	Т	Ч
<b>г, Г</b>	б	Г	Д	Д	З	Л	Ж
<b>ч, Ч</b>	П	К	Т	Т	С	Т	Ч
<b>ш</b>	П	К	Т	Т	С	Т	Ч





## A fragment of the acceptable combination of consonants system in Tuvan

Таблица 3

	<b>Т</b>	<b>Д</b>	<b>Л</b>	<b>Н</b>	<b>С</b>	<b>З</b>	<b>Ч</b>	<b>Ш</b>	<b>Ж</b>
<b>С</b>	<i>кестик</i>	-	-	-	<i>бассын</i>	-	<i>касчыр-</i>	-	-
<b>Т</b>	<i>эъттиг</i>	-	-	-	<i>четсе</i>	-	<i>бутчак</i>	-	-
<b>К</b>	<i>өөктээр</i>	-	-	-	<i>сукса-</i>	-	<i>көрүкчү</i>	<i>акша</i>	-
<b>Ш</b>	<i>чештин-</i>	-	-	-	<i>ышсыг</i>	-	<i>дашчы</i>		-
<b>Р</b>	<i>эрте</i>	<i>көрдү</i>	<i>баарлыг</i>	<i>силерниң</i>	-	<i>барза</i>	<i>кадарчы</i>		<i>хааржак</i>
<b>Л</b>	-	<i>малдың</i>	-	-	-	<i>элзит-</i>	<i>болчур</i>		-
<b>М</b>	-	<i>амдан</i>	-	<i>эмне-</i>	-	<i>хүлүмзүр-</i>	<i>кымчы</i>		-
<b>Н</b>	-	<i>ындыг</i>	-	<i>хүннер</i>	-	<i>хүнзе-</i>	<i>кинчи</i>		-
<b>Ң</b>	-	<i>аңдар-</i>	-	<i>деңне-</i>	-	<i>даңзы</i>	<i>аңчы</i>		-
<b>Й</b>	<i>ойтаяр</i>	<i>шайдан</i>	<i>чайлаг</i>	<i>ойнаар</i>	-	<i>дайзын</i>	-		<i>хоорайже, далайжы</i>
<b>Г</b>	-	<i>дагда</i>	<i>суглук</i>	<i>дагны</i>	-	<i>аарыгзыыр</i>	<i>садыгчы</i>		<i>дагже</i>





### 3. Phonological rules

that include only phonological conditions (independent from the morphological division of the word form).

#### 3.0. $V_1-V_2 \rightarrow V_1V_1$

*ойна-Ар* → *ойнаар* ‘to play’

*сөглe-Ар* → *сөглээр* ‘to say’

*чыры-Ыр* → *чырыыр* ‘to shine’

#### 3.0.1. - → ∅

#### 3.1. $V_1\Gamma,\Gamma V_2 \rightarrow V_1V_1$

*оГЫл* > *оол*, *аГЫс* > *аас*, *таваГ-Ым* → *таваам* ‘my plate’; *уруг-Ы* → *уруу* ‘his child’; *белеГ-Ым* → *белээм* ‘my present’; *кес=eГ-Ы* → *кезээ* ‘his part’; *даг-Ы* → *даа* ‘his mountain’; *суг-Ы* → *суу* ‘his water’; *саг-Ын* → *саан* ‘having milked’, *каг-Ар* → *каар* ‘to leave behind’, *чуг-Ыр* → *чуур* ‘to wash’, *кеГ-Ыр/кел-Ыр* → *кээр/келур* ‘to come’, *беГ-Ыр* → *бээр* ‘to give’.

#### 3.2. $V_1C_{fortis}V_2 \rightarrow V_1C_{lenis}V_2$

*(ам-Ы)* → *амы* → *адЫ* ‘his name’, *час-Ын* → *чазЫн* ‘in springtime’, *аак-Ы* → *аазЫ* ‘his consequence’: *соок аазы* ‘consequences of the frost’; *эКЫн* ‘shoulder’ → *эГЫн* (→ *эГин* → *эгин*), *уШ=Ын* ‘stomach’ → *иЖин*.





### 3.3. Palatal and labial vowel harmony

Таблица 4

$V_{n-1} \backslash V_n$	А	Ы
$V_{back}$	а	Ы
$V_{front}$	е	И
$V_{front\ lab}$	е	Ү
$V_{back\ lab}$	а	у

**3.4.** All the symbols (uppercase Cyrillic letters) that survived at this stage transform into the corresponding lowercase letters and bring back the orthographical form of the Tuvan word form (before the use of graphical rules).

(*am-ЫI* →) *адЫI* → *ады* ‘his name’, (*час-ЫIn* →) *чазЫIn* → *чазын* ‘in springtime’, *эГин* → *эгин* ‘shoulder’.





#### 4. «Graphical sandhi»

recording the phonetic combinations like *й+V* using iotated Cyrillic letters:

**4.1.** *йу* → *ю*, *йа* → *я*, *йе* → *е*.

*хой* ‘sheep’ – Poss.1Pl (*хой-ЫвЫс* →) *хойуvus* → *хоюvus* ‘our sheep’;

*ой-* ‘to cut through, to punch through’ – Prosp (*ой-КАш* → *ой-ГАш* →) *ойгаш*, Infinitive (*ой-Ар* →) *ойар* → *ояр*, ConvPast (*ой-Ын* →) *ойун* → *оюн*;

*өй-* ‘to roll’, Prosp (*өй-КАш* → *өй-КАш* →) *өйгеш*, Infinitive (*өй-Ар* →) *өйер* → *өер*, ConvPast (*өй-Ын* →) *өйүн* → *өйүн*.

**4.2.** *ее* → *ээ*:

(*кел-Ыр* →) *кеер* → *кээр* ‘to come-Fut’





## Conclusion

- ✓ Segmental processes that determine the synchronous appearance of a Tuvan word form are a complex multilevel system that emerged as a conglomeration of phonetic processes in different times that were gradually pushed down to deeper levels of the language, and also washed out by analogous processes (cf. "Partial restoration of the historical appearance" ГТЯ 118).
- ✓ An automatic morphological analyzer can bypass some of these phenomena with the help of "lexical exceptions" by placing the relevant information in the lexical base dictionary entries, but a significant number of them should still be processed by the algorithm.
- ✓ Creating the algorithm for processing segment transformations for the parser has led to discovering significant gaps in description of the segmental behavior of a number of morphemes and types of morphemes in existing dictionaries and grammars of the Tuvan. These gaps can be eliminated by corpus and field studies.

