

The Ursai Language

John Sarris Burke

May 30, 2017

Contents

1	Introduction	3
2	Phonology	4
2.1	The Eight Phonemes	4
2.2	Stress	4
2.3	Sonority	4
2.4	Phonotactics	5
2.4.1	Nasals	5
2.4.2	Semivowels	5
2.4.3	Consonant Breaking	5
2.4.4	/k - t/ Realizations	5
2.4.5	/s - h/ Realizations	5
2.4.6	R-colored Vowels	5
2.4.7	Diphthongs	5
2.5	Romanization	5
2.5.1	Common	6
2.5.2	Restricted	6

List of Figures

1	The Eight Phonemes	4
2	Sonorance Hierarchy	4
3	Common Romanization	6

1 Introduction

The Ursai Language is an alien language spoken by the Ursai, a far off extraterrestrial race of spacefaring Aliens in a variety of more jocular side projects of the author. The Aliens, when in proximity of Earth, enjoy abducting cows, convincing humans to build pyramids and other stone structures to make a "World Battery", teaching people how to lay rocks really really good, and other nonsensical activities so that they can have a solid chortle back on the main bridge. All this said, Ursai is a fully fledged, human speakable language, because a corner like this in side projects is a fun diversion and it makes use of it a little more fun and interesting.

Eight Phoneme Challenge Ursai evolves from a Constructed Language Game called the Eight Phoneme Challenge. The game restricts participants to making a language with only eight distinct phonemes, which is far less than any natural language on Earth works with. Any given phoneme may have any number of phonetic realizations; for example, /**p**/ could be [**f**] word initially, [**b**] after nasal sounds, and [**p**] otherwise. There is no real driving reason why an alien group is restricted in such a way; it was done simply for fun.

2 Phonology

Despite being limited to only eight phonemes, Ursai manages to get a fair deal of variety by being liberal with what can be a syllable nucleus, letting individual phonemes get stretched way past their typical sound spaces, and forcing a couple phonotactic rules in. Though not pointing to any particular source language, it does look to many. For the sake of entertainment, when such an inspiration is present, the source will be noted.

2.1 The Eight Phonemes

Phoneme	Ursai IPA	Realization
a - ɛ	/a/	[a] except for Consonant Breaking 2.4.3
i - j	/i/	[j] for semivowel realizaion 2.4.2
u - w	/u/	[w] for semivowel realizaion 2.4.2
r - ɐ	/r/	See Sonority 2.3
m - n - ŋ	/n/	[n] default, Homorganic, see 2.4.1
s - h	/s/	[h] word initially in isolation
k - t	/k/	[t] when adjacent to /s/
p	/p/	only as [p]

Figure 1: The Eight Phonemes

The table left shows all eight phonemes in Ursai with some notes on their realizations which will be expanded in detail later. Furthermore, it defines an "Ursai IPA" for easy explanations through the rest of the document. Ursai is host to three vowels, two of which can be semivowels, a rhotic that has two roles, a highly flexible nasal, one fricative, and two stops, one of which is a fun reference to the Hawaiian Language[1]. More than

half of them can act as a syllable nucleus, which is described in the next section 2.3. Furthermore, the nasal has a variety of realizations depending on environment, which is covered in detail in section 2.4.1.

2.2 Stress

Stress is a point of lexical difference. Single syllable words have no stress, but if they act closely with another word that is multisyllabic, they can shift the stress in that word forward or backward. In multisyllabic words, stress is by default on the penultimate syllable. Some words may have it on the ultimate or, more rarely, the antipenultimate.

2.3 Sonority

The table to the right is a quick reference the the sonority hierarchy in Ursai. The most sonorant sounds are at the top, and sonority decrease as one goes down. It is important to note that /u/ and /i/ are regarded as having equal sonority. Furthermore, both /r/ and /n/ can act as the nucleus of a syllable.

Given that several of the phonemes can act as the nucleus of a syllable, the Sonority Hierarchy defines which ones will be it in a given setting. The phoneme in a given syllable that has the highest sonority will be the nucleus. Others that have sonorant versions in the environment of more sonorant sounds will take on the more consonantal realization. If two sounds are of equal sonority, the last one that can act as the nucleus does so. For example, the sequence /sra/ will always be realized as [sra] and never as [sɐ.a]; /niu/ will always be [nju], /rkuan/ will be [ɐ.kwan], following how nasals function in section 2.4.1, and so on.

Phoneme	Sonorance Notes
/a/	Only as a Vowel
/i/ & /u/	Vowel or Glide
/r/	Sonorant or Consonant
/n/	
/s/	Only as Consonant
/k/	
/p/	

Figure 2: Sonorance Hierarchy

2.4 Phonotactics

2.4.1 Nasals

The Nasal Phoneme in Ursai has the widest variety among all the sounds. It has three realizations, labial - [m], alveolar - [n], and velar - [ŋ]. Furthermore, each of these have sonorant realizations: **m̥**, **n̥**, and **ŋ̥**. By default, this phoneme resolves as [n]. Word initial before a vowel, intervocalically, between a vowel and syllabic /r/, and next to /s/ are such environments. However, in a word final position after a vowel, it is velar, [ŋ]. [ŋ] also arises when adjacent to [k]. Finally, [m] will only arise when adjacent to /p/. Finally, any mutations to the nasal sound must be derived from the word it is a part of; sounds across word boundaries cannot alter it.

2.4.2 Semivowels

The phonemes /u/ and /i/ can become /w, j/ respectively when adjacent to an /a/. In these cases, a diphthong is always formed with /a/ as the heavier element.

