SYLLABIFICATION RULES IN PASHTO

JEHANGIR ZAMAN KHAN

ABSTRACT

This paper will cover the discussion about the syllabification rules in Pashto. There are lots of ways of arranging of syllables in Pashto but some constraints are also there so that some syllables do not start words. The paper is discussing what pattern of consonant and vowel can be followed by particular patterns and why some patterns cannot occur in continuation. These points will be discussed in this paper in detail. Pashto is one of the East Iranian groups of languages, which includes, for example, Ossete (North Ossetian, south Ossetian, Caucasus Russian Republic) and Yaghnobi (Tajikistan). Syllabification rules and algorithms for that rules for Pashto are proposed   in this paper. This paper deals with the syllabification of Pashto and syllabification in general.

1. INTRODUCTION

Pashto is mostly spoken in the Afghanistan, Pakistan and Iran. There are two major dialects of Pashto: Western Pashto spoken in Afghanistan and in the capital, Kabul, and Eastern Pashto spoken in northeastern Pakistan. Most speakers of Pashto speak these two dialects. Two other dialects are also distinguished: Southern Pashto, spoken in Baluchistan (western Pakistan and eastern Iran) and the one in Kandahar, Afghanistan (Afghan-network).

In syllabification the location of some syllable in word is also of great importance. Syllable shows different behavior in a form of articulation features and pronunciation if it occurs in word initial, final or intermediary position. Similarly a phoneme place in the syllable also effects the duration and pronunciation of the phone. We can achieve syllabification by looking at different patterns of syllables meaning that we can experiment with monosyllabic, disyllabic and polysyllabic words. In syllabification, the location of syllable in polysyllabic words is of extreme importance because changing the position of some syllable will change its properties (e.g. pronunciation). So determination of boundary of syllables in a word is of great significance.

Generally, nucleus (vowel or diphthong) is proceeded and followed by consonant (onset or coda) or both so it will look like (C*VC*) but both the C's in the given pattern are optional. The asterisks before and after the vowel conforms that there might be any number of consonants before and after the vowel in syllable but there are some limitations as well that tells the validity of a word in a language.

2. LITERATURE REVIEW

Canonical rules and algorithms have been established for the syllabification and are applicable to most of the languages around the globe for syllabification. These rules and algorithms suggest: The Maximum Onset Principle (MOP) always like to place the consonant or consonant cluster in the onset of syllable whenever it is possible to place it in the onset but in Maximum Coda Principle (MCP) we place consonants in the coda up to the possibility.

In Sonority Sequencing Principle (SSP) the sonority of the syllable is increasing (in onset)[1] towards the nucleus and decreasing after (in coda)[2] (Kenstowicz, 1994). Given these rules the sonority is at the peak at the nucleus, increasing in onset and decreasing in coda.

Hogg (1987, p. 53) suggested that in certain cases, a consonant becomes member of both the onset and the coda of two syllables. That consonant is said to be a geminate and the process is known as gemination.

3. METHODOLOGY

3.1 Subject

Data was obtained from the native speakers of Pashto and also from the dictionary (Khan, 1990) of Pashto and was syllabified by four different speakers (including the Author)[3] of different ages but all the test was done on male speakers because no female was available for the experiment(pronunciation). On the basis of these experiments almost all the syllable templates available in Pashto were acquired. About 150 monosyllabic, disyllabic and polysyllabic words were experimented to find out the syllable categories in Pashto. For the words of Pashto see the Appendix.

3.2 Procedure

All the four people were asked to pronounce the words of Pashto, which were selected for the analysis and were containing almost all the syllable templates. The boundaries of syllable were marked according to the pronunciation of the team. There were some words that were pronounced with different syllable patterns by different speakers. Examples are in Table 1.

Table 1 contains Pashto words and the pronunciation of two members of the team that were helping the author in the research.

TABLE 1 Data recorded for the analysis

Words	Member A	Member B
mə̃gaz	mə̃gaz	məngaz
zəmbik (galaxy)	zəmbik	zənbik
xanadani	xanadani	xandani
suk	suk	sok
umbarək	umbarək	ombarək

3.3 Data Recording and Analysis

Data was recorded for analysis and to remove ambiguities. This is shown above in Table 1. The voice was recorded and processed on Praat and X-waves. By observing formants through spectrogram, controversies about syllable structure and boundary were removed.

