APPENDIX

Sumerian-Ugric Protowords and Regular Sound Changes

Peter Z. Revesz

School of Computing
University of Nebraska-Lincoln

1. Introduction

Revesz (2019) has shown that Sumerian is an Ugric language by presenting a set of regular sound change rules for a proposed West-Ugric branch of the Uralic language family, which is shown in Figure 1. The name West-Ugric implies those Ugric languages that are located west of the Ural Mountains, that is Hungarian, Minoan and Sumerian in contrast to the Ob-Ugric languages of Khanty and Mansi that are spoken to the east of the Ural Mountains. The sixteen basic regular sound change rules shown in Table 3 were identified by using associative rule data mining on a set of cognate words from the first volume of EDSL and the Pennsylvania Dictionary of Sumerian (2021).

In this impressive third volume, Simo Parpola agrees that Sumerian is an Ugric language, although there are some minor differences in view. First, he argues that Sumerian is closer to the Ob-Ugric languages of Khanty and Mansi instead of Hungarian. Second, he assumes a broader set of sound changes than the sixteen regular sound change rules in Revesz (2019).

A major problem with a broader set of sound change rules is that it makes it difficult to reconstruct protowords because each word could be derived from several alternative protowords. In contrast, a narrower set of regular sound change rules makes the reconstructions more credible. Hence a general acceptance of Sumerian as an Ugric language cannot be achieved by making dictionaries alone but credible protoword reconstructions based on regular sound change rules are also needed.

Uralic linguists often criticized previous Sumerian etymological dictionary authors for lack of regular sound changes among the proposed cognates in their dictionary entries. Moreover, Uralic linguists take a fairly strict meaning of ’regular’ when it comes to sound changes. It would be disappointing if their reception of this third volume were as negative as the reception of the previous two volumes. The goal of this appendix is to help the situation by addressing some of their concerns. In particular, we aim to do the following:

Goal: Find protowords for the Ugric-related entries in the EDSL. For each entry the protoword together with the regular sound changes should be able to explain all the cognate words in that entry.

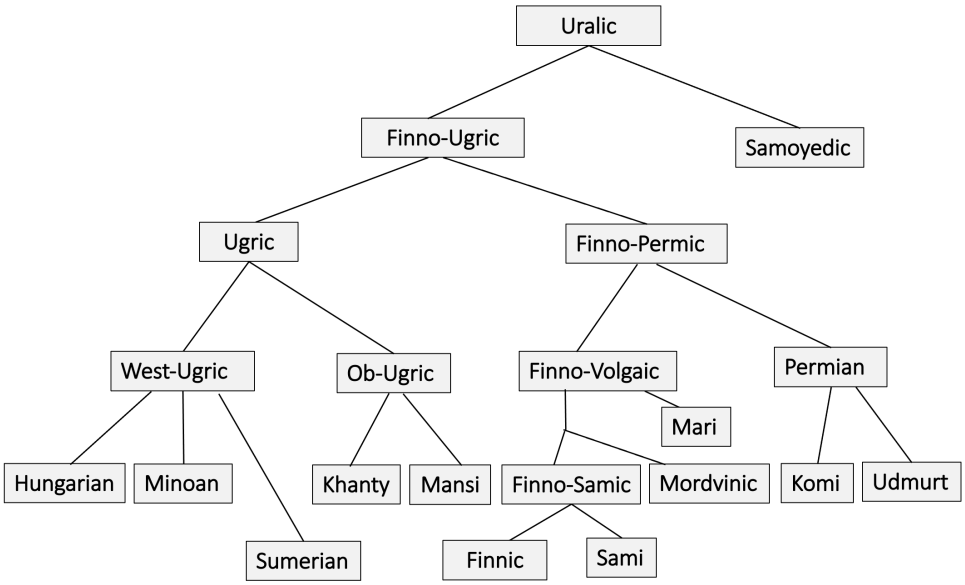


FIGURE. 1. *The Uralic language family tree following the computational phylogenetic study of Honkola et al. (2013) but extended by the West-Ugric branch following Revesz (2017, 2019).*

The rest of this appendix is organized as follows in order to accomplish the stated goal. Section 2 gives an overview of Ugric phonology. Section 3 reviews the 16 regular sound change rules of Revesz (2019). Section 4 applies these rules to the sample words provided by Professor Parpola. Section 5 gives a satistical analysis of the results. Section 6 presents a brief overview of the new horizons of Uralic linguistics created by the identification of Sumerian as an Ugric language. Finally, Section 7 gives some concluding remarks and directions for further work.

2. Ugric Phonology

We believe that the Proto-Ugric language had 16 different groups of consonants as shown in Table 1. These groups were derived by a careful consideration of what are the common phonemes in the various Ugric languages currently and in their known histories. For Hungarian phonological development we relied on Kiss and Pusztai (2018). These 16 different groups of consonants can be further grouped into five larger groups:

Sibilants: This group of phonemes includes /š/, /s/, /č/, and /ś/ and are found as the first three data rows of Table 1.

Plosives: This group of phonemes includes /d/, /t/, /d'/, /t'/, /mp/, /p/, /h/, /k/ and is found in data rows 4-10 of Table 1. The voiceless velar fricative /x̱/ is included in this group because it is a spirantized /k/.

Liquids: This group of phonemes includes /l/, /l'/ and /r/ and are found in data rows 11-12 of Table 1.

Nasals: This group of phonemes includes /m/, /n/, /ń/ and /ŋ/ and are found in 1data rows 13-14 of Table 1.

Semivowels: This group of phonemes includes /w/ and /j/ and are found in 1data rows 15-16 of Table 1.

The above grouping allows a deeper analysis of regular sound chages than a simple alphabetical listing because, as we will see in the next section, changes that happen to a member of any group also tend to happen to other members of the same group.

TABLE. 1. *Proto-Ugric phonemes grouped into 16 smaller and five larger groups. Our phonetic notations, which are shown in the second column, differ slightly from the International Phonetic Alphabet (IPA) notation, which can be found in the fifth column.*

#	Our Notation	Group	Type	IPA Notation	English example
1	š, s	sibilants	postalveolar and alveolar sibilant fricatives	ʃ, s	**sh**y, **s**igh
2	č		postalveolar sibilant affricate	tʃ	**ch**ase
3	ś		palatized /s/, absent in Hungarian, Khanty	sj	con**s**ume
4	d	plosives	voiced dental plosive	d	**d**ye
5	t		voiceless dental plosive	t	**t**ie
6	d', t'		palatized /d/ and /t/	ɟ, c	**d**ew, **t**une
7	mp		/m/ and /p/ combination, precursor of H /b/		
8	p		bilabial plosive	p	**p**ie
9	ḫ		spirantized /k/, precursor of H /h/, S /g/	x	lo**ch**
10	k		voiceless velar plosive	k	s**k**y
11	l, l'	liquids	voiced alveolar lateral approximant, palatized /l/	l, lj	**l**ie, **l**ute
12	r		voiced alveolar approximant	ɹ	**r**ye
13	m	nasals	voiced bilabial nasal	m	s**m**ile
14	n, ń, ŋ		voiced alveolar, palatal and nasals	n, ɲ, ŋ	ca**n**, ca**ny**on, si**ng**
15	w	semivowels	voiced labio-velar approximant semivowel	w	**w**ine
16	j		voiced palatal approximant semivowel	j	**y**ou

In addition to the Proto-Ugric phonemes of Table 1, there are several additional phonemes that occur in specific Ugric languages. When we use them, we denote these apparently later introduced phonemes as follows: /b/, /f/, /g/, /ɬ/, /ŧ/, /v/, /z/ and /ž/. These notations agree with the International Phonetic Alphabet (IPA) notation except for the voiced velar lateral /ɬ/, voiceless dental fricative /ŧ/, and the voiced postalveolar fricative /ž/, which are respectively denoted as /ɬ/, /θ/ and /ʒ/ by the IPA. In the rest of this appendix, we will refer to all Ugric words following

a phonetic transcription where the consonants are transliterated as shown in Table 2.

TABLE. 2. *Transliteration of Ugric consonant phonemes into our notation.*

#	Our Notation	Hungarian	Khanty	Mansi	Sumerian
1	š s	s sz			
2	č	cs			
3	ś			ś	
4	d		ɒ, δ	δ	
5	t				
6	d’ t’	gy ty			
7	mp				
8	p		ṕ	в	
9	ḫ		x, χ	χ, χ́	
10	k			ǩ	
11	l l’	 ly	ļ, л, ʌ	ʟ	
12	r				
13	m				
14	n, ń, ŋ	 ny ng, nk	ŋ	ŋ ŋₒ	 ĝ
15	w		u̯	û̯	
16	j				
	g		γ	ɢ, γ	
	ŧ		ϑ	ϑ	
	v		β	β	
	z				ṣ
	ž	zs			

We also transliterate the Ob-Ugric vowels, that is those that are common to Khanty and Mansi as follows:

â, å, a̋, oå → a
öä, eä, ę, ē, ɛ → ä
ė, ɜ → ə
o̯ → o
üö → ö
į, ü → ü.

Furher, the Khanty vowels as follows:

ā, á, à, ă, a̋, ȧ, ɒ → a
a̋, oå → å
ä̋, äₒ → ä
ō, ó, ò, o̯, o͡, °, ȯ, o̯, ɔ̄, ɔ, ɵ → o.

The Mansi vowels as follows:

$k_u \rightarrow kw$
$\varepsilon \rightarrow e$
$ı, ɪ̣ \rightarrow i$
$ụ \rightarrow u$
$ū, ɯ \rightarrow ü$
$^{u}w \rightarrow w$.

