Utilization of Nganasan digital resources:
a statistical approach to vowel harmony

Fejes, László
Hungarian Academy of Sciences
Research Institute for Linguistics
`fejes@nytud.hu`

Abstract

According to the wide-spread belief, although Nganasan has vowel harmony, the harmonic class of a given stem is unpredictable, completely lexicalized. The research made on two different digital sources of Nganasan (a lexicon of a morphological analyzer with harmonic class of the stems tagged and a morphologically annotated corpus) shows that in most of the cases the harmonic class of stems is well predictable based on the vowels in it. Nganasan vowels belong to two harmonic classes except for one neutral vowel.

Kivonat

A széles körben elterjedt felfogás szerint a nganaszanban ugyan van magánhangzó-harmónia, de a tövek harmóniaosztálya megjósolhatatlan, teljes mértékben lexikalizálódott. Két különböző nganaszan digitális forráson (egy morfológiai elemző tőtárán, illetve egy morfológiailag annotált korpuszon) elvégzett elemzés azonban azt mutatja, hogy hogy az esetek többségében a tövek harmóniaosztálya megbízhatóan megjósolható a tőben szereplő magánhangzók alapján. A nganaszan magánhangzók – egy semleges magánhangzó kivételével – két harmóniaosztályba sorolhatóak.

In the past decade, we have observed an intensive growth of computational linguistic projects on Uralic languages. Some researchers build annotated corpora for their own research, some others develop morphological analyzers for practical tasks (spellchecking etc.). Since these instruments (analyzers, annotated corpora) give a

basic analysis of linguistic structure, even linguists who are not experts of the given language are able to do research based on the annotation. However, although many of these sources are open or at least accessible in some way, it is very rare that someone makes use of the sources built by others for their own research.

I would like to present how I utilized two different Nganasan digital resources to solve a problem of Nganasan vowel harmony. Neither of these projects had been intended to analyze phenomena connected to Nganasan vowel harmony. Furthermore, I am not an expert on Nganasan. However, thanks to digital resources, I could discover regularities of Nganasan morphophonology unknown even for experts.

1 Nganasan vowel harmony: completely lexicalized harmonic class?

The Nganasan vowel inventory contains eight vowels and two diphthongs (see Table 1, based on Helimski 1998, 483 and Várnai 2002, 33).

	Palatal		Velar	
	illabial	**labial**	**illabial**	**labial**
high	*i*	*ü*	*ı*[1]	*u*
mid	*e*		*ə*	*o*
low			*a*	
diphthongs	*^{i}a*			*^{u}a*

Table 1: The vowel system of Nganasan

Nganasan has two kinds of vowel harmony (Helimski 1998, 492-493; Várnai 2002, 55-60). One of these is a palatovelar harmony with alternating pairs *i : ı* and *ü : u*. The palatal variant is used when we find high palatal vowels (*i* or *ü*) in the preceding syllable, the velar variant is used in any other case. Since palatovelar harmony works only in cases when the other kind of vowel harmony (see below) also works, there is always an *i : ı : ü : u* alternation. Palatovelar harmony is very regular; therefore, it is not interesting for us.

The other kind of harmony is much more complex. The alternating pairs we find in this kind of harmony are presented in Table 2 (the variants given in parentheses are the results of palatovelar harmony described above).

---

[1] This vowel is represented in different ways in different publications: *ï* (Helimski 1998), *i̧* (Várnai 2002; Wagner-Nagy 2002a), *ɨ* (Brykina et al.) etc.

after U stems	after I stems
u (ü)	*ı (i)*
a	*ı (i)*
a	*ⁱa*

Table 2: Suffix internal vowel alternation according to vowel harmony

In this alternation, the alternating pairs differ in a different property in each of the three subtypes. In the first case, it seems to be a roundness harmony. In the second case, the main difference is in vowel height. In the third case, the first component of the diphthong differs in height and palatality from the alternating vowel, while the second component is exactly the same. In the following, I call this type of harmony U/I harmony, while I call the first type of harmony discussed above palatovelar harmony.

Whereas palatovelar harmony in Nganasan is obviously a new phenomenon, U/I harmony is a relict of the Uralic palatovelar harmony. Helimski (Helimski 1998, 490) states that as a result of changes in the vowel system, the patterns of vowel harmony have also changed. Following Helimski (Helimski 1998, 490–491), literature on Nganasan morphophonology states that today it is unpredictable whether a given stem belongs to U stems or I stems (Várnai 2002, 56, Várnai and Wagner-Nagy 2003, 322, Katzschmann 2008, 333–334).