The equipment used in recording was a mic with a frequency response of 80 to 14,000 Hz with a 16 bit sound card and high quality speakers.

4. RESULTS

The already presented syllabification methods and rules like sonority sequence and maximum onset principal were used with Pashto language to know about its syllabification. This tells us how syllables are restricted by syllable sonorant and obstruents.

It is observed that the groups of two or three consonants can start a word like s, sp, st, sk, sxw but some consonant clusters like st and sk can never occur at the middle of the word.

4.1 Syllable Structure

Pashto allows at most three consonant clusters in the onset and at most two in the coda. 'ə' has a special nature of existing without onset and coda consonant. Words having three consonants in onset normally have no or at most one coda.

Table 2: Syllable Template for Pashto

Examples	Syllable template	Pashto Representation
ə	V	He
əs	VC	Horse
ḥəjasaṱəl	CV	Care for grace
Dəz	CVC	Fire
ɣla	CCV	Stealing
mərg	CVCC	Death
ʃʈək	CCVC	Spontaneous Problem
sxwe	CCCV	Bull
sxwen	CCCVC	Fat guy

The possible templates of Pashto along with its words are given in the Table 2.

4.2 Syllable Restriction

In Languages there are always restriction whether some consonant or class of consonant can occur in the coda or onset of syllables and in which position and order it can occur.

The Principle of Maximum Onset (MOP) preference is followed by the Pashto language e.g. in the word "prewɑtal", all the possible words consonants goes to the onset of the syllable so the syllable boundary becomes as "Pre.wɑ.tal" Similarly Pashto also satisfies Sonority Sequencing Principle (SSP) as well, like in all the syllables of Pashto the sonority is at its peaks on the nucleus and is increasing in onset from border to the nucleus and sonority is decreasing in coda towards the syllable boundary.

The Principle of Maximum Coda (MCP) is not satisfied for the Pashto syllabification. In the analysis few facts were found that if there is just one consonant in the onset then there is no restriction on it, it can be any consonant. In case of more than one consonant in onset, there are some constraints that are second of the two will always be an approximant or if it is a stop then that must be followed by a fricative. The case of 'ʃɽ̃' is also a special case in Pashto. This tells that a fricative can be followed by a retroflex in the onset of Pashto syllables. Examples are given in the Table-D.

Table 3: Onset information in Pashto syllables.

Class	Cluster of onset Consonants	Actual Word
liquid+glide[5]	lw	lwedəle
Stop+glide	tr	tror
Frecative + stop	xk	xkule
Stop+glide	pr	pregdəl
Fricative + stop	sp	spe
Fricative + retroflex	ʃɽ	ʃɽɑk
Fricatives + Approximants	ɣw	ɣwɑxe

If there are two consonant in coda then the boundary consonant must have to be a stop but in very special cases it could be retroflex. The consonant adjacent to the nucleus are usually glides, fricatives and nasals. Examples are given in the Table 4.

Table 4: Coda information in Pashto syllables

| Class	Cluster of coda Consonants	Actual Word
nasal+stop	nd	bərbənd
glides+stop	rg	mərg
Nasal+Fricative	nɽ	ʈɑnɽ

The above Tables shows that the Pashto satisfies the Maximum Onset Principle (MOP) and Sonority Sequence Principle (SSP) with the way that sonority increases in onset from boundary and decreases in coda to the boundary. This is violated in just one example of "xkule". It is a special case in Pashto where extra syllabification is being done.

4.3 Gemination

In Pashto gemination takes place when a consonant comes in the middle of two short vowels. As in the word "aga" when this 'g' comes in the middle of two 'a' it geminates itself to be the coda of the first and onset of the second syllable. The reason for this is that a short vowel mostly has a consonant following it and in cases when there is another short vowel next to the consonant the consonant becomes onset and coda of both vowels causes gemination.

4.4 Syllabification through Templates

From the analysis it is noticed that Pashto is directed from left to right.
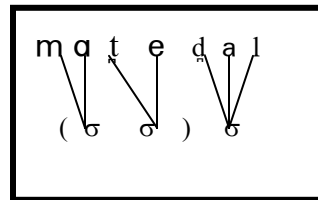

mɑteɖal (being broken)
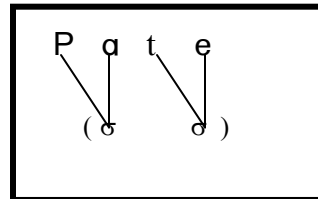
pɑt e   (hidden)

So to check whether a word lies in Pashto language the template will be matched from Left to Right. The validity of any word will be confirmed by checking it with the given template. To find the number of syllables in the word, possible template pattern will be checked and will tell us how many syllables are in that word.