Finally, from the Sumerian EDSL dictionary words we removed the diacritic accents, the subindexes, the determinatives ($^{d, gi, ĝiš, ku_6, ú, urudu, uzu}$, etc.), the hyphens after verbal stems, and the grapheme boundaries, that is, we rewrote a.a → a not aa, e.e. → e, i.i → i, u.u → u. However, we changed ka_5.a → kaj.

3. Regular Sound Change Rules

We used associative data mining to find the initial set of regular sound change rules (Revesz 2019b, 2020). This automated process returned candidate sound change rules such that each rule was supported by several examples. This process eliminates from consideration those rules that do not have enough supporting examples. This is usually an advantage, but it may overlook some rarer regular sound changes that were added manually if they fit the pattern of their group. Table 3 shows the regular sound change rules.

The blank regions in Table 3 are not applicable. For example, Rule 1 states the regular changes for the phones /š/ and /s/. These are both omitted in Hungarian when they occur in word initial position but can change to any of the fricatives /š/, /s/, /z/ in either word medial or word final positions. In Sumerian, these regularly change to the same fricatives in any position. The Proto-Ugric form is generally preserved in the Khanty and the Mansi languages, although the word intitial /š/ and /s/ can change to a /t/ sometimes. The /t/ changes to /z/ in Hungarian only in some word final positions.

Below we use the notation $w_1 \sim w_2$ means that the two words w_1 and w_2 are cognates and $*w_1 > w_2$ means that the hypothetical word w_1 is the origin of the attested word w_2. We also use the abbreviation H for Hungarian, K for Khanty, M for Mansi, PU for Proto-Ugric, and Su for Sumerian. We will also represent all words in phonetic transcription instead of using their exact spelling. A good example of Rule 1 is the following based on entry 2943 in the EDSL:

PU */šäppä/ > H /epe/ ~ K /siv/ ~ M /tep/ ~ Su /zebba/ (Rules 1 and 8)

Here the word initial /š/ changed to Hungarian by the omission of this consonant, while in Khanty it changed to /s/ and in Sumerian to /z/. The existence of /s/ in Khanty supports the idea that the PU word had an initial /š/ and the change from that to /t/ occured later with the Mansi after the Ob-Ugric languages were separated from each other. Simultaneously, the above also is a good example of Rule 8. Here the medial PU /pp/ changes to /p/ in Hungarian, /v/ in Khanty, /p/ in Mansi, and /bb/ in Sumerian.

TABLE. 3. *The 16 regular sound change rules from (Revesz 2019) extended with subcases.*

#	Proto-Ugric	Hungarian	Khanty	Mansi	Sumerian	*PU ~ Hungarian ~ Sumerian (EDSL entry) The rules do not apply in the blank regions.	
						Initial	**Medial or Final**
1	š, s	- (Initial)	l, t (Initial)	t (Initial)	š, s, z	*šenk# ~ ég ~ šeĝ6 (2340)	
		d', š, s, z, - (Me/Fi)	s, t, - (Me/Fi)	s, t, t', - (Me/Fi)			*pesä ~ fészek ~ pisaĝ (1998)
2	č	č, d', š, s, ž	č, s, ŧ	š, s	š, s, z	*čur ~csurgó ~ sur (2277)	*aruča ~ rozs ~ arz-ig (161)
3	ś	š, s, z	s	s	š, s, z	*śäweU ~ sző ~ sá (2059)	*ńeśe ~ nyöszörög ~ eš9 (714)
4	d	d, l	l, t	l	d	*dereŋ ~ derék~ dirig (510)	*sid ~ szid ~ šid (2381)
5	t, tt	d, t (In/Me/Fi)	t	t	d, š, t	*tes ~ tesz ~ tuš (2617)	*kota ~ ház ~gud (875)
		z (last Syl.)					
6	d', t'	d', j, l', t' (Me/Fi)	j, t'	l', t'	d, r		*at'a ~ atya ~ ad (39)
		š, s (last Syl.)					*pät' ~ fos ~ bed (317)
7	mp	b	mp	mp	b		*kump ~ hab ~ gúb (867)
8	p, pp	b, f (Initial)	P (Initial)	p	b, bb, p	*punaU ~ fon ~ pan (1953)	
		p, v, - (Me/Fi)	p, w (Me/Fi)				*sep ~ szép ~ zé-eb (2945)
9	ḫ	h	h, k	h, k	h, -	*ḫaś# ~ hasad ~ ḫaš (1129)	*kuhu ~ köhög ~ huhu (1163)
10	k, kk	h (Ini+back V)	h, g, k	h, k (Initial)	g, h, k, -	*kara ~ harap ~ kur8 (1476)	
		g, k, - (other)		h, g, k, w (Me/Fi)		*kiń ~ kin ~ kíĝ (1386)	*suk ~ zúg ~ šeg12 (2141)
10b	kt, (kd)	t, (d)	t	t	gg, kk		*tekte# ~ tetű ~ tagga (2481)
11	l	l	l	l	l, r	*lep ~ lop ~ lib (266)	*čilag ~ csillog ~ zalag (2926)
11b	l'	l', d', j	j, l', ł	l'	l	*l'uk# ~ lyuk ~ lug (1600)	*hal' ~ hagy ~ hal (1109)
12	r	r	r	r	r	rog ~ rokon ~ rugu (2050)	*huru ~ harag ~ úrgu (2818)
13	m	m (Initial)	m	m	b, m (Initial)	*mar ~ mar ~ mir (1083)	
		m, v, - (Me/Fi)			m (Me/Fi)		*nime ~név ~ nam (1825)
14	n	n, ń	n	n, ń	n	*neŋi ~ néné ~ nin9 (1883)	*pan ~ fon ~ pan (1952)
14b	ń	ń	ń	ń	n, -	*ńaŋar ~ nyír ~ naĝar (1822)	*munaU ~ mony ~ mu (1726)
14c	ŋ	g, k, ng, nk, d', ń, v	g, m, ŋ, ŋg, ŋk, w	g, k, m, ŋ, ŋh, ŋk	g, j, m, n, ŋ		*jäŋeU ~ jég ~ aj (3)
14d	nt	d, t	nt	nn, nt	d		*jant# ~ játék ~ enedi (666)
15	w	v, -	w, -	w, -	b, m, -	*wis ~ víz ~ eš (715)	*kiw# ~ hév ~ kím (1383)
16	j, -	d', j, -	j, -	j, -	-	*juče ~ juh ~ u8 (2644)	*sa ~ szíj ~ zà (2895)
16b	jk	g, h, -	g	g, w	ŋ		*pojkaU ~ fiú ~ buŋa (362)

An example of Rules 2 and 12 is the following based on entry 2281 in the EDSL:

PU */čor/ > H /čor-og/ ~ K /sori/ ~ M /čor-x/ ~ Su /sur/ (Rules 2 and 12)

Here the word initial /č/ is preserved in Hungarian and Mansi and changes to /s/ in Khanty and Sumerian. In addition, the word final /r/ is preserved in all four languages.

An example of Rules 3 and 15 is the following based on entry 2059 in the EDSL:

PU */śäwe/ > H /ső/ ~ K /sö-g/ ~ M /sä-g/ ~ Su /sa/ (Rules 3 and 15)

Here the word initial /ś/ changes to /s/ and the medial /w/ is omitted in all four languages. In addition, both Khanty and Mansi add the same /g/ suffix. This suffix may have been already present in the Proto-Ob-Ugric language which was spoken by the common ancestors of the Khanty and the Mansi. The missing of this suffix in both the Hungarian and the Sumerian words suggests that they are closer related, and it makes sense to group them into a common West-Ugric branch that is separate from the Ob-Ugric branch.

An example of Rules 4 and 8 is the following based on entry 1990 in the EDSL:

PU */pede/ > H /fül/ ~ K /pət/ ~ M /pel/ ~ Su /bed/ (Rules 8 and 4)

Here the word initial /p/ changed to /f/ in Hungarian, /b/ in Sumerian and preserved as /p/ in both Khanty and Mansi. The medial /d/ changed to /l/ in Hungarian and Mansi, /t/ in Khanty, and /d/ in Sumerian.

Finally, let us consider the following example of Rules 10 and 14 based on entry 1386 in the EDSL:

PU */kin/ > H /kín/ ~ K /kün/ ~ M /hün/ ~ Su /kin/ (Rules 10 and 14)

Here the initial /k/ is preserved in all languages except Mansi, where it is leniated to /h/. In addition, the word final /n/ is preserved in all four languages. In the application of Rule 2, we can see that there is a front vowel /i/ after the initial /k/. Hence the Hungarian word correctly preserves the /k/ initial phoneme in this case.

Only a few regular sound change rules of Table 3 allow the of consonants. Intuitively, rules should be rare. There is a huge difference between and more typical regular sound changes such as palatalization. For example, consider the English words *win, wit* and *with*. Suppose that a person with a heavy accent would pronounce these words as /wiń/, /wit'/ and /wid'/. That would be strange but perfectly understandable. However, another person who drops the last consonants would be incomprehensible. It would be easier to communicate with that person by asking him to write down what he wants to say with *wi* for *win*, *wi*$_2$

for *wit*, and *wi₃* for *with*. In fact, Sumerian scribes did something similar because Sumerian had many similar sounding words that were distinguished by different cuneiform signs. Nevertheless, many consonant rules are still not warranted because there is a danger that many words that are not true cognates will become derivable from hypothetical protowords that never existed. However, Table 3 contains some regular sound change rules that allow the of consonants. For example, Rule 14b allows the of Sumerian medial /ń/ as illustrated by the following examples.