Nonetheless, I supposed that the vowels of the stem may determine which harmonic class a specific stem belongs to. I decided to check this hypothesis in digital resources by statistical methods, using simple Perl scripts to see whether there is a connection between the quality of the vowels and the harmonic class of the stem.

2   Vowel statistics based on a lexicon

The first resource I used for my research was the lexicon of the Nganasan morphological analyzer (`http://www.morphologic.hu/urali/`). This program was developed between 2001 and 2009 by Beáta Wagner-Nagy, Zsuzsa Várnai, Sándor Szeverényi and Attila Novák. Although the development of the analyzer also served as a test field of the morphophonological rules described earlier, U/I harmony was not involved. The lexicon was based on *Chrestomathia Nganasanica* (Wagner-Nagy 2002b) and on a Russian–Nganasan–Russian dictionary (Kosterkina et al. 2001).

Attila Novák was kind to send me the lexicon file, so I was able to compare the vowel skeleton of the stems to their harmonic class. In the lexicon file, verbs are listed in their infinitive forms. Since the suffix of the infinitive alternates according

to U/I harmony, it was easy to determinate which harmonic class the given verb stem belongs to. About the 7% of the 3359 suffixable (verb, noun, adjective and numeral) stems were not tagged to belong to any of the harmonic classes, 19 stems were tagged I/U or U/I, both meaning that the stem can be suffixed with suffixes belonging both to the U and the I class: all these 19 stems were ignored. In the case of nouns, adjectives and numerals, most of the stems were explicitly tagged as U stems or I stems. Some stems belonging to some other parts of speech (adverbs, postpositions etc.) were also suffixable and tagged as belonging to one of the harmonic classes. All stems assigned to a certain harmonic class were taken into account, independently of the fact to which part of speech they belonged to.

2.1 Method

For statistics, I wrote a Perl script. First of all, I took the lexical items, I removed unnecessary information, so that only the stem form, the stem type and the meaning remained. I also modified transcription to make it more readable, but here I will not go into the details. The only thing I note is that I used å instead of ⁿa and ä instead of ⁱa to make pattern matching easier.

   For every lexical item I generated a type, which contained only the harmonic class of the stem and the vowel skeleton (consonants were removed from the word form). I used two hashes, and the generated type served as a key in both of them. In the first one, the value was always a string, which contained the stems belonging to the given type. In the second one, the value was a number, and it contained the number of the lexical items belonging to the given type.

   For example, the lexicon contained the word *kiriba* 'bread' in the form presented in Figure 1.

```
kiribaU[N:kenyér];en:;
```

Figure 1: Example line from the lexicon of the morphological analyzer

   In this row, `kiriba` is the stem form, `U` indicates that this is a U stem, `N` is for noun, `kenyér` is the meaning in Hungarian, *en:* should stand for the meaning in English, but in most of the cases it is missing. The type of `kiriba` was `U: iia` indicating this is a U stem and it contains three vowels in the given order. There were two more words among U words with the same vowel skeleton: *kirbiśa(n-)* 'knife, scissors, razor' and *ṫiiďa-* 'to hide'. Therefore, the value associated with the key `U: iia` in the first hash was what is presented in Figure 2 below and 3 in the second hash.

```
kirbibśa 'kés,| olló,| borotva'
kiriba 'kenyér'
t̃iiďa- 'elrejt'
```

Figure 2: Value associated with U: iia in the first hash

When I needed the number of stems matching a given pattern, I took all the keys from the second hash, and the sum of the values associated to the keys matching the required pattern was the number I needed. E.g. when I needed U stems with an *a* in the last syllable, I used the regular expression /^U: .*a$/. However, if I needed specific examples, I used the first hash, and listed the values associated with the keys matching the pattern.

2.2 Results

First of all, I took all the monosyllabic stems. If vowel quality is connected to the stem class in any way, it is supposed to be clear when there is only one vowel in the stem. The results presented in Table 3 showed that most of the monosyllabic stems containing a rounded vowel, *a* or the diphthong ^ua belong to U stems.

	U stems	I stems
o	5	2
ü	7	1
a	12	0
^ua	2	0

Table 3: Vowels tending to occur in U stems in monosyllabic words

The only rounded vowel that did not fit in this picture was *u*: although 5 monosyllabic stems with *u* belong to U stems and only 4 to I stems, the difference was not enough to prove connection between *u* and the U class. However, it seemed to be clear that most of the unrounded vowels are connected to the I class, as it is demonstrated in Table 4.

Nonetheless, monosyllabic stems with *ə* showed no significant difference:[2] 3 of them belonged to the U class and 5 of them to the I class.

---

[2]As a rough approach, I considered that data are significant if one of the two options has at least a two-thirds majority.

	U stems	I stems
i	1	13
e	0	4
ı	2	6
ⁱa	0	3

Table 4: Vowels tending to occur in I stems in monosyllabic words

As a next step, I also examined polysyllabic stems containing the same vowel. The picture did not radically change: the results showed that rounded vowels (including the diphthong *ᵘa*) and *a* are connected to U stems, and unrounded vowels are connected to I stems: even stems with *u* showed the preference of U stems. The monosyllabic and polysyllabic stems together show a clear preference in the cases presented in Table 5.

	U stems	I stems
u	41	6
ü	53	2
a	54	1
i	4	75
ı	2	23

Table 5: Vowels tending to occur in U or I stems in mono- and polysyllabic words

Since non-first syllable *e* and *o* are rare, usually occurring in fresh loanwords, I could not find any polysyllabic words with just *e*. However, there was one (native) word with two *o*s: *ďoot* 'ember, firebrand', and it was an I stem. Despite that fact I considered *o* to be connected to U stems: maybe it is not the best solution, but data show that non-first syllable *o* occurs mostly in U stems, irrespective of other vowels in the stem: *raboťij* 'worker', *kresto* 'cross', *ťemodanə* 'suitcase', *kilogrammə* 'kilogram', *kilomətra* 'kilometer', *honu'o* 'plait, braid' etc. Non-first syllable diphthongs are not so rare as non-first syllable *e* or *o*, but I could find just one stem containing more than one diphthong: *mⁱaiðⁱa-* 'to be married, to have a husband' (an I stem, as we expect).

However, it seemed that *ə* is not connected to any of the stem types: I found altogether 49 mono- and polysyllabic U stems containing no other vowel but *ə*, while 38 such stems were I stems. My conclusion was that Nganasan U/I harmony is basically a roundness harmony. I found two exceptions: the vowel *a*, although phonetically unrounded, occurred in U-stems almost without exceptions. Therefore I decided to take

it as a phonologically rounded vowel. So, in the following when I write "rounded vowel", the set also includes *a* and diphthong *ᵘa*. The vowel *ə* was ambiguous, it did not trigger either of the stem classes. All other unrounded vowels were connected with I stems.

If we have a look at Table 2 again, we can see that after U stems we find vowels which occur mostly in U stems, while after I stems we find vowels which usually occur in I stems. That is exactly what we are expecting when speaking about vowel harmony. Nonetheless, we expect that it will work even with stems containing different vowels. I distinguished the following types of stems:

- **U**: containing exclusively rounded vowels (*u, o, ü, a* or diphthong *ᵘa*);

- **Ux**: containing exclusively rounded vowels (*u, o, ü, a* or diphthong *ᵘa*) and neutral *ə*;

- **I**: containing exclusively unrounded vowels (*i, e, ı* or diphthong *ⁱa*);

- **Ix**: containing exclusively unrounded vowels (*i, e, ı* or diphthong *ⁱa*) and neutral *ə*;

- **UI**: containing both rounded (*u, o, ü, a* or diphthong *ᵘa*) and unrounded vowels (*i, e, ı* or diphthong *ⁱa*), but not neutral *ə*;

- **UIx**: containing both rounded (*u, o, ü, a* or diphthong *ᵘa*) and unrounded vowels (*i, e, ı* or diphthong *ⁱa*) and also *ə*;

The results justified my supposition (Table 6).

	U stems	I stems
U	562 (96%)	24 (4%)
Ux	895 (90%)	94 (10%)
I	12 (5%)	230 (95%)
Ix	81 (16%)	417 (84%)
UI	276 (52%)	252 (48%)
UIx	213 (48%)	228 (52%)

Table 6: Correlations between the vowel skeleton and harmonic class of stems

These numbers show that if a stem contains exclusively rounded or exclusively unrounded vowels, the stem type is highly predictable: just every twentieth bet will be wrong. However, the presence of a neutral vowel will spoil our chances. If the stem

contains rounded and neutral vowels, every tenth bet will be wrong; if it contains unrounded and neutral vowels, we loose every sixth bet. Nonetheless, even these chances can be considered very good; it shows a high predictability of stem type based on the vowels the stem contains. Moreover, when the stem contains both rounded and unrounded (and possibly neutral) vowels, it seems to be completely unpredictable whether the stem belongs to the U class or the I class.

If we look at the composition of the stems, we can see that 2315 stems (70%) are (internally) harmonic (that is they belong to the U, Ux, I or Ix category), and just 969 stems (30%) are disharmonic. (The 87 stems containing just ə are ignored here.re.) That shows that also inside stems, rounded vowels prefer to occur with other rounded vowels and unrounded vowels with other unrounded vowels.