The above two are the examples of finding syllables in a word. That helps us to find the direction of some language.
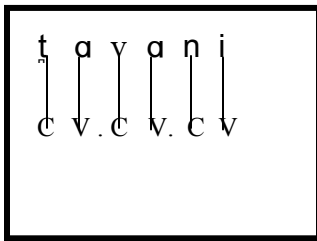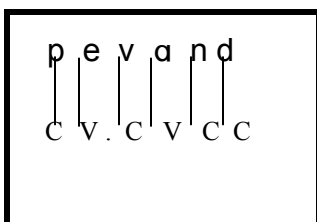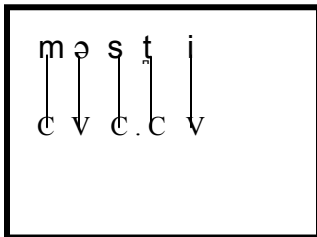

ṭavani (to harm)

The above example gives us the possible templates in a word. The algorithm for syllabification is given in the next part.
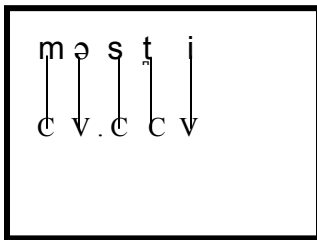
məsṭi(having fun)

pevand(binding)

4.5 Algorithm for Syllabification

To find out the membership of a word in Pashto there are two ways. Both can be applied to it but the one suggested in this paper is more suitable because this is purely based on Pashto. This method is know as the template matching method in which given a word, we have to match it with the given pattern of syllable and will decide about its membership. The other one is the conventional method(which is not being discussed) using MOP with SSP.

In the above example of "məsṭi" there are two possible ways to make its syllables but we will prefer the pattern cvc.cv rather than cv.ccv because we mentioned in the gemination part that every short vowel is always followed by a consonant. So when ever there is a short vowel we will select the pattern that have at least one consonant in the coda and if there is a long vowel then we will lookup to next syllable before deciding about the one that is under consideration and will decide about the pattern considering MOP and SSP. As in the case of "pevand" we can have two option as cvc.vcc and cv.cvcc so we will select the second one due to the mentioned properties.

4.6 Template Matching Algorithm for Syllabification

1. Match Syllable template from left to right with the word whose validity as Pashto word is checked.

2. If more than one templates matches for the same word then there are two possibilities. If nucleus is a short vowel, then add the upcoming consonant in coda i.e. CVC template should be followed. Otherwise, if it is a long vowel then SSP is checked for the pattern selection.

3. Match all the possible templates available for the give word. Matching will be stopped when all syllables will be matched or there is no match in the template.

4. If a word fits in a template then it is a valid Pashto word else it is not.

5. DISCUSSION

There were lots of controversies among the native speakers while working on the syllabification of words. These have been resolved with the acoustic analysis of the words. The analysis of the word "mɑngɑz" was actually "mãgɑz" and there is no nasal 'n' there. There are some words that break the rule of SSP but they are still words of Pashto like "spe" and "xkule". In these word the 'sp' 'xk' are onset and in onset there must always be increasing sonority towards the nucleus. But it fails in this case so these are considered as exceptions in Pashto.
There are very few words that can have three consonants "sxwa", but some people think they are not the words of Pashto. This is still a controversial issue.

6. REFERENCE

Khan K. 1990 Pashto Urdu Dictionary Compiled by: Pashto Academy University of Peshawar.

Kenstowicz, M. 1994. Phonology in Generative Grammar; Cambridge: Backwell Publishers. (250-298)

Hogg, R. & McCully, C. B. 1987. Metrical Phonology: A Course book; Cambridge: Cambridge University Press. (31-60).

Masica, C. 1991. The Indo-Aryan Languages; Cambridge: Cambridge University Press.

Kiraz, G and Bernd M. "Multilingual syllabification using weighted finite-state transducers"; Bell labs, Lucent

Technologies, Murray Hill, NJ 07974 USA. Afghan-network.org
www.Afghan-network.org.

Donna J.N 1996. *Linguistic An Introduction* chap#2.