Figure 2a shows the derivation of the Ugric words in entry 1726. Here we can assume **mon* as the Proto-Ugric protoword because it fits well with the Ob-Ugric languages. Moreover, according to the regular sound change Rule 14, Proto-Ugric word final /n/ could change to /ń/ yielding the Hungarian word *moń*. If we assume that this change took place at an early stage on the development from Proto-Ugric to Hungarian, then it could presumably already have occurred in West-Ugric leading to the West-Ugric protoword **moń*. Then word final /ń/ could be omitted to yield Sumerian *mu* by Rule 14b.

Figure 2b shows Sumerian word *ku* in entry 1421 and the UEW's Ugric entries related to **kuńa* 'to sleep.'

Figure 2c shows the derivation of the Ugric words in entry 1417 to which we add Mansi *künš-kaškej* 'ant' listed in the UEW (Rédei 1988). Here we assume that the hypothetical Ugric protoword is **konši*. Then the West-Ugric **końši* is obtained by Rule 14. The Hungarian *hańd'a* and the Sumerian *kiši* are obtained by Rules 10, 14b, and 1, respectively in the word initial, medial and final positions from the West-Ugric protoword.

Figure 3a shows the derivation of some words related to bread. The Hungarian words /keńér/ `bread' and the Sumerian word *gar₃* `a designation of breads' according to the Pennsylvania Sumerian Dictionary was already grouped as cognates in Revesz (2019), which predicted that more examples of the medial /ń/ will be found.

Figure 3b shows the derivation of the Ugric words in entry 2536. This example is interesting because the /ń/ results medially in the diphthong /ue/ by Rule 14b. This diphthong is resolved by a /b/ insertion following Rule 15. In the previous example, the application of Rule 14b resulted only in /ee/, which is easy to resolve as either a short or long /e/ without the use of Rule 15.

Figure 4a shows the derivation of the Ugric words in entry 1710. This example shows the Sumerian of a word final /ń/ consonant. On the other hand, it can also change in Sumerian into an /n/. Hence both Sumerian /mi/ and /min/ occur simultaneously.

Figure 4b shows the derivation of some Finno-Ugric words that mean `to calm' in entry 2514 together with the Hungarian *tunya* `sluggish.' Presumably, */tuń/ also originally meant 'to calm' because the *-a* is an adjective forming suffix. Following Rule 14b, Sumerian again shows two cognates, one with an of the word final /ń/ and another with a change to /n/.

Finally, Figure 4c shows the derivation of Ugric words in entry 2856. This is a rare example of a word initial /ń/. Here the assumed Ugric protoword is */ńuj/. In the first step, there is a word final /j/ by Rule 16, resulting in a West-Ugric protoword */nu/. To that protoword a frequentative suffix /š/ seems have been

added already in the West-Ugric period. This was changed into Hungarian *nyúz* and Sumerian *uš* by Rule 1. The of word initial /ń/ seems to be a regular sound change in Sumerian, as can be seen in entry 714 with Hungarian *nyöszörög* /ńösörög/ and Sumerian *eš₉*.

Another interesting regular sound change is Rule 15, which can change Proto-Ugric /w/ to Sumerian /m/. This occurs mostly in word initial positions as shown by the following examples based on the EDSL entries 1005, 1659 and 1751, respectively:

PU */wala/ > H /való/ ~ K /wal/ ~ M /al/ ~ Su /mal/ (Rules 15 and 11)

PU */wača/ > H /üső/ ~ K /wati/ ~ M /wasi-g/ ~ Su /maš/ (Rules 15 and 2)

PU */wul/ > H /al/ ~ K /wol/ ~ M /wol/ ~ Su /mul/ (Rules 15 and 11)

All the above examples the protoword had a back vowel following the word initial /w/. In case of the protoword has a word initial /w/ followed by a front vowel, Sumerian has an elision of /w/ by Rule 15 as shown by the following examples based on the EDSL entries 56, 715, 2627, 2823, 2860 and 2886, respectively:

PU */wäŋa/ > H /ék/ ~ K /aŋ/ ~ M /woŋḫ-ap/ ~ Su /aga/ (Rules 15 and 14c)

PU */weti/ > H /víz/ ~ K /wot/ ~ M /wüt/ ~ Su /eš/ (Rules 15 and 5)

PU */wäje/ > H /vaj/ ~ K /woj/ ~ M /waj/ ~ Su /u/ (Rules 15 and 16)

PU */weri/ > H /vér/ ~ K /wer/ ~ M /wür/ ~ Su /uri/ (Rules 15 and 12)

PU */wäśə/ > H /vés/ ~ K /woš/ ~ M /us/ ~ Su /uš/ (Rules 15 and 3)

PU */wäč/ > H /vöčök/ ~ K /wasi/ ~ M /was/ ~ Su /uz/ (Rules 15 and 2)

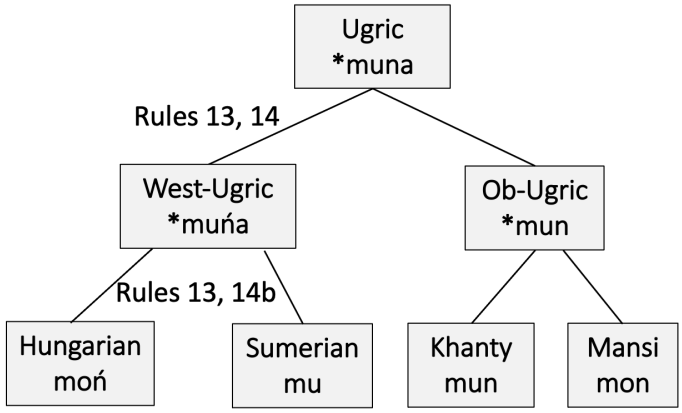
In the third example above, we chose the protoword */wäje/ because of several Finno-Ugric cognates with front vowels such as Komi *vi*, Mari *üj*, and Udmurt *vöj*. These examples do not support the UEW's reconstruction of the Finno-Ugric protoword for the cognates in this group as **woje* with a front-syllabic back vowel. We believe that **wäje* is a better supported reconstruction.

The elision of word final /w/ is also common in Sumerian as shown by the following examples based on the EDSL entries 346, 1163, 1928, 2059 and 2402, respectively:

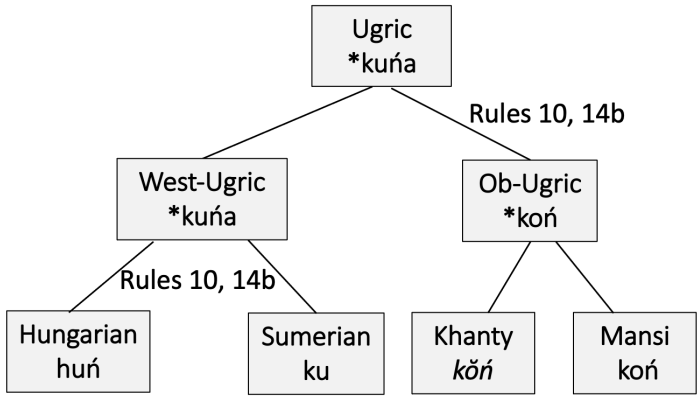
PU */puw/ > H /fú/ ~ K /puw/ ~ M /pu/ ~ Su /bu/ (Rules 8 and 15)

PU */kuḫuw/ > H /köhög/ ~ K /ḫut/ ~ M /kakwi/ ~ Su /ḫuḫu/ (Rules 10, 9, 15)

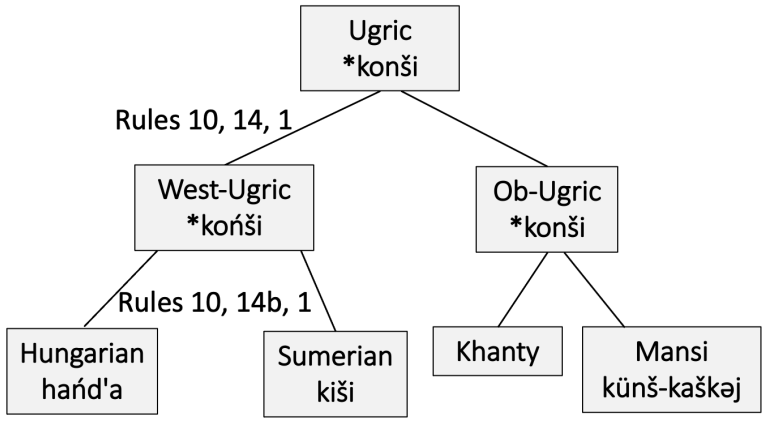
PU */fuw/ > H /fúv/ ~ K /paj/ ~ M /pat/ ~ Su /pa/ (Rules 8 and 15)



(a) Ugric words in entry 1726.

