SYLLABIFICATION RULES IN PUNJABