LING GU4120 Language Documentation and Field Methods
Progress Report 1: The Sound System of Sakha

Eleanor Lin, Columbia University

Friday, March 3, 2023

Yakut is a Turkic language spoken by approximately 378,000 people in the following administrative divisions of Russia, which lie in Siberia: the Republic of Sakha (also called Yakutia), Irkutsk and Magadan provinces, Khabarovsk Krai, and Krasnoyarsk Krai. The endonym for the language is Sakha (Yakut 2023).

The consultant for this paper is Platon Shamaev, who is a native speaker of Sakha and also speaks English. He moved to the United States from Russia recently and currently works as a lawyer in New York, where he is affiliated with Columbia University's Human Rights Advocates Program.

Consonants

I propose that Sakha has the eighteen consonant phonemes listed in Table 1. Consonants are shown in parentheses when there is not enough data to confirm distinctiveness through (near)-minimal pairs.

	Bilabial	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Plosive	(p) b	t d			k (g)		
Affricate			(tʃ) ʤ				
Nasal	m	n			ŋ		
Trill		r					
Tap or Flap							
Fricative		s				χ ʁ	(h)
Approximant				j			
Lateral Approximant		l					

Table 1: Sakha consonant phonemes. Consonants for which (near)-minimal pairs have not been found are shown in parentheses.

The following near-minimal set illustrates consonant phonemes /χ/, /t/, /r/, /n/, /d/, /ʁ/, /l/, /j/, /m/, and /ŋ/ between two instances of the vowel /ɑ/. Each example lists the following, ordered left to right: the consonant illustrated, the phonetic transcription of the word, the gloss, and the date on which the word was elicited. Note that (4) is a Yakutized Russian loanword. (The English loanword 'banya' comes from the same Russian word, and its spelling approximately reflects the Russian pronunciation).

(1) /χ/   [sɑχɑ]   'Sakha'   January 19, 2023

(2) /t/   [ɑtɑχ]   'leg'   January 19, 2023

(3) /r/   [χɑrɑχ]   'eye'   January 19, 2023

(4) /n/   [bɑnɑ]   'bathroom'   February 9, 2023

(5) /d/   [dʒɑdɑnɑ]   'poor'   January 26, 2023

(6) /ʁ/   [mɑʁɑn]   'white'   January 26, 2023

(7) /l/   [χɑlɑn]   'sky'   January 19, 2023

Note that /l/ is velarized (realized as [ɫ]) unless otherwise noted.

(8) /j/   [ɑjɑn]   'trip'   February 23, 2023

(9) /m/   [tɑmɑχ]   'drop'   February 23, 2023

(10) /ŋ/   [sɑŋɑs]   'aunt'   February 16, 2023

Gaps in the consonant inventory

In this section, I illustrate the remaining consonants listed in Table 1, which do not fit into the near-minimal set (1)-(10).

Elicitations have not yet yielded any examples of /dʒ/, /k/, /b/, or /s/ occurring in the context /ɑ_ɑ/ shown in (1)-(10). However, the following near-minimal set, with each consonant in a word-initial context before /ɑ/, demonstrates that /dʒ/, /k/, /b/, and /s/ contrast with one another, and with /m/, /χ/, /t/, and /d/, which were shown to be distinctive in (9), (1), (2), and (5), respectively. More data is needed to confirm whether /dʒ/, /k/, /b/, and /s/ also contrast with the other consonants illustrated by (1)-(10).

(11) /dʒ/   [dʒɑχtɑr]   'woman'   January 19, 2023

(12) /k/   [kɑɾɑχutɑ]   'tears'   February 23, 2023

(13) /b/   [bɑltɪ]   'younger sister'   January 19, 2023

(The process which results in [ɪ] in (13) is discussed in the section on phonological rules.)