2.4.3 Consonant Breaking

Ursai has a strict consonant cluster rule: only two sounds acting as consonants may occur in sequence. This means that a sequence like [skwa] is illegal. No words native to Ursai break this rule internally; however, across word boundaries, clusters of three or four consonants are possible. When this happens, an epenthetic [ɛ] is placed at the beginning of the second word. Thus a word sequence of /unk harna/ is realized as [uŋk ɛ.sã.na]. To note, this was inspired by Arabic[2].

2.4.4 /k - t/ Realizations

Like Hawaiian[1], Ursai features a /k - t/ phoneme, namely it covers what is two very different sounds in English. /k/ arises everywhere except when this is adjacent to /s/.

2.4.5 /s - h/ Realizations

Ursai features a /s - h/ phoneme, where it is [s] always except when word initial before a vowel. /s/ also mutates [k] into [t]. This was inspired by Ancient Greek[3] in part because it developed initial aspiration from a historic /s/ that was weakened.

2.4.6 R-colored Vowels

Ursai's rhotic, /r/, will cause an immediately preceding vowel to become rhotic, similar to some varieties of Mandarin[4]. As a result, there are three possible rhotic vowels: [ã], [ĩ], and [ũ]. All of these count as a VC sequence and contribute to clusters that may need to be broken [2.4.3].

2.4.7 Diphthongs

The two semivowels [j] and [w] work to form the diphthongs in Ursai. The most common ones are the ones that have /a/ at the core, though /i/ and /u/ as a core happen as well. It should be noted that the semivowels count as a full consonant in clusters and their use can trigger consonant breaking [2.4.3].

2.5 Romanization

Unlike many constructed languages, the romanization system for Ursai does not aim to be phonemic, but rather aims for clarity. Using a phonemic scheme, while possible, forces the readers of the romanized system to first internalize the synchronic sound change rules and phonotactics, and then to correctly

decode the sounds from the phonemic system on the fly. Since Ursai has so few sounds, instead a more phonetic system is used so that the reader may read more naturally at the price of a little lexical transparency.

2.5.1 Common

Graph	Sound or use	Phoneme	Notes
⟨a⟩	[a]	/a/	
⟨e⟩	[ɛ]	/a/	Consonant Breaking Only
⟨i⟩	[i]	/i/	Vowel Only
⟨j⟩	[j]	/i/	Semivowel Only
⟨u⟩	[u]	/u/	Vowel Only
⟨w⟩	[w]	/u/	Semivowel Only
⟨r⟩	[ɾ]	/r/	Consonant Only
⟨r̥⟩	[ə̞]	/r/	Sonorant Only, Optional
⟨n⟩	[n]	/n/	Consonant Only
⟨n̥⟩	[n̥]	/n/	Sonorant Only, Optional
⟨m⟩	[m]	/n/	Consonant Only
⟨m̥⟩	[m̥]	/n/	Sonorant Only, Optional
⟨g⟩	[ŋ]	/n/	Consonant Only
⟨g̥⟩	[ŋ̥]	/n/	Sonorant Only, Optional
⟨s⟩	[s]	/s/	
⟨h⟩	[h]	/s/	
⟨k⟩	[k]	/k/	
⟨t⟩	[t]	/k/	
⟨p⟩	[p]	/p/	

Figure 3: Common Romanization

mally a little tricky to get easy access to anyhow. Optionally, next to /**k**/ in can remain as simply ⟨**n**⟩ since it will be clear from an English background.

The remaining consonants are taken from the Full IPA based on the strict phonetic realization. This is done since some of them can become so different from the original phoneme and ease is the target rather than phonemic transparency.

A final point is stress. Words with no stress marking follow the patterns that are considered default in 2.2. However, when a sound does not follow this, acute and grave accents are used. An acute accent, such as is ⟨**kanú**⟩, tells the reader that the stress is on the marked syllable and that it is where the stress for this word falls naturally. A grave accent, such as in ⟨**ikpanì**⟩ is indicative of where the stress is to be said, but also indicates that it has be shifted by some other word or element from outside it to that syllable.

2.5.2 Restricted

Realizing that in certain settings not all the glyphs used above may be available, digraphs may be substituted for the sonorant consonants. A digraph of ⟨**oC**⟩ where **C** is the sonorant is permissible since **o** is unused and can be considered as a garbage letter. Furthermore, in such a conext, all accent marking for stress may be ignored, though, if possible, some other means should be availed to try to clarify where it may be since unmarked vowels and sonorants also indicate a certain expected stress pattern.

The common mode of romanization is enumerated in the table to the left. The different letters are reflective of both the phonetic realization and then their use.

For letters that in English represent non-vowels acting as sonorants in Ursai, an underdot, such as **ṛ**, represents that letter as a sonorant. It is permissible, but not preferred, to skip the underdot in this way.

⟨**e**⟩ is a weakened form of /**a**/ that is present only when consonant breaking [2.4.3] occurs. It is written explicitly so that readers won't have to remember a phonotactic rule while reading, especially aloud. ⟨**e**⟩ is used to indicate that it is only a linking sound not present except for phonological rules.

The semivowels are represented as they are in the full IPA. This is to differentiate it from the full vowel forms even though the semivowels will only appear next to /**a**/ or another one of the vowels that can become a semivowel.

⟨**g**⟩ is used for the velar nasal since it would be unused otherwise and **ṅ** is nor-

References

- [1] Hawaiian Phonology, *Wikipedia*, 3 May 2017, https://en.wikipedia.org/wiki/Hawaiian_phonology
- [2] Arabic Phonology, *Wikipedia*, 25 May 2017, https://en.wikipedia.org/wiki/Arabic_phonology#Phonotactics
- [3] Ancient Greek Phonology, *Wikipedia*, 26 May 2017, https://en.wikipedia.org/wiki/Ancient_Greek_phonology#Fricatives
- [4] Erhua, *Wikipedia*, 26 May 2017, <https://en.wikipedia.org/wiki/Erhua>