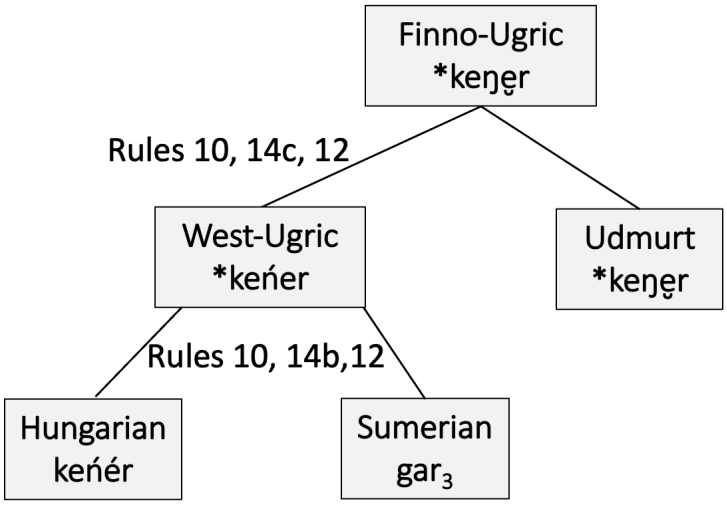


*(b) Sumerian word ku in entry 1421 and UEW's Ugric entries related to *kuńa.*

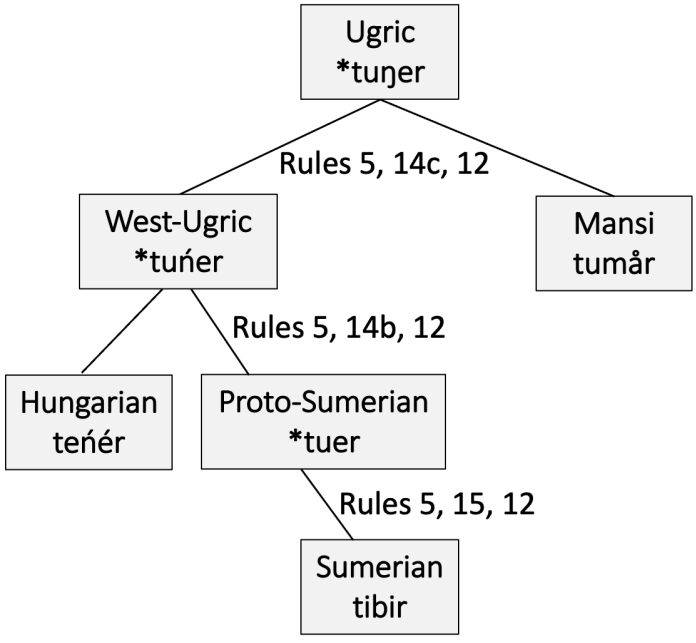


(c) Ugric words in entry 1417 and UEW's Mansi künš-kaškej 'ant'.

FIGURE. 2 *Regular sound change Rule 14b with Sumerian of medial /ń/.*

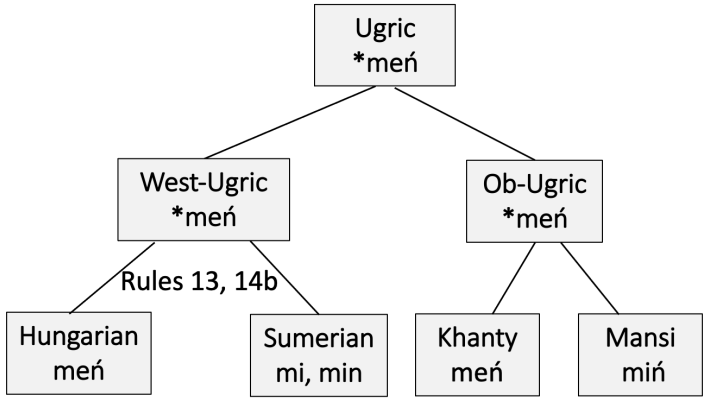


(a) Some words related to bread that were *listed as cognates in Revesz (2019).*

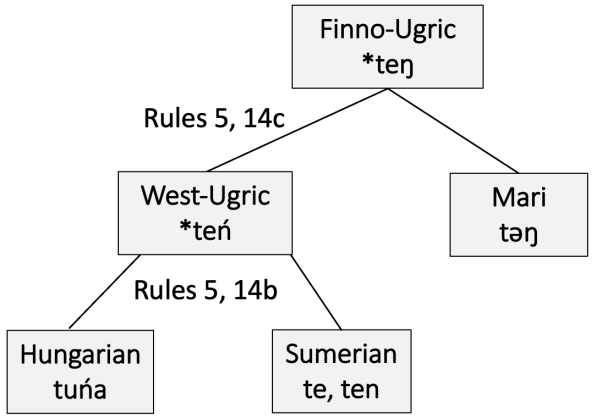


(b) Ugric words in entry 2536.

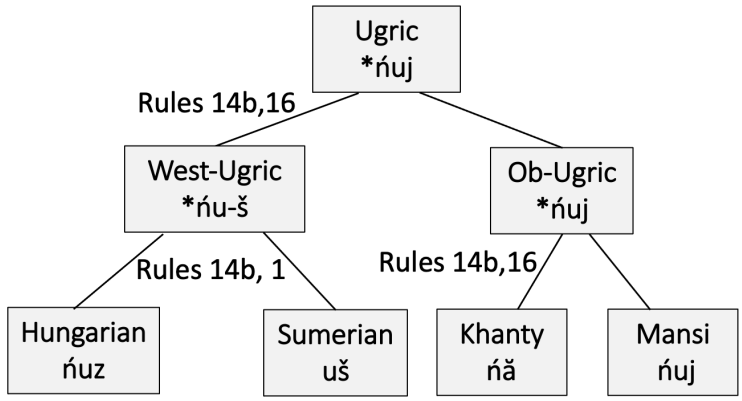
FIGURE. 3 *Regular sound change Rule 14b with Sumerian of medial /ń/ where there is a following /r/ phoneme.*



(a) Ugric words in entry 1710.



(b) Some words related in entry 2514 and Hungarian tunya 'sluggish.'



(c) Ugric words in entry 2856.

FIGURE. 4. *Regular sound change Rule 14b applied to word final and word intitial /ń/..*

PU */ śäw/ > H /ső/ ~ K /sew/ ~ M /sä/ ~ Su /sa/ (Rules 3 and 15)

PU */ šew/ > H /ev/ ~ K /te/ ~ Su /šu/ (Rules 1 and 15)

The word medial /w/ can change in Sumerian to /b/ or /m/ as shown by the following examples based on the EDSL entries 1383, 2155 and 2213, respectively:

PU */kiwe/ > H /hév/ ~ Su /kim/ (Rules 10 and 15)

PU */śäŋwa/ > H /šeg/ ~ M /saŋw/ ~ Su /sigba/ (Rules 3, 14c and 15)

PU */śiwe/ > H /sív/ ~ K /saw/ ~ Su /sub/ (Rules 3 and 15)

As the above examples show, there is a greater regularity in the sound changes than is apparent from Table 3. In other words, not all the regularity that we noticed can be crammed into Table 3, and we could express only by giving some concrete examples. This incompleteness is especially true for the medial consonant clusters because there are numerous cases of those. We plan to continue to explore the subtle interactions of the medial consonant clusters in the future.

4. The Need for Ugric Protoword Reconstructions

It is important to consider the issue of reconstructing the Ugric protowords. Earlier dictionaries concentrated on the task of reconstructing Finno-Ugric protowords, which are often different because Proto-Ugric developed further after it separated from Proto-Finno-Permic. For example, many researchers reconstruct a /ć/ phoneme, that is a palatized voiced dental non-sibilant affricate, for the Proto-Finno-Ugric language. However, such a phoneme does not exist in any of the Ugric languages. Hence it seems likely that the /ć/ phoneme has disappeared in Proto-Ugric, simplifying to either a /ś/ phoneme or some other sibilant depending on the word context. Therefore, regular sound change rules that are specific to the Ugric languages should not include the /ć/ sound. Some other phonemes that may have been present in Proto-Finno-Ugric but were eliminated in Proto-Ugric are /ḱ/ and /k^{w}/. Although some traces of these consonants may have been left in a few dialects of Khanty, we cannot be sure that those words not borrowed from Samoyedic languages. Hence our regular sound change rules do not include these consonants.

Since several Proto-Finno-Ugric phonemes have been eliminated from Proto-Ugric, it may seem that the set of Ugric regular sound changes is simpler than the set of Finno-Ugric regular sound changes. However, that is not the case because at the same time new phonemes were introduced. For example, the /h/ phoneme seems to have been introduced in Proto-Ugric because it occurs in all the Ugric languges. Therefore, now we need specific regular sound chage rules that involve a Proto-Ugric /ḫ/ phoneme. This is accomplished by Rule 4 in Table 3.

In addition, some phonemes which either did not or rarely appeared in the word initial position, started to be more frequent. This seems true for the /r/ phoneme as shown in Table 4. The Ugric vocabulary growth added 29.4 percent new /r/ initial words, while only 18.5 percent new /t/ initial words according to the *Uralisches etymologisches Wörterbuch* (UEW) of Rédei (1988), which is considered a major source on protowords within Finno-Ugric languages.

The above suggests that a relatively large ratio of /r/ to /t/ initial phoneme words in a vocabulary is a feature of the Ugric branch within the Finno-Ugric languages. According to Table 4, the /r/ to /t/ initial phoneme ratio is 0.24 in reconstructed Proto-Uralic and Finno-Ugric words, it reaches 0.27 when we add the reconstructed Proto-Ugric words, and it is 0.31 in Sumerian. Table 4 also shows the /r/ to /m/ ratios. Both the ratios suggest that Sumerian is closer to Proto-Ugric than to Proto-Finno-Ugric. This also supports the idea that Sumerian belongs to the Ugric group of languages.