(14) /s/   [sɑŋɯ]   'shoulder'   January 19, 2023

(15) /m/   [mɑχtɑl]   'thank you'   January 19, 2023

(16) /χ/   [χɑrɑχ]   'eye'   January 19, 2023

(17) /t/   [tɑmɑχ]   'drop'   February 23, 2023

(18) /d/   [dɑlɑɦɑ]   'small bridge'   February 23, 2023

More data is also needed to determine whether [p], [tʃ], and [g] are consonant phonemes.

The only non-loanword citation form in which [p] has appeared in a non-word-final context is [olopɔs], 'chair.' For the relevance of the word-final position, see **Allophonic Variation and Contextually Limited Contrast** below. Briefly, since Sakha devoices obstruents word-finally, observations of [p] word-finally may underlyingly be /b/, and therefore cannot serve as evidence of /p/ as a phoneme.

[tʃ] has been observed in some of the following contexts:

(19) [boltʃʊt]   'hunter'   February 21, 2023

(20) [intʃɑʁɑj]   'wet'   January 26, 2023

(21) [mitʃɛr]   'smile'   January 19, 2023

(22) [tʃɯjtʃɑχ]   'bird'   January 19, 2023

(23) [ytʃygɛj]   'good'   January 26, 2023

[g] has also appeared in a variety of contexts, non-exhaustively illustrated below:

(24) [tuguj] 'what?' January 19, 2023

(25) [ytʃygɛj] 'good' January 26, 2023

(26) [ɑnnigɑr] 'under' February 21, 2023

(27) [bɛhɛgɛ] 'we' February 9, 2023

(28) [bygyn] 'today' February 21, 2023

In addition to (27) and (49), /h/ has also been observed in contexts such as the following:

(29) [hɔs] 'room' February 9 & 16, 2023

(30) [ɛhɛgɛ] 'you' (2nd person plural subject) February 9, 2023

Vowels

Sakha has the eight vowels shown in Figure 1, with binary roundness, height, and backness contrasts which can be described with features [round], [high], and [back]. /i/, /yː/, /ɯ/, and /u/ are [+high]. /ɛ/, /ɔ/, /œ/ and /ɑ/ are [-high]. There are also five diphthongs /yœ/, /iɛ/, /ɯɑ/, and /uɔ/. Our data also provides evidence for length contrasts for /ɑ/, /i/, /ɛ/, and /ɯ/. Near-minimal pairs that demonstrate mutual contrastiveness with all other vowels are not available for vowels shown in parentheses.
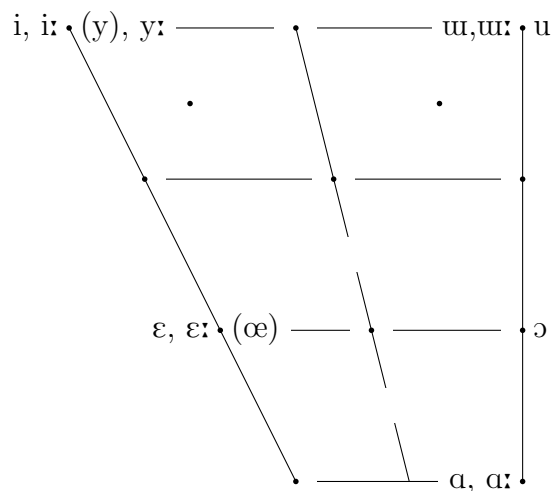

Figure 1: Sakha vowels. A complete set of near-minimal pairs is not available for vowels shown in parentheses.

The following near-minimal set illustrates all vowel phonemes of Sakha (except /œ/) in a word-initial position before the consonant /t/, the diphthongs /uɔ/ and /yœ/, and length contrasts for /i/, /ɑ/, and /ɯ/.