Bisyllabic words with both rounded and unrounded vowels showed that the stem type is more closely related to the last than to the first syllable (Table 7, first two rows).

bisyllabic stems	U stems	I stems
{u, o, a, ü, ᵘa} + {i, e, ı, ⁱa}	43 (31%)	94 (69%)
{i, e, ı, ⁱa} + {u, o, a, ü, ᵘa}	66 (79%)	18 (21%)
{u, o, a, ü, ᵘa} + ə	93 (87%)	14 (13%)
ə + {u, o, a, ü, ᵘa}	70 (85%)	12 (15%)
{i, e, ı, ⁱa} + ə	13 (18%)	59 (82%)
ə + {i, e, ı, ⁱa}	12 (25%)	36 (75%)

Table 7: Harmonic classes of bisyllabic disharmonic stems

Although in polysyllabic words (words of three or more syllables) overall statistics did not show any evident correlation with the first, penultimate or last syllable, it became clear that the final syllables have a key role in deciding the stem type. If we examine bisyllabic stems with a rounded vowel and a neutral vowel, we find no striking difference between the penultimate and last position (Table 7, third and fourth rows). However, when we find an unrounded and a neutral vowel in a bisyllabic word, the penultimate syllable seems to be more relevant (Table 7, last two rows).

In polysyllabic disharmonic stems, when the final two vowels are both rounded or both unrounded, the stem type is highly predictable. Nonetheless, if we find a rounded and an unrounded vowel in the final two syllables, their order does not help to predict the stem type (Table 8).

When we find a neutral and a round vowel in the last two syllables of a polysyllabic disharmonic stem, it helps to predict the stem type, especially when the rounded vowel is in the last syllable. However, when we find a neutral and an unrounded vowel in the last two syllables of a polysyllabic disharmonic stem, although it helps

two final syllables	U stems	I stems
{u, o, a, ü, ᵘa}	91 (91%)	9 (9%)
{i, e, ı, ⁱa}	14 (16%)	71 (84%)
{u, o, a, ü, ᵘa} + {i, e, ı, ⁱa}	39 (53%)	35 (47%)
{i, e, ı, ⁱa} + {u, o, a, ü, ᵘa}	72 (48%)	77 (52%)

Table 8: Harmonic classes of polysyllabic stems with harmonic and disharmonic vowels in the last syllables

to predict the stem type, our prediction will be wrong more often than in the case of rounded vowels. Moreover, the stem type is more predictable if the unrounded vowel is in the penultimate syllable than when it is in the last one (Table 9).

two final syllables	U stems	I stems
{u, o, a, ü, ᵘa} + ə	78 (66%)	41 (34%)
ə + {u, o, a, ü, ᵘa}	21 (87%)	3 (13%)
{i, e, ı, ⁱa} + ə	24 (27%)	66 (73%)
ə + {i, e, ı, ⁱa}	35 (38%)	56 (62%)

Table 9: Harmonic classes of polysyllabic stems with a harmonic and a neutral vowel in the last syllables

It seems that the roundedness of the vowels in the first part of the word counts even when the last two vowels are neutral. The order of the rounded and unrounded vowels makes a difference even when the two last syllables are neutral. Note, however, that the number of examples is very low here. Therefore, the reliability of statistics is disputable (Table 10).

two neutral final syllables	U stems	I stems
{u, o, a, ü, ᵘa} ...+ ə + ə	125 (79%)	34 (21%)
{i, e, ı, ⁱa} ...+ ə + ə	33 (35%)	61 (64%)
{u, o, a, ü, ᵘa} + {i, e, ı, ⁱa} ... + ə + ə	2 (20%)	8 (80%)
{i, e, ı, ⁱa} + {u, o, a, ü, ᵘa} ... + ə + ə	4 (57%)	3 (43%)

Table 10: Harmonic classes of polysyllabic stems with a rounded and a neutral vowel in the last syllables

We can conclude that in Nganasan we have two harmonic classes of vowels and a

neutral vowel. The vowels in the stem play a great role in the decision which variant of an alternating suffix will be attached to the stem: the variant which contains the vowel belonging to the same harmonic class as the vowels of the stem. When the stem contains vowels belonging to both classes, vowels closer to the suffix (i.e. closer to the end of the stem) play a more significant role. In most of the cases the stem category is well predictable, in other cases (when it contains exclusively neutral vowels or vowels belonging to both harmonic classes) we can predict that the stem category is unpredictable.

It is clear that Nganasan vowel harmony works in a very different way from Finnish or Hungarian vowel harmony, or vowel harmony known from any other Uralic (or Turkic) language.[3] However, since the main principle of the examined phenomenon is that the vowels of the suffix must belong to the same group as the vowels of the stem, it must be considered a type of vowel harmony.

3 Vowel statistics based on a corpus

When I started to spread news on my findings, Beáta Wagner-Nagy suggested checking my results against different sources on Nganasan, namely on the Nganasan Spoken Language Corpus or NLSC (Brykina et al.). While the Nganasan Morphological Analyzer was tested on a text collection of one fieldworker from one informant (or NLSCLabanauskas 2001), NSLC contains texts collected by different fieldworkers, from different informants at different times.

3.1 Method

I downloaded 54 xml files, each containing an annotated text. First, using a Perl script, I removed xml tags and other irrelevant information. I got a file which contained different annotation tiers in different lines. With the help of another Perl script, I merged every word in one line with its annotation. In the new files, every line contained a word segmented to morphs, then a tab character followed by the meaning of the stem in English and Russian, separated by a pipe (vertical line) from each other, and separated by a hash mark form the morpheme tags following it. Annotation was followed by a tab character and the reference number in parentheses. (Reference numbers were kept to help debugging.) Consequently, the example above was converted to the string presented in Figure 3.

The hashtag was necessary since sometimes the gloss for meaning also contained a hyphen like in Figure 4.

---

[3] The analysis of these differences are out of scope of the current study.

```
kontu-gu-mə      take.away|увезти#-IMP-1SG.O      (694)
```

Figure 3: Word form and analysis put in one line

```
Abamu-nu      Ust'-avam|Усть-Авам#-GEN.1PL
```

Figure 4: Hyphen in meaning

In some marginal cases, the glosses in Russian and in English were different or the annotation was missing. I gave a feedback on these cases to the developers of the corpus, and I ignored these instances. It was not a considerable loss, because most of the ignored word forms did not contain any suffixes.

My next step was merging all the word forms occurring in the text into one text file. In this file, word forms were presented similarly to the previous one, but reference numbers were omitted and the frequency of the form (e.g. the total number of occurrences in the 54 texts) stood at the end of the line. Since our example word form occurred only once, the line containing it will look like in Figure 5.

```
kontu-gu-mə      take.away|увезти#-IMP-1SG.O      1
```

Figure 5: Word form and analysis with its frequency

In this file word forms were listed in a quasi-alphabetic order. It makes it easy to notice that the same stems are sometimes glossed in a different way. For example, above the line presented in Figure 5, we find the line presented in Figure 6

```
kontu-gu-mə      carry|отнести#-IMP-1SG.O      1
```

Figure 6: Word form and analysis with its frequency

We can also find the stem *kontu-* glossed with the meaning 'lead/отвести' nearby. We can also find the stem *kondu-*, most of the times glossed as 'take.away/увезти', but as 'kill/убить', too. This is important because the next step was the identification of the stems. Every stem allomorph with a given meaning was considered to be a separate stem. Therefore, although both *kontu-* and *kondu-* in all of their meanings belong to the same verb, they counted as five different stems (see Figure 7; numbers at the end of the lines indicate the number of word forms in which the given stem allomorph occurs with the given meaning – it is ignored how many times the given word forms occur).

This is important because this method inavoidably modifies the results of statistics.

```
kondu 'kill|убить'        3
kondu 'take.away|увезти'      7
kontu 'carry|отнести'      1
kontu 'lead|отвести'      1
kontu 'take.away|увезти'      22
```

Figure 7: Word form and analysis with its frequency

However, the probability that a stem which proves our expectations will be overrepresented is the same that a stem which contradicts our hypothesis will be overrepresented. Therefore, we can hope this result will reflect roughly the same as a more precise approach, namely, if we tried to associate all the allomorphs of a lexeme with all glossed meanings to a single stem.

To get the stems organized as presented above, I used another Perl script. This segmented the word forms and made two files: the list of stems (as presented above) and a list of suffixes (as presented in Figure 8).

From this file, I generated another list in which allomorphs are ordered according to their functions (Figure 9).[4]

The next step was to define which morphs undergo vowel harmony and which harmonic class they belong to. Since Nganasan morphophonology is very complex, I categorized allomorphs manually. Allomorphs containing no vowels undergoing harmony (such as [1SG.O] or [IMP] *ŋəə*) were simply left out of the list. Only those allomorphs were taken as undergoing harmony which had at least one pair belonging to the other harmonic class. For example, *-gu-, -gü-, -ku-* and *-kü-* undergo palatal harmony, and according to the literature (e.g. Helimski 1998, 493, Várnai 2002, 57–59) this happens only when it also undergoes rounding harmony. However, (despite the relatively high number of the above mentioned allomorphs, especially with *u*), none of the allomorphs *-gɨ-, -gi-, -kɨ-* or *-ki-* were attested, I took *-gu-, -gü-, -ku-* and *-kü-* as allomorphs not undergoing rounding harmony. Because of the same reason, *-ŋəi-* was also taken as a suffix not undergoing vowel harmony. However, other allomorphs beginning with *ŋ* were regarded as belonging to one or other harmonic class (see Figure 10).[5]

---

[4]The Nganasan Spoken Language Corpus uses the character *ɨ* to indicate velar illabial high vowel. For technical reasons, these are substituted by *ï* in the presented codes. For the same reason, *ʔ* is replaced by *ʔ* and *'* is replaced by a vertical straight apostrophe.

[5]In fact, not all these allomorphs are the allomorphs of the same morpheme. The allomorphs beginning with *k* or *g* are always used in first person forms, therefore a tag something like [1IMP] or [IMP1] would be more appropriate for them. The allomorph *-ŋəə-* is used for third persons in the subjective and

```
[...]
gu [DRV] 1
gu [DUR] 2
gu [IMP] 33
[...]
gü [IMP] 8
[...]
mə [1SG.O] 66
mə [1SG] 2
mə [ACC.1SG] 28
mə [ACC.SG.1SG] 46
mə [DRV] 1
mə [EXCL.[NOM.SG]] 3
mə [EXCL] 4
mə [MOM] 3
mə [NOM.1SG] 8
mə [NOM.SG.1SG] 93
mə [SG.1SG] 3
[...]
```

Figure 8: List of suffixes (segments)

Allomorphs tagged in a slightly different way but clearly belonging to the same morpheme were also included in the list (see Figure 11).

Allomorphs clearly belonging together but containing different zero allomorphs were also taken into account, see Figure 12.

Allomorphs undergoing alternation but containing vowels belonging to different harmonic classes were indicated with an exclamation mark (see Figure 13). Later these allomorphs were not taken into account, but they were collected to another list for further studies. They may be a result of an annotation mistake, a slip of the tongue, signs of disintegration of vowel harmony (maybe not independently of language loss) etc.

---

objective conjugation, *-ŋəi-* occurs in third person reflexive conjugation: these are two distinct morphemes again. The allomorphs *-ŋu-* and *-ŋü-* are used in second person subjective and objective conjugation, *-ŋa-* in second person reflexive conjugation, while *-ŋi-* and *-ŋı-* occur in second person forms of all the three conjugations: two additional morphemes with some overlapping allomorphs again. So we have to speak about six different morphemes, tagged in different ways (c.f. Helimski 1998, 505, Wagner-Nagy 2002a, 110). Nonetheless, this problem does not influence my statistics.

```
[1SG.O] mə 66
[1SG.O] n'ə 1
[...]
[IMP] gu 35
[IMP] gü 8
[IMP] ku 51
[IMP] kü 5
[IMP] ŋa 5
[IMP] ŋi 2
[IMP] ŋu 21
[IMP] ŋü 3
[IMP] ŋəi 5
[IMP] ŋəə 25
[IMP] ŋï 24
[IMP…] ŋu 1
```

Figure 9: List of suffixes ordered by tags (segments)

```
U[IMP] ŋa 5
I[IMP] ŋi 2
I[IMP] ŋu 21
U[IMP] ŋü 3
I[IMP] ŋï 24
```

Figure 10: Suffixes having allomorphs with vowels belonging to both harmonic
classes were taken as morphemes undergoing vowel harmony

```
U[IMP…] ŋu 1
```

Figure 11: Harmonizing suffix with a uniquely formed tag with its harmony class
tagged

```
U[LOCN.[ACC.SG]] rəmu 1
I[LOCN.[GEN.SG]] d'əmï 1
U[LOCN.[NOM.SG]] rəmu 1
```

Figure 12: List of suffixes with zero allomorphs with their harmony class tagged

```
I[LOC.PL] tini 28
![LOC.PL] tinü 1
U[LOC.PL] tünü 1
```

Figure 13: List of suffixes with their harmony class tagged and a non-categorizable allomorph among them

Based on this definition of harmonic classes of suffixes, all the suffixes were replaced by their harmonic classes in our file. E.g. our example in Figure 5 was first reassigned into a form presented in the first line of Figure 14 and then the suffixes were changed into their harmonic classes, as it is presented in the second line. NB, although at first sight -*gu*- may seem to belong to the U harmonic class, it is neutral (see above). However, the form presented in the third line contains two harmonic suffixes, therefore it is changed into the form presented in the fourth line of the same figure.

```
kontu[take.away|увезти]-gu[IMP]-mə[1SG.O]
kontu[carry|отнести]-N-N
kontu-ra-?a take.away|увезти#-PASS-PF.[3SG.S]
kontu[take.away|увезти]-U-U
```