Table 3 implies that the /l/, /m/, /r/ and /t/ phonemes are fairly stable within the Ugric branch because the regular sound changes tend to preserve them. Hence their ratios changed slowly with the addition of new words and dropping of some old words rather than by a major sound shift. It is also intuitive that new word proposals using infrequent phonemes (in this case /r/) are more likely to stick than new word proposals with frequent phonemes, which may sound too similar to already existing words. This explains the gradual growth of words with /r/ word initial within the Ugric branch.

TABLE. 4. *Word initial phoneme statistics comparing the EDSL and the UEW.*

Phoneme	UEW Uralic + FU	UEW Ugric add	UEW total	EDSL Sumerian
l	40	6	46	122
m	66	17	83	186
r	17	5	22	50
t	70	13	83	162
r/t ratio	0.24		0.27	0.31
r/m ratio	0.258		0.265	0.269

As an example for the need for Ugric protoword reconstructions let us consider entry 1990 again. The EDSL gives */ped'ä/, while UEW gives */pel'a/ as the protoword for this group where the entries mean 'ear'. However, neither of these is satisfying because neither of them can derive the Hungarian word 'fül'.

The two reconstructions agree with the word initial /p/, which by the application of Rule 8 yields the required /f/. The problem is with the word medial reconstructions. The EDSL's /d'/ leads by Rule 6 to /d'/, /j/, /l'/ or /t'/, while the UEW's /l'/ leads to /l'/, /d'/ and /j/. Therefore, neither case led to the required Hungarian /l/.

In contrast, our protoword reconstruction for the 1990 entries is */pede/. Here the applying Rule 4 yields the required Hungarian /l/. In addition, the Sumerian /bed/ can be also derived by applying Rule 8 word initially and Rule 4 word medially to yield /bede/. Hence we got a perfect match for the first three

phonemes. Of course, the extra word final vowel is not a concern because it tends to be dropped in many Uralic languages.

After this motivating example for Ugric protoword reconstructions, we give actual reconstructions of a set of Ugric-related sample entries from the EDSL using Table 3's regular sound change rules for Ugric languages in the next section.

5. The Ugric-Related Data Set of EDSL

During my visit to the University of Helsinki in the Fall 2019 semester, I met Professor Simo Parpola, and I offered to test the Ugric entries in his updated *Etymological Dictionary of the Sumerian Language (EDSL)* by my regular sound change rules and work on Ugric protoword reconstructions. He kindly shared with me 210 entries from his updated dictionary. Each entry was complete in the sense that it contained a presumably cognate word fom each of Hungarian, Khanty, Mansi, and Sumerian.

This provided an opportunity to make Ugric protoword reconstructions from four Ugric languages, whereas all earlier attempts were based only on the last three languages and with considerably fewer complete entires even for those languages. For example, Honti (1997) listed only 84 entries that had cognate words from each of Hungarian, Khanty and Mansi. Subsequently, the Ugric protowords project commenced right away in the Fall of 2019 and continued after my return to the University of Nebraska-Lincoln.

While this was a great opportunity to work with 210 complete (four languages) entries instead of with only 84 complete (three languages) entries, some of the 210 entries were marked as being 'less certain' in the EDSL. Hence these 210 entries had to be verified by the Ugric regular sound change rules in Table 3 while simultaneously trying to reconstruct their protowords.

The 210 EDSL entries were rewritten into our phonetic notation as shown in columns 2-5 in Tables 5-7 and as explained in the previous section. We separated the word roots and the suffixes by hyphens because the regular sound changes are expected to explain only the changes in the word roots.

When an EDSL entry contained several cognate words in the same language, then we chose among those words the one that seemed to best fit the cognate words in the other languages. We also made some minor changes in the Hungarian words in entry 0065 where we replaced *okos* 'smart' with *okul* 'learn', in entry 0337 where we replaced *pežeg* with its synonym *buzog,* in entry 418, where we added *dadog* 'stutter, stammer, falter,' and in entry 439, where we added *türök*, which is an older form of *tülök* 'horn' (Zaicz, 2006).

Next, we examined each of the 210 entries trying to find a protoword that together with the regular sound change rules of Table 3 can explain all the Hungarian, Khanty, Mansi and Sumerian word forms. Table 5 lists those entries where we were successful in our search. In particular, for the words listed in Table 4, we give the following analysis details:

1. A hypothetical Proto-Ugric protoword, which is always preceded by a star symbol. The Proto-Ugric word is assumed to have existed for the common

proto-Ugric language that is the common ancestor of Hungarian, Khanty, Mansi and Sumerian.

2. For each Proto-Ugric protoword, we also indicate a source (So). The source may be the EDSL (E), the UEW (U) or our proposed reconstruction (#). We propose new reconstructions only when the ones given in either E or U do not fit the regular sound changes in Table 1.

3. We also list the applicable regular sound change rules. The regular sound change rules are divided into three groups.

   The first group is the word intitial (Init.) regular sound changes. These are applicable only if the Proto-Ugric word starts with a consonant. If the Proto-Ugric word starts with a vowel, then this is not applicable, which is indicated by a - symbol.

   The second group is the words medial regular sound changes. These are applicable only if the Proto-Ugric word contains a medial consonant. These are indicated by - if they are not applicable.

   The third group is the word final (Fin.) regular sound changes. These are applicable only if the Proto-Ugric word contains a final consonant. These are also indicated by - if they are not applicable.

We kept in Table 5 each entry where at least the Hungarian and the Sumerian words could be explained by the regular sound change rules. In all but eighteen cases at least one of the Ob-Ugric languages also satisfied the regular sound rule changes. Hence all of the entries show a cognate relationship between Sumerian and Hungarian and most of them also with Khanty and Mansi too. We will consider a statistical analysis in the next section. Table 5 shows in parenthesis those words that did not fit according to the regular sound change rules.

TABLE. 5. *Ugric entries that satisfy the regular sound change rules of Table 3.*

Entry No.	Sumerian	Khanty	Mansi	Hungarian	Proto-Word		Sound Change Rules		
					Form	So	Init.	Medial	Fin.
0003.	aj	jəŋk	jaŋk	jég	*jäŋe	U	16	14c	-
0028.	aba	apa	apa	apa	*apa	E	-	8	-
0039.	ad	aťi	aťa	aťa	*aťa	E	-	6	-
0055.	ag	waŋk	waŋk	vág	*wäŋ	#	15	-	14c
0056.	aga	aŋ	woŋ̮-ap	ék	*wäŋa	#	15	14c	-
0065.	aŋal	omal-əŋ	akel-əŋ	okul	*oŋal	#	-	14c, 11	-
0102.	ama	aŋka	ama	aṅa	*aŋa	#	-	14c	-
0116.	ene	əna	en	ön	*ene	#	-	14	-
0135.	ar	ari	jar	ár-t	*ar	#	16	-	12
0136.	ar	ar	jeri	ár	*ara	#	-	12	-
0161.	arz-ig	arəs	ras-ək	rož	*aruča	#	-	12, 2	-
0189.	auna	aŋ-ən	äŋ-ən	iń	*eŋe	#	-	14c	-
0217.	bala	(pal-əs)	wal	vála-s	*wala	#	15	11	-
0241.	bar	pirä	pärä	far	*perä	U	8	12	-
0242.	bar	pera	pola- ḫ	faro-k	*pärä	E	8	12	-
0254.	bar	par	par	for-gáč	*päre	U	8	12	-