(31) /i/ [itiː] 'hot' January 19, 2023

(32) /iː/ [iːt] 'bring up, raise' February 16 & 23, 2023

(33) /y/ [yːt] 'milk' February 16 & 23, 2023

(34) /ɛ/ [ɛt] 'meat' February 9 & 23, 2023

(35) /ɑ/ [ɑt] 'horse' January 19 & February 23, 2023

(36) /aː/   [aːt]   'name'   January 26 & February 23, 2023

(37) /ɯ/   [ɯt]   'dog'   January 19 & February 23, 2023

(38) /ɯː/   [ɯːt]   'release me' (imperative mood)   February 23, 2023

(39) /u/   [utaχ]   'beverage'   January 26, 2023

(40) /ɔ/   [ɔt]   'grass'   February 16 & 23, 2023

(41) /uɔ/   [uɔt]   'fire'   January 19 & February 23, 2023

(42) /yœ/   [yœt]   'backwater creek'   February 23, 2023

While there are no examples that fit into the near-minimal set above to illustrate diphthongs /iɛ/ and /ɯɑ/, evidence for these diphthongs comes from both Platon's own enumeration of some Sakha diphthongs in isolation (during elicitation on January 19) and from the examples below:

(43) /ɯɑ/   [tɯɑl]   'wind'   January 19, 2023

(44) /ɯɑ/   [ɯrɯɑ]   'song'   January 19, 2023

(45) /iɛ/   [biɛs]   'five'   January 19, 2023

The diphthongs /yœ/, /iɛ/, /ɯɑ/, and /uɔ/ all consist of a sequence of a [+high] vowel followed by a [-high] vowel, suggesting that Sakha diphthongs in general may move from [+high] to [-high]. Furthermore, each diphthong features vowels with matching roundedness and frontness. This is consistent with the **Backing** and **Rounding Harmony** rules presented in the next section, **Allophonic Variation and Contextually Limited Contrast**.

The following two minimal pairs demonstrate that length is contrastive for /ɛ/:

(46) /ɛ/   [ɛbɛ]   'river'   February 16, 2023

(47) /ɛː/   [ɛbɛː]   'grandmother'   February 16, 2023

(48) /ɛ/   [ɛhɛ]   'bear'   January 19 & February 16, 2023

(49) /ɛː/   [ɛhɛː]   'grandfather'   January 19 & February 16 & 23, 2023

While we have not yet observed /œ/ in a word-initial context before /t/, leaving a gap in the minimal set presented in (31), /œ/ has been observed in (50) and (52), which form minimal pairs with (51) and (53), respectively. Therefore, we can at least conclude that /œ/, /ɑ/, and /y/ contrast, although further data is needed to confirm the contrast between /œ/ and other vowels.

(50) [tœbœ]   'head'   January 19, 2023

(51) [tɑbɑris]   'friend'   January 26, 2023   NB: Yakutization of Russian for 'friend'

(52) [tœrdys]   'fourth'   February 16, 2023

(53) [tyrgɛn]   'fast'   January 26, 2023

More data is also needed to determine whether length is contrastive for /y/, /œ/, /u/, and /ɔ/. Platon describes the [yː] in 'milk' (33) as long, but we have no minimal pairs to confirm whether length is contrastive for this vowel.

Allophonic Variation and Contextually Limited Contrast

In this section, I give an overview of allophonic variation and neutralization in Sakha.

Tapping

/r/ → [ɾ] / ._

/r/ becomes [ɾ] at the beginning of a syllable.

Due to phonotactic constraints, consonants may occur only at the beginning or end of a syllable. The following examples (with syllable boundaries marked) illustrate that /r/ is realized as [r] at the ends of syllables:

(54) [ɑj.mɑχ.tɑr]   'relatives'   February 16, 2023

(55) [ɑr.dɑχ.tɯr]   'it is raining'   January 19, 2023

(56) [bœr.dʊk]   'flour'   February 16, 2023

(The process which produces [ʊ] in the surface representation of 'flour' is discussed in **Word-Final Lowering**.)

[ɾ] occurs as an allophone of /r/ syllable-initially:

(57) [bɑ.ɾɯ]   'all'   February 9, 2023

(58) [bɯ.ɾɑs]   'doctor'   February 21, 2023

(59) [χɑ.ɾɑ]   'black'   January 26, 2023

Word-Final Devoicing

[-sonorant] → [-voice] / _#

Obstruents devoice word-finally.

For example, as shown in (13), (18), and (6), respectively, /b/, /d/, and /ʁ/ are distinctive in non-word-final contexts (with (2) and (1) providing near-minimal pairs for /t/ and /d/, and /χ/ and /ʁ/). However, word-finally, we only find /p/, /t/, and /χ/, suggesting that the voicing contrast for obstruents is neutralized at the ends of words:

(60) [sœp]   'yes'   January 26, 2023

(61) [ɑt]   'horse'   January 19 & February 23, 2023

(62) [ɑjmɑχ]   'relative'   February 16, 2023

Alternation Between [s] and [h]

The consonants [s] and [h] appear to alternate in Platon's idiolect. On multiple occasions, Platon would produce [h] in an initial utterance of a word, then replace [h] with [s] in subsequent utterances, as below, where all instances of [s] were produced alternately as [h] at least once during elicitation:

(63) [min utujɑ sitɑbɯn]   'I am sleeping'   February 9 & 23, 2023

(64) [sɑlgɯmːɑ]   'in the air'   February 21, 2023

(65) [sɛttɛ]   'seven'   January 19 & February 16, 2023

Since [h] usually surfaced only in initial utterances and not in subsequent repetitions of the same word, [s] may be characteristic of more careful or formal speech, and [h] may be a dialectical variant of [s]. (Platon has informed us of differences in pronunciation of several words in different Sakha dialects and registers.) But a complicating factor is that /s/ is replaced by /h/ in the following Yakutized Russian loanword (transcription from Wiktionary):

(66) Sakha: [bɑhibɑ]   Russian: [spɐsʲibə]   'thank you'   January 19, 2023

It is unclear if the phonological process at work in (66) is related to the alternations in (63)-(65).

(66) also provides a possible example of **Cluster Reduction**: it appears that the initial [s] in the Russian word for 'thank you' is deleted. Sakha does not allow for consonant clusters within the same syllable (see **Phonotactic Constraints**), but consonant clusters do occur across syllable and morpho-syntactic boundaries, as in the set of examples starting with (86).

Additionally, there may be a phonotactic constraint on [p] appearing word-initially, accounting for the voicing [p] → [b] in the Yakutization of (66). However, there does not seem to be a consistent voicing rule, as the following loanwords keep [p] voiceless:

(67) [pɑkɑ]   'goodbye' (Russian loanword)   February 9, 2023

(68) [pɑɾk]   'park' (English loanword)   February 9, 2023

The next two rules, Backing Harmony and Rounding Harmony, can be illustrated simultaneously through alternations in the vowel appearing in the surface realization of the plural marker, and are therefore presented together.

Backing Harmony

$$\left[\begin{array}{c} +\text{syllabic} \end{array} \right] \rightarrow \left[\begin{array}{c} \alpha\text{back} \end{array} \right] / \left[\begin{array}{c} +\text{syllabic} \\ \alpha\text{back} \end{array} \right] C_0\_ \quad \text{Bounding domain: word}$$

A vowel matches the backness of the first vowel preceding it, within the same word.

Rounding Harmony

$$\left[\begin{array}{c} +\text{syllabic} \end{array} \right] \rightarrow \left[\begin{array}{c} \alpha\text{round} \end{array} \right] / \left[\begin{array}{c} +\text{syllabic} \\ \alpha\text{round} \end{array} \right] C_0 + C_0\_$$

At morpho-syntactic boundaries, a vowel matches the roundness of the nearest preceding vowel.

Evidence that Rounding Harmony is edge-sensitive comes from the following words, which feature rounded and unrounded vowels:

(69) [yœɾɛχ]   'education'   February 9, 2023

(70) [kumɑχ]   'sand'   February 23, 2023

(71) [kutujɑχ]   'mouse'   January 19, 2023

Since each of the above examples features an unrounded vowel following a rounded vowel within a single morpheme, indicating that Rounding Harmony does not apply, it appears that Rounding Harmony only applies at morpho-syntactic boundaries. Now consider the following plural forms, where the vowel in the plural suffix, which has underlying representation /-l$\left[\begin{array}{c} +\text{syllabic} \\ -\text{high} \end{array} \right]$r/, matches the preceding vowel in backness and roundness:

(72) [i] and [ɛ] are both $\left[\begin{array}{c} -\text{back} \\ -\text{round} \end{array} \right]$:   [kɛhi-lɛr]   'people'   January 19, 2023

(73) [ɑ] is $\left[\begin{array}{c} +\text{back} \\ -\text{round} \end{array} \right]$:   [aʁɑ-lɑr]   'fathers'   January 26, 2023

(74) [œ] is $\left[\begin{array}{c} -\text{back} \\ +\text{round} \end{array} \right]$:   [tœbœ-lœr]   'heads'   February 21, 2023

(75) [ɔ] is $\left[\begin{array}{c} +\text{back} \\ +\text{round} \end{array} \right]$:   [dʒɔn-nɔr]   'people'   January 19, 2023

The next set of rules account for alternations in the initial consonant of the plural suffix, which can be realized as [t], [d], [l], or [n] depending on context.

Stop Formation

[l] → [-approximant] / [-syllabic]＿   unless [l]＿＿

[l] becomes a stop when following a consonant, unless that consonant is also [l].

Thus, in the following examples, where the stem ends in a vowel or [l], the [l] in the plural suffix is underlying:

(76) [ili-lɛr]   'arms'   January 19, 2023

(77) [kɯhɯl-lɑr]   'red ones'   January 26, 2023

In contrast, the following examples feature words where the stem ends in a glide or obstruent, and thus the plural suffix begins with a stop derived from underlying /l/:

(78) [kugɑχ-tɑr]   'ears'   January 19, 2023

(79) [bɑj-dɑr]   'rich ones'   February 9, 2023

Voicing Assimilation

[-syllabic] → [αvoice] / $\begin{bmatrix} \text{-syllabic} \\ \alpha\text{voice} \end{bmatrix}$+＿

Consonants at morpheme boundaries take on the voicing of the preceding consonant.

The above rule must be edge-sensitive due to the following example, where [t] fails to become voiced after [l]:

(80) [ɑltɑ]   'six'   January 19, 2023

(78) illustrates Voicing Assimilation. Underlying /l/ first undergoes Stop Formation, followed by devoicing, since /l/ is preceded by [χ], a voiceless consonant.

Nasalization

[-continuant] → [+nasal] / [+nasal]+＿

Stops at morpheme boundaries become nasal following a nasal consonant.

The following examples illustrate Nasalization. First, /l/ in the plural suffix undergoes **Stop Formation**; then, it undergoes **Nasalization** in the environment after [n]:

(81) [miːn-nɛr]   'soaps'   February 16, 2023

(82) [χɑllɑn-nɑr]   'skies'   January 26, 2023

The following plural forms are irregular and violate the Backing Harmony, Rounding Harmony, and Stop Formation rules:

(83) [ʤɑχtɑr]   'woman'   [ʤɑχtɑl-lɑr]   'women'   Expected: *[ʤɑχtɑr-dɑr]   January 19, 2023

(84) [kyn]   'sun'   [kyn-nɛr]   'suns'   Expected: *[kyn-nœr]   January 26, 2023

(85) [uɔt]   'fire'   [uɔt-tɑr]   'fires'   Expected: *[uɔt-tɔr]   January 19, 2023

The above words are all relatively common, and it is not unusual for irregular forms to be those that are most commonly used. It is plausible that these plural forms are simply exceptions to the proposed phonological rules.