Figure 14: Annotations realigned to the morphs they belong to and changed to their harmonic classes

Interestingly, I have found about 200 word forms in which I could identify suffixes belonging to both of the harmonic classes, e.g. see Figure 15. In this case, my classification contradicts the literature (e.g. Wagner-Nagy 2002a, 78), according to which the plural lative suffix always contains *i*, therefore it is not a harmonic suffix. However, I took it as an alternating suffix because I could find forms -*tü*- and -*ntü*- as in the word forms shown in Figure 16.

```
ma-ti-tü tent|чум#-LAT.PL-3SG
ma[tent|чум]-I-U
```

Figure 15: Stem with two suffixes belonging to different harmonic classes

It is worth noting that the word form meaning 'to his tents' has two different forms here: *matitü* and *matütu* and the suffix is never attested with vowels *ı* or *u*. A similar case is attested with the coaffix of perfect verbs, which should also contain *i* all the

```

```
ma[tent|чум]-tü[LAT.PL]-tu[OBL:3SG]
ku[which|какой]-ni[LOCPRON]-ə[ADJZ]-ntü[LAT.PL]-ndü [OBL.2PL]
ŋabtə[hair|волосы]-tü[LAT.PL]-tü[OBL.3SG]
```

Figure 16: Plural lative suffixes with *ü*

time ( Wagner-Nagy 2002a, 101). But it also occurs with *a*, *ü* and *ı* and several kinds of diphthongs (see below) in the annotated corpus.

In any case, I ignored word forms with suffixes (seemingly) belonging to different harmonic classes and used just the forms in which all the suffixes belonged to the same harmonic class (or were neutral).

## 3.2   Results

I made two kinds of statistics on these forms. The first shows the composition (vowel skeleton) of the stems before the suffixes of the two harmonic classes. I used harmonic categories based on the previous study. However, the corpus contained two diphthongs, unknown by the literature ( Helimski 1998, 483, Várnai 2002, 33) and not used by the morphological analyzer: $^{i}a$ and $^{ü}a$ (for pattern matching, substituted by i̧ and ű, respectively). As a hypothesis, I treated $^{i}a$ as unrounded (as $^{i}a$ is) and $^{ü}a$ as rounded (as $^{u}a$ is). The distribution shown by the statistics is presented in Table 11.

|        | U class   | I class   |
|--------|-----------|-----------|
| **U**  | 395 (87%) | 59 (13%)  |
| **Ux** | 305 (78%) | 85 (22%)  |
| **I**  | 26 (11%)  | 205 (89%) |
| **Ix** | 52 (25%)  | 153 (75%) |
| **UI** | 148 (51%) | 142 (49%) |
| **UIx**| 42 (49%)  | 44 (51%)  |
| only ə | 39 (59%)  | 27 (41%)  |

Table 11: Number of stems with a given vowels skeleton type before suffixes belonging to different harmonic classes

The results show again that harmonizing suffixes with rounded vowels tend to occur after stems with rounded vowels and harmonizing suffixes with unrounded vowels tend to occur after stems with unrounded vowels. The presence of a neutral vowel weakens this tendency. After stems containing vowels belonging to both or none of the harmonic classes the harmonic class of the suffix is unpredictable.

To check whether the categorization of each vowel is proper, I also made statistics for the separate vowels. I counted how many times they occur in stems suffixed by the rounded and unrounded variants of harmonizing suffixes. (If the same vowel occurs twice or three times in the same stem, it is counted as two or three occurrences, respectively.)

| | U class | I class |
|---|---|---|
| *u* | 360 **(73%)** | 135 (27%) |
| *o* | 204 **(76%)** | 64 (24%) |
| *a* | 570 **(80%)** | 135 (20%) |
| *ü* | 314 **(82%)** | 71 (18%) |
| $^u a$ | 53 **(91%)** | 5 (9%) |
| $^ü a$ | 6 **(86%)** | 1 (14%) |
| *i* | 218 (33%) | 417 **(67%)** |
| *e* | 42 (24%) | 135 **(76%)** |
| *ı* | 44 (17%) | 222 **(83%)** |
| $^i a$ | 28 (35%) | 51 **(65%)** |
| $^ı a$ | – | 2 **(100%)** |
| *ə* | 540 **(58%)** | 383 (42%) |

Table 12: Number of different vowels before suffixes belonging to different harmonic classes

The data presented in Table 12 show that at least two thirds of any vowel occurs in stems before a suffix belonging to the same harmonic class as I classified them. Although neutral *ə* also occurs more times before U class suffixes, the range here is considerably lower than the lowest value for vowels classified as belonging to a harmonic class (seven twelfth instead of eight).

## 3.3   Some problems

There were almost 250 stems which occurred with suffixes belonging to both of the harmonic classes. In the statistics above, they were presented as two different stems belonging to different harmonic classes. E.g. the word *basa* 'iron, money' was tagged as an U stem in the morphonological analyzer. In the corpus, in most of the cases it is followed by suffixes belonging to the U class (or neutral), but in some cases it is followed by the I class variant of the destinitive suffix (Figure 17).

However, similarly to the case of the plural lative suffix and the perfect coaffix seen above, the destinitive suffix is attested only with vowels *i, ü* and *ı*, therefore it is

```
basa-ði-n'i? iron|железо#-DST-ACC.PL.1PL
basa[iron|железо]-I-I
basa-ði-n'ü? money|деньги#-DST-ACC.PL.1PL
basa[money|деньги]-I-U
basa-ði-t'ü iron|железо#-DST-ACC.PL.3SG
basa[iron|железо]-I-U
```

Figure 17: Suffixed forms of *basa* 'iron, money'

disputable whether it is a harmonizing suffix. Moreover, the (nominative-accusative-genitive) plural form with a first person plural possessor (PL1.PL) also occurs with vowels *i* and *ü* only. If we removed all these suffixes from the list of the harmonizing suffixes, rounding harmony would seem more regular.

Nonetheless, there are cases which are more complicated. Sometimes we find suffixes of different harmonic classes a certain stem, see Figure 18.

```
ŋad'a-ðu younger.sibling|младший.брат/сестра#-NOM.SG.3SG
ŋad'a-ðï younger.sibling|младший.брат/сестра#-NOM.SG.3SG
```

Figure 18: The same form with suffixes belonging to different harmonic classes

In this case, the suffix occurs with all the four harmonizing vowels and I have found no explanation for the lack of vowel harmony in the second case. All such cases need a detailed analysis.

Moreover, to evaluate the results, it will be necessary to check the distribution of seemingly irregular forms among the different informants. If we find that many of these forms come from a smaller group of informants, we have to suggest that they are further on the way of language loss (cf. Helimski 1998, 493) and therefore data from them are not reliable for researching Nganasan vowel harmony in its intact form. However, data from these people are a useful source for the research of language loss.

4   Conclusion

Contrary to what we find in the literature on Nganasan morphophonology, Nganasan U/I vowel harmony proved to be quite well predictable by a statistical analysis of Nganasan data. It is basically a rounding harmony, although phonetically unrounded *a* belongs to rounded vowels and *ə* does not belong to either of the classes. The two very different digital sources of Nganasan show basically the same tendency: stems

containing rounded vowels tend to be suffixed with harmonizing suffixes containing rounded vowels and stems containing unrounded vowels tend to be suffixed with harmonizing suffixes containing unrounded vowels.

Acknowledgments

References

Maria Brykina, Valentin Gusev, Sándor Szeverényi, and Beáta Wagner-Nagy. ???? Nganasan Spoken Language Corpus (NSLC). Archived in Hamburger Zentrum für Sprachkorpora. Version 0.1. http://hdl.handle.net/11022/0000-0001-B36C-C. Accessed 2017-06-16, published2016-12-23.

Eugene Helimski. 1998. *The Uralic Languages*, Routledge, London and New York, chapter Nganasan, pages 480–515.

Michael Katzschmann. 2008. *Chrestiomathia Nganasanica.* Norderstedt, Hamburg.

N. T. Kosterkina, A. Č. Momde, and T. Ju. Ždanova. 2001. *Slovar' nganasansko-russkij i russko-nganasanskij.* Prosvesčenije, Sankt-Petersburg.

Kazys Labanauskas. 2001. *Nganasanskaja fol'klornaja hrestomatija.* Number 6 in Fol'klor narodov Tajmyra. Tajmyrskij okružnyj centr narodnogo tvorčestva, Dudinka.

Zsuzsa Várnai. 2002. *Chrestomathia nganasanica*, SZTE Finnugor Tanszék – MTA Nyelvtudományi Intézet, Szeged – Budapest, chapter Hangtan, pages 33–69.

Zsuzsa Várnai and Beáta Wagner-Nagy. 2003. Magánhangzó-harmónia a nganaszanban. *Nyelvtudományi Közlemények* 100:321–337.

Beáta Wagner-Nagy. 2002a. *Chrestomathia nganasanica*, SZTE Finnugor Tanszék – MTA Nyelvtudományi Intézet, Szeged – Budapest, chapter Alaktan, pages 71–126.

Beáta Wagner-Nagy. 2002b. *Chrestomathia nganasanica*, SZTE Finnugor Tanszék – MTA Nyelvtudományi Intézet, Szeged – Budapest, chapter Szójegyzék, pages 274–291.