0264.	bar	paj-ət	pal'	fad'	*pad'a	E	8	6	-
0266.	bar	pał	pel	fél	*pele	U	8	11	-
0269.	bar	palə-k	pal	fél	*pälä	U	8	11	-
0282.	bar	por-təm	por-t	forr	*par-t	#	8	-	12
0303.	bara	parə-ḫ	por	fura	*para	#	8	12	-
0313.	bi	päj	pöj	fő	*peje	U	8	16	-
0317.	bed	paťˀ	pot'	foš	*pät'	#	8	-	6
0319.	bid	(pet)	(put)	vet	*wet	#	15	-	5
0337.	biz	pusi	pəs-ḫt	buz-og	*pisi	E	8	1	-
0346.	bu	puw	pu	fú	*puw	#	8	-	15
0357.	buluḫ	(poləḫ-tə)	(puteg)	buďog	*pul'uk	#	8	11b	10
0361.	bun	poŋg-ət	puŋe-t	bog	*puŋa	#	8	14c	-
0362.	buŋa	pag	pug	fiú	*pojka	U	8	16b	-
0367.	bur	porki	porḫi-t	forog	*porki	#	8	12	-
0379.	buru-d	pori-t	peru-t	fúr	*pura	U	8	12	-
0409.	dag	laŋk	lok	lak	*daŋa	#	4	14c	-
0413.	dag	tag	tak	tag-ol	*taka	#	5	10	-
0418.	dag	tag	latḫ-at	dadog	*dak(dak)	#	4	(10b)	10
0425.	dal	(təgət)	til	toll	*tal	#	5	-	11
0439.	dar	tar-pi	sor-pi	tür-ök	*tar	#	5	-	12
0455.	dib	tep	läp	lép	*dapu	#	4	8	-
0475.	dili	tälä-ŋ	tal	tele	*tele	#	5	11	-
0565.	duḫ	toki	toḫ-r	dug-gat	*duki	#	4	10	-
0607.	e	i	ä	e-z	*e	#	-	-	-
0609.	e	(ej)	(ej)	hé	*ḫe	#	9	-	-
0610.	eg	jegi	je	jó	*jo-(ke)	U	16	10	-
0643.	ellaŋ	loŋk	(saŋk-w)	lád'-ék	*láŋa	#	11	14c	-
0666.	enedi	jant-əg	jänn	ját-sik	*jänti	#	16	14d	-
0686.	ere	irə-t	ərə-t	ere-d	*ere	#	-	12	-
0701.	era	jarə-ŋ	jorä-ŋ	erő-š	*jera	#	16	12	-
0715.	eš	wot	wüt	víz	*wis	#	15	-	1
0767.	gazum	kočəm	kwäššöm	köč-ög	*köč-əm	#	10	15, 2	13
0777.	gazi	kosi	kosi	kese-g	*kesi	#	10	1	-
0785.	gena	kena	kinä	keńe-š	*kene	#	10	14	-
0835.	gir	kir	ker	ker	*kere	#	10	12	-
0836.	gir	kar	kär	here	*kara	#	10	12	-
0855.	gun	kun	koan	hón	*kon	#	10	-	14
0866.	gubud	käpät	(kapal)	követ	*käpät	#	10	8	5
0875.	gud	ḫot	(kwäl)	ház	*kota	U	10	5	-
0919.	gul	ḫał	kol	hal	*kola	U	10	11	-
0933.	gum	kom-taḫ	ḫom	hom-lok	*kuma	U	10	13	-
0941.	gun	ḫanti	kont	had	*kun-(ta)	U	10	14(d)	-
1001.	ŋa	mä	äm	eng-em	*eŋe	#	-	14c	-
1005.	mal	wal	al	való	*wala	#	15	11	-
1033.	men	men	men	men	*mene	U	13	14	-
1045.	geš	ḫot-tə	ḫut	híz-ik	*katə	U	10	5	-
1109.	ḫal	kaj	kul'	haďˀ	*ḫal'	#	10	-	11b
1129.	ḫaš	kos	ḫus-at	haš-ad	*ḫaś	#	9	-	3
1163.	ḫuḫu	(ḫut)	kakwi	köhö-g	*kuḫuw	#	10	9	15
1237.	uli	oli	(uľ)	olaj	*olij	#	11	-	16
1245.	imi	(sawi)	(jəm)	homo-k	*komi	#	10	13	-
1290.	ka	kej	käj, kij	kíj-ó	*kij	#	10	-	16
1300.	kad	kat, kät	kit	két	*ket	#	10	-	5
1339.	kiri	kärə-t	(kwärä-kt)	kerí-t	*kerä	U	10	12	-
1343.	kas	kasi	ḫus	hős	*kaśi	#	10	3	-
1349.	kaš	kos	kuš	húďˀ	*kučə	#	10	-	2
1351.	kaš	kas	käs	küz-d	*käśə	#	10	-	3
1353.	keš	(kač)	kät	köt	*kät	#	10	-	5
1383.	kim	(kajmə)	(kim)	hév	*kiwe	#	10	15	-
1386.	kin	kün	ḫün	kín	*kin	#	10	-	14
1400.	kir	käri	kir	keri-ng	*kerä	U	10	12	-

1417.	kiši	(kačŋi)	(ḫoši)	hańd'a	*konši	#	See Figure 2c.		-
1444.	kul	kol	ḫul	hall	*kule	U	10	11	-
1446.	kul	ḫuli	(ḫoľg)	hal-ad	*kulə	U	10	-	11
1561.	leum	(lewḫ)	(leŋkw)	lem-ez	*lemu	#	11	-	13
1574.	lil	lil	lili	léle-k	*lili	#	11	11	
1600.	lug	ľok	(loḫ)	ľuk	*l'uk	#	11b	-	10
1604.	luḫ	logi	(low-t)	lúg	*luk	#	11	-	10
1612.	lum	lum	lum	láng-ol	*luŋ	#	11	-	14c
1645.	mar	marə-t	marə-ḫt	már-t	*marə	U	13	-	12
1659.	maš	wati	wasi-g	üső	*wača	#	15	2	-
1669.	men-de	meŋ	men	mik	*miŋ	#	13	-	14c
1670.	me	mej	me	mi	*məj	U	13	-	16
1710.	min	meń	miń	meń	*mińä	U	See Figure 4a		-
1720.	mu	(mi)	mig	meg-ad	*mek	#	13	-	10
1726.	mu	mun	mon	moń	*muna	U	See Figure 2a.		-
1751.	mul	wol	wol	al	*wul	#	15	11	-
1775.	mur	mori	mur	mar	*mura	U	13	12	-
1777.	mur	mar-əŋ	mir-iŋ	mor-aj	*mura	U	13	12	-
1778.	mur	mor-əg	mur	mor-og	*moro	E	13	12	-
1787.	murra	mar-dat	mor-t	mér	*merə	U	13	12	-
1824.	nam	nem	nem	nem	*nem	E	14	-	13
1825.	nam	nam	nam	név	*nime	U	14	13	-
1882.	nin	nen	(ne)	néni	*neni	#	14	14	-
1922.	pa	(ponək)	(päŋk)	fej	*päj	#	8	-	16
1923.	pa	päk	pek	fék	*päke	#	8	10	-
1928.	pa	(paj)	(pat)	fúv	*fuw	#	8		15
1931.	pad	pul	pul	fal-at	*pada	E	8	4	-
1936.	pad	put	pot	faz-ék	*pata	U	8	5	-
1941.	pag	pəg-at	pug	fog	*påkå	E	8	10	-
1952.	pan	pan	pon	fon-at	*puna	U	8	14	-
1953.	pan	pan	pon	fon	*puna	U	8	14	-
1990.	bed	pət	pel	fül	*pede	#	8	4	-
2002.	puta	pətta	(pol)	buta	*pättä	E	8	5	-
2050.	rugu	rag	rog	rok-on	*raka	#	12	10	-
2059.	sa	sew	sä-	ső	*šäw	#	3	-	15
2067.	sad	sot-om	(sor)	čata	*čatta	#	2	5	-
2070.	saŋ	soŋk	(saŋkw)	šág	*čaŋ	#	2	-	14c
2072.	saŋ	(jäŋk)	seŋk	seg	*šäŋ	#	3	-	14c
2073.	saŋ	tuŋk	(son)	šenki	*čeŋe	#	2	14c	-
2077.	saŋi	(sagən)	(senkve)	šüke-t	*šeŋe	#	2	14c	-
2097b	saŋuš	(ńawiš)	(šonöŋš)	sokáš	*šaŋuš	#	3	14c	1
2100.	sam	sam	sam	sám	*šam	#	3	-	13
2141.	si	səj	soj	zaj	*šije	#	3	16	-
2155.	sigba	(sojga)	saŋw	šeg	*šäŋwa	#	3	14c, 15	-
2164.	sil	sil	sil	sel	*šälä	#	3	11	-
2168.	sila	sal	süla	silá-nk	*šila	#	3	11	-
2213.	sub	saw	(säkw)	sív	*šiwe	#	3	-	15
2226.	sugu	soḫ	süg	sűk	*šuku	#	3	10	-
2238.	suḫur	sogər	sukər	šügér	*šuker	#	3	10	12
2239.	suḫur	šekare	saḫar	čokor	*čukur	#	2	10	12
2252.	sul	sel	sel	sél	*šele	#	3	11	-
2281.	sur	sori	čor-ḫ	čor-og	*čor	#	2	-	12
2282.	sura	sor	soro	šör	*šura	#	3	12	-
2310.	šara	sor	sur	sára-d	*šara	#	3	12	-
2340.	šeŋ	soŋgi	(saḫ)	ég	*šeŋi	#	1	14c	-
2357b	sir	sir	sir	ser	*šir	#	3	-	12
2371.	šeš	(jeji)	šeš	öč	*šeč	#	1	-	2
2394.	šita	sottə	sot	čat-ol	*čotta	#	2	5	-
2399.	šu	(tit)	(tet)	ujj	*šuj	#	1	-	16
2402.	šu	te	(teg)	ev	*šew	#	1	-	15
2403.	šu	tuj	tul	ujj	*šuj	#	1	-	16

2426.	šukur	(cogör)	(šoḫura)	tőr	*tukur	#	5	10	12
2446.	šuru	soṛ	sor	sar	*śarə	#	3	12	-
2448.	šuruš	(särij)	šerit	čerez	*čerit	#	2	12	5
2466.	tab	tap	tap	tap-od	*tap	#	5	-	8
2481.	tagga	teutem	(takkum)	tetű	*tekte	#	5	10b	-
2494.	tara	tarög	torg-	di-dereg	*terek	#	5	12	10
2498.	tar	(sart)	(surkät)	tár	*tar	#	5	-	12
2512.	te	(təḫ)	lüg	tő	*te	#	5	-	-
2627.	u	woj	waj	vaj	*wäje	#	15	16	-
2796.	ura	uraḫ	wureḫ	ara-t	*wurak	#	16	12	10
2811.	uraš	(waröš)	(kwores)	őrš	*uraš	#	-	12	1
2823.	uri	wer	wür	vér	*weri	U	15	12	-
2860.	uš	woš	us	vés	*wäśə	#	15	3	-
2866.	uš	oš	us	és	*äś	#	-	-	3
2886.	uz	wasi	was	vöč-ök	*wäč	#	15	-	2c
2915.	zal	soli	soll	sól	*śoli	#	3	11	-
2924.	zal	čał	šäľ	šaj-nál	*čal’	#	2	-	11b
2941.	zara	sari	sir-ḫ	čir	*čari	#	2	12	-
2943.	zebba	siw	tep	epe	*šäppä	#	1	8	-
2961.	zid	sat-at	šit	sít	*śit	#	3	-	5
3025.	zur	sari	šarr	šír	*śari	#	3	12	-

When in an entry the Hungarian and Sumerian words satisfy the regular sound change rules but neither the Khanty nor the Mansi satisfies them, then we can suspect that the word is West-Ugric and belongs to a lower level in the language family tree.