Backing Harmony, Rounding Harmony, Voicing Assimilation, and Nasalization can also be seen in the agreement markers in the following predicate nominal and attributive clauses. The agreement marker for the first person singular subject is /-bɪn/, from which [-bɯn], [-bɯn], [-pyn], and [-mʊn] are derived below.

(86) [ytʃygɛj-bɯn]   'I am good'   January 19 & 26, 2023

(87) [mɯn utʃutɑl-bʉn]   'I am a teacher'   February 9, 2023

(88) [mɯn ynkyhyt-pyn]   'I am a dancer'   February 9, 2023

(89) [mɯn ustuʤɔn-mʊn]   'I am a student'   February 9, 2023

In (86), /I/ in first person singular subject agreement marker /-bIn/ is realized as [ɪ], which is [-back, -round], since the preceding vowel [ɛ] is [-back, -round]. Thus this example is consistent with the Backing Harmony and Rounding Harmony rules. Furthermore, /b/ remains voiced since it is preceded by voiced consonant [j], demonstrating consistency with Voicing Assimilation. (87) demonstrates Rounding and Backing Harmony for vowels which are [+back, -round]. (88) demonstrates devoicing, as underlying /b/ becomes [p] following voiceless consonant [t]. In addition to Backing Harmony and Rounding Harmony, (89) demonstrates Nasalization, as underlying /b/ becomes nasal [m] following [n].

Word-Final Lowering

$$\begin{bmatrix} \text{+syllabic} \\ \text{-long} \end{bmatrix} \rightarrow \begin{bmatrix} \text{-tense} \end{bmatrix} / \_C_0\#$$

Short vowels become lax in the final syllable of a word.

Since only the [+high] Sakha vowels are [+tense], **Word-Final Lowering** is only seen applying to [+high] vowels such as /i/ and /u/, as illustrated below.

(90) UR: /i/   SR: [ɪ]   Example: [mɯn]   'I'   January 19, 2023

(91) UR: /u/   SR: [ʊ]   Example: [bœr.dʊk]   'flour'   February 16, 2023

(92) UR: /i/   SR: [ɪ]   Example: [bɑltɪ]   'younger sister'   January 19, 2023

Phonotactic Constraints

The alveolar sonorant consonants /r/, /l/, and /n/, along with /ʁ/, /g/, /ŋ/, /ɾ/, /p/, and /j/, never appear word-initially in non-loanwords. Voiced obstruents never appear word-finally, which is ascribed to **Word-Final Devoicing**.

The following syllable types are attested: V, VC, Vː C, CV, CVː, CVC, and CVː C. V syllables include the following:

(93) [u]   'water'   January 19, 2023

VC and Vː C syllables are illustrated by the near-minimal set for vowels, starting with (31).
The following words illustrate CV syllables, with syllable boundaries marked for clarity:

(94) [sɛ.ɾɪ]   'war'   February 21, 2023

(95) [χɑ.ɾɑ]   'black'   January 26, 2023

(96) [kɛ.hɪ]   'person'   January 19, 2023

Examples of CVː syllables are the words 'grandmother' (47) and 'grandfather' (49).
The following are examples of CVC syllables:

(97) [bɛt]   'nice'   January 26, 2023

(98) [kyn]   'sun'   January 19, 2023

The number 'one' is an example of a CVː C syllable:

(99) [biːr]   'one'   January 19, 2023

Prosody

Stress is consistently word-final. Both rising and falling intonational contours have been observed for declaratives, but elicitations thus far have been limited to citation forms and simple sentences. Further elicitation of spontaneous speech is needed to gather natural prosodic data.

References

Spasibo. 2023. *Wiktionary.* (`https://en.wiktionary.org/wiki/%D1%81%D0%BF%D0%B0%D1%81%D0%B8%D0%B1%D0%BE`) (Accessed 2023-3-3.)

Yakut. 2023. *Ethnologue.* (`https://www.ethnologue.com/language/sah/`) (Accessed 2023-2-26.)