Some comments need to be made regarding the words in Table 5. First, it seems that some parenthesized words in some entries may be cognates. The explanation here is that regular sound change rules are one-step rules, but a chain of applications of these rules may indeed lead to a derived word. This may seem as a limitation of the regular sound change rules, but it is not a true limitation because each application of a regular sound change rule implies a certain period of time. Hence words that would need several rules for their derivations are further apart than those that would need the application of only a single rule.

As the examples in Figure 2 show, the number of derivation steps measure distances of relationships. If we ignore the number of steps in the derivations, then words closer to the root cannot be distinguished from those that are further from the root in the derivation tree. In other words, we already know that these set of languages are related and the words are likely cognates, but we also would like to know the structures of the derivation trees.

In entry 2512, the Sumerian word has in the EDSL a longer form *ten,* and it means ’basis, base, stalk of a plant.’ Therefore, this Sumerian word may be cognate with the Hungarian word *tengely* meaning ’axis’ because an axis is the base of a wheel.

Finally, Table 6 lists those entries where the words do not satisfy the regular sound change rules.

TABLE. 6 *Ugric entries not satisfying the regular sound change rules of Table 3.*

Entry	Sumerian	Khanty	Mansi	Hungarian	Comment
0012.	a	jəs	jos, jəs	őš	Sumerian misses word final fricative.
0015.	a	jas	jos	íz	Sumerian misses word final fricative.
0213.	bala	pälit	pelt	vált	Sumerian misses word final t or d.
0271.	bar	wort	vort	párt	Sumerian misses word final t or d.

0554.	dugbad	soḫmat	lakwäl	sökken	
0590.	dungu	čug	seŋko	tömeg	
0738.	gake	kagi	kökeḫ	kakukk	
0779.	gi	ki, kit	kit	küld	Sumerian misses word final d.
0901.	guŋer	ḫoŋgar-eŋ	kiwr-eŋ	homor-ú	
0939.	gun	künt	kont	gond	Sumerian misses word final t or d.
1029.	ŋi	jij	ji	éj	
1043.	ŋeš	ket	kät	kéz	
1051.	ŋeš	pis	jes	és	
1054.	ŋeš	još	jäš	jeď	
1056.	ŋeš	ḫüs	vesi	fas	
1058.	ŋeš	jäš	jäš	véš	
1120.	ḫar	tar	šär	sőr	
1389.	kin	känš	kins	vonz	Sumerian misses word final fricative.
1459.	kuniga	kunoḫ	ḫon	kán	
1465.	kur	kur	kur	kór	
1485.	kuš	koč, koť	kwot, koät	héj	
1545.	lal	lawḫ	low	lov	
1641.	manukal	mäŋoko	meŋukə	manó	
1852.	niga	ńaťga	laška	laššú	
1943.	pag	päk	paka	fáj	
1950.	pal	pala	pal	pad	
1956.	pap	pop	pup	pap	
2008.	ra	rog	rag	rod'	Sumerian misses a word final consonant.
2026.	rig	rig, riw	ragg, rap	rep-ül	
2054b	sa	saw	saw	haŋ	Sumerian misses a word final consonant.
2071.	sag	sak	sag	haj	
2156.	siga	sikko-ŋ	šińə-ŋ	čino-š	
2223.	sug	tog, toḫ	to	tó	
2250.	sul	sel	šilt	žír	
2251.	sul	sol	sol-val	šó	Wanderword: English *salt*, etc.
2319.	šed	togət	seť	sövet	
2332.	šedgi	soləḫ	serg	siseg	
2373.	šeš	söť	šes	šúr-ol	
2500.	tar	tarə, ŧarə-ḫ	tori-ḫul	šér-inc	
2549.	tila	ťali, čali	taľeḫ	cél	
2582.	tuk	čok, šuk	šöw	ńűg	
2673.	udug	utəm, uləm	ulem	álom	
2680.	udug	adak	üösḫ	ördög	
2692.	uga	jaug	soj	čuka	
2696.	ugu, agu	uḫ, og	uḫ, aḫ	ad'	
2706.	ugudig	ogöntiw	ontep	od'velő	
2783.	ur	jor, ŧor	jekwer	d'ökér	
2826.	urta	wort	urt	őr	
2895.	zag	sagə, sigə	saḫi	síj	
2954.	zi	ziw	tou	sel	
2968.	zig	säw	sow	suv	

In Table 6, we gave comments regarding those cases where the Sumerian word misses some required ending. If these words are still cognates, then some explanation needs to be given about why they apparently miss the final consonants. One possible reason for missing many different final consonants may be due to the preference for vowel endings by the Akkadian speakers who adopted the Sumerian words. It seems less likely that the simultaneous omission of all types of final consonants was a natural internal development of the Sumerian language.

6. Statistical Analysis

For the 210 entries of the EDSL, we could find 159 Ugric protowords that satify the regular sound change rules of Table 3 for Sumerian and Hungarian and all but eighteen also simultaneously satisfied either Khanty or Mansi. Hence our analyis verified that **159 out of 210 or 75.7 percent of the considered EDSL entires contained Sumerian-Ugric cognate words**. Since 159 Sumerian words were shown to be cognates with Ugric words by the use of a conservative set of regular sound change rules, it is now beyond a reasonable doubt that Sumerian is an Ugric language.

In addition, nine more Sumerian words could have a cognate origin, but currently their status is unclear because they seem to miss some final consonant. Perhaps the complete forms of these words could be found in the earliest Sumerian texts, and their status could be reinvestigated afterwards.

The cognate nature of some other words may have complex explanations that are beyond the reach of our set of regular sound change rules. If no regular sound change rule can be found to apply, then it does not mean that the words in that entry are not cognate. It may turn out that another regular sound change rule is needed to be added to our list. However, adding more rules is warranted only if several occurrences of the same type of change are found in the same context and without a conflict with the existing rules. We refrained from adding rules where there was not a seeming regularity that was attested by several plausible examples. Since we worked with small sets of entries both in this appendix and in our earlier publications (Revesz 2019, 2019b), we may have overlooked some infrequent but regular sound changes.

We hope that further investigations will not only extend the regular sound change rules in Table 3 but also will clarify the placement of Sumerian within the Ugric branch. We presented arguments for classifying Sumerian as a West-Ugric language instead of an Ob-Ugric language. However, this is a research topic that needs more investigation.

It bears mentioning that Simo Parpola's idea was to look at all possible dictionaries to try to collect all possible cognates that looked to him cognates. His work culminates earlier efforts to collect evidence of the Sumerian language's Finno-Ugric, Ugric or Hungarian relationship, including earlier dictionaries and grammatical analyses by Badiny (1974), Bobula (1951), Csőke (1974), Gosztony (1975), Götz (1989), Varga (1942), and Zakar (1971), to give only an incomplete list.

After this wide and quixotic search of many years, it is greatly satisfying that many of 159 words that we could finally verify are also listed as Uralic, Finno-Ugric or Ugric-origin words by traditional Uralic linguists. This actually can be a cause of celebration among the Uralic linguists because it also verifies that their search for cognate words within the already accepted Uralic languages was fairly complete. In fact, out of the 159 verified cognates, there are 33 new cognates (see Table 7) that are not listed in the UEW (Rédei 1988).

TABLE. 7. *New Ugric cognates verified by the regular sound change rules of Table 3.*

Entry No.	Sumerian	Khanty	Mansi	Hungarian
0065.	aŋal	omal-əŋ	akel-əŋ	okul
0161.	arz-ig	arəs	ras-ək	rož
0357.	buluh	(poləh-tə)	(puteg)	buďog
0418.	dag	tag	lath-at	dadog
0439.	dar	tar-pi	sor-pi	tür-ök
0609.	e	(ej)	(ej)	hé
0701.	era	jarə-ŋ	jorä-ŋ	erő-š
0785.	gena	kena	kinä	keńe-š
1163.	huhu	(hut)	kakwi	köhö-g
1237.	uli	oli	(uľ)	olaj
1245.	imi	(sawi)	(jəm)	homo-k
1383.	kim	(kajmə)	(kim)	hév
1386.	kin	kün	hün	kín
1561.	leum	(lewh)	(leŋkw)	lem-ez
1604.	luh	logi	(low-t)	lúg
1777.	mur	mar-əŋ	mir-iŋ	mor-aj
1778.	mur	mor-əg	mur	mor-og
2002.	puta	pətta	(pol)	buta
2067.	sad	sot-om	(sor)	čata
2077.	saŋi	(sagən)	(senkve)	šüke-t
2097b	saŋuš	(ńawiš)	(šonöŋš)	sokáš
2100.	sam	sam	sam	sám
2226.	sugu	soh	süg	sűk
2238.	suhur	sogər	sukər	šügér
2282.	sura	sor	soro	šör
2394.	šita	sottə	sot	čat-ol
2448.	šuruš	(särij)	šerit	čerez
2466.	tab	tap	tap	tap-od
2494.	tara	tarög	torg-	di-dereg
2498.	tar	(sart)	(surkät)	tár
2811.	uraš	(waröš)	(kwores)	őrš
2924.	zal	čał	šäľ	šaj-nál
2941.	zara	sari	sir-h	čir

These new cognates can tell more about the Proto-Ugric and the Proto-West-Ugric stage of languages as well as cultural development. Some of these are onomatopoeic words that are routinely ignored in comparisons, but it is an interesting question whether some onomatopoeic words, such as those that describe how the animals around us are presumed to speak, actually reflect something about ourselves. However, probably the greatest debate will be generated by cultural words like Sumerian *arzig, saŋuš, sad, sam, sura* and *uraš*. While the avoidance of onomatopoeic words is understandable, the avoidance of words reflecting a higher-than-presumed cultural development may be unjustified. Now at least a serious debate can begin about whether these are native developments or wanderwords. The Sumerian words *suhur* and *dar* are notable too because their cognates in the other languages suggest that the Proto-Ugric homeland was in an area where these animals coexisted before the rise of the Sumerian civilization.

Some of the words are classified as borrowed words from Turkic and other languages. The presumed direction of borrowings was already questioned by some linguists, for example by Marácz (2018), and these cognate words will surely add material to that continuing debate. In all the heated discussion about

loanwords, let us keep in mind the adage that ”the wiser is not the one who teaches, but the one who learns.”

7. New Horizons of Uralic Studies

Simo Parpola’s work opens a new horizon in Uralic studies. The most important implication of the family tree drawn in Figure 1 is that the hypothetical age of the entire family tree needs to be shifted earlier in every branch. Most scholars used to doubt that the Uralic language family existed during the Ice Age, but in light of the Sumerian connection and the necessary extension of the evolutionary history of the Uralic language family, I now believe that it is possible that the history of the Uralic language family stretches back to the Ice Age. That possibility raises some questions about the history of the Uralic language family that need to be investigated by a multidisciplinary collaboration of linguists, archaeologists, art historians, paleobotanists, anthropologists and archaeogenetic researchers.

Earlier in this volume, Simo Parpola listed all the archaeological cultures that are usually associated with Uralic language speakers. Hajdú (1964) thought to have located the Uralic homeland near the Ural Mountains based on cognate plant names in various Uralic languages, but he did not account for paleobotanic changes on a large time scale going back to the Ice Age. Janhunen (2009) presented an interesting Uralic family tree where the Mansic group, which is composed of Hungarian and Mansi, branched off from the remaining Proto-Finno-Ugric group as early as 2500 BC. However, in light of Sumerian being an Ugric language, even Janhunen’s estimate is too late.

Wiik (1997) and Makkay (2003) argued that the Uralic language speakers migrated north from the Black Sea to the Baltic area following the direction of rivers. Based on art motif comparisons, Revesz (2019c) argues that the migration was not only northward but also southward towards the Aegean area, including Crete. In fact, some of the Minoan texts can be read as a Proto-Ugric language (Revesz 2016b, 2016c, 2017, 2017b).

Of course, that ancient migration may also have led to the Ural Mountains where the above mentioned Uralic-associated archaeological cultures were established.

8. Conclusions

Revolutions, unlike sound changes, are not regular occurrences in linguistics. The addition of Sumerian to the Uralic language family is a revolutionary development. It raises a tsunami of exciting questions for computational and comparative linguists, whether they specialize in morphology or phonology, and for assyriologists and sumerologists, who can take a deeper look at some cuneiform texts that were previously incomprehensible and now translate them with the help of the EDSL. To all of these researchers, my advice is: *Carpe diem!*

References

Badiny, Ferenc J., 1974. *The Sumerian Wonder*, School for Oriental Studies of the University of Salvador, Buenos Aires.

Bobula, Ida 1951. Sumerian affiliations; A plea for reconsideration, manuscript, Washington D.C., USA.

Cser, Ferenc and Darai, Lajos, 2005. *Magyar folytonosság a Kárpát-medencében*, Budapest: Frig.

Csőke, Sándor 1974. *Sumér-Finn-Mongol-Török Összehasonlító Nyelvtan*, Buenos Aires: Magyar Őskutatás Publishing.

Csúcs, Sándor 2019. *Miért Finnugor Nyelv a Magyar?* Budapest: Tinta Publishing.

Daggumati, Shruti and Revesz, Peter Z. 2021. A method of identifying allographs in undeciphered scripts and its application to the Indus Valley Script, *Humanities and Social Sciences Communications*, 8, 50.

Gosztony, Kálmán 1975. *Dictionaire D'etymologie Sumerienne et Grammaire*, Paris: De Boccard.

Götz, László 1989. *Keleten Kél a Nap (The Sun Rises on the East)*. Budapest: Püski Publ.

Hajdú, Péter 1964. Über die alten Siedlungsraume der uralischen Sprachfamilie. *Acta Linguistica Hungaricae*, 14: 47–83.

Honkola, T.; Vesakoski, O.; Korhonen, K.; Lehtinen, J.; Syrjänen, K.; Wahlberg, N. 2013. Cultural and climatic changes shape the evolutionary history of the Uralic languages, *Journal of Evolutionary Biology*, 26, 6: 1244–1253.

Honti, László 1997. *Az Ugor alapnyelv kérdéséhez, (On the question of the Proto-Ugric language)*, Budapest: KEL Print.

Honti, László 1979. Characteristic features of Ugric languages (observations on the question of Ugric unity), *Acta Linguistica Academiae Scientiarum Hungaricae*, 29, 1/2: 1-26.

Janhunen, Juha. 2009. Proto-Uralic–what, where, and when. *The quasquicentennial of the Finno-Ugrian society* 258: 57-78.

Kiss, Jenő and Pusztai, Ferenc. eds., 2018. *A Magyar Nyelvtörténet Kézikönyve (Handbook of Hungarian Language History)*, Budapest: Tinta Publishing.

Krantz, Grover 1988. *Geographical Development of European Languages,* NY: P. Lang.

Makkay, János 2003. An archeologist speculates on the origin of the Finno-Ugrians, *Mankind Quarterly* 43, 3: 235-272.

Marácz László, 2018. Revisiting the theory of the Hungarian vs. Chuvash lexical parallels, In: A. Marcantonio ed., The State of the art of Uralic studies: Tradition vs. innovation.

Nagy, P. L. Olasz, J. Neparáczki, E. et al. 2021. Determination of the phylogenetic origins of the Árpád Dynasty based on Y chromosome sequencing of Béla the Third. *European Journal of Human Genetics* 29: 164-172.

Pennsylvania Sumerian Dictionary (PSD), 2021. Available: http://psd.museum.upenn.edu

Rédei, Károly editor, 1988. *Uralisches etymologisches Wörterbuch*, Budapest: Akadémiai Kiadó.

Revesz, Peter Z. 2016. Bioinformatics evolutionary tree algorithms reveal the history of the Cretan Script Family, *International Journal of Applied Mathematics and Informatics*, 10, 1: 67-76.

— 2016b. A computer-aided translation of the Phaistos Disk, *International Journal of Computers*, 10: 94-100.

— 2016c A computer-aided translation of the Cretan Hieroglyph Script, *International Journal of Signal Processing*, 1: 127-133.
— 2017. Establishing the West-Ugric language family with Minoan, Hattic and Hungarian by a decipherment of Linear A, *WSEAS Transactions on Information Science and Applications*, 14: 306-335.
— 2017b. A Translation of the Arkalochori Axe and the Malia Altar Stone, *WSEAS Trans. on Information Science and Applications*, 14, 1: 124-133.
— 2019. Sumerian contains Dravidian and Uralic substrates associated with the Emegir and Emesal dialects, *WSEAS Transactions on Information Science and Applications*, 16: 8-30.
— 2019b. Using data mining algorithms to discover regular sound changes among languages, *Proceedings of the 23rd International Conference on Circuits, Systems, Communications and Computers*, MATEC Web of Conferences, **292**, no. 03018,
— 2019c. Art motif similarity measure analysis: Fertile Crescent, Old European, Scythian and Hungarian elements in Minoan culture, *WSEAS Transactions on Mathematics*, 18: 264-287.
— 2019d. Minoan archaeogenetic data mining reveals Danube Basin and western Black Sea littoral origin. *International Journal of Biology and Biomedical Engineering*, 13: 108-120.
— 2020. Minoan and Finno-Ugric regular sound changes discovered by data mining, *Proceedings of the 24th International Conference on Circuits, Systems, Communications and Computers*, IEEE Press, 241-246.
— 2020b. A vowel harmony testing algorithm to aid in ancient script decipherment, *Proceedings of the 24th International Conference on Circuits, Systems, Communications and Computers*, IEEE Press, 35-38.
— 2021. Data mining autosomal archaeogenetic data to determine Minoan origins, *Proceedings of the 25th International Database Engineering and Applications Symposium*, ACM Press, 46-55.

Szabó, István Mihály 2004. A magyar nép eredete – Az uráli népek eurázsiai-amerikai őstörténete, Budapest: Mudus Publisher.

Varga, Zsigmond 1942. *Ötezer Év Távlatából*, reprinted 2020, Budapest: Gede Testvérek.

Wiik, Kalevi 1997. The Uralic and Finno-Ugric phonetic substratum in Proto-Germanic, *Linguistica Uralica* 33, 4: 258-280.

Zaicz, Gábor, editor, 2006. *Etimológiai Szótár: Magyar Szavak és Toldalékok Eredete, (Etymological Dictionary: Origin of Hungarian Words and Affixes)*, Budapest: Tinta Publishing.

Zakar, András 1971. Sumerian-Ural-Altaic Affinities, *Current Anthropology*, 12, 2: 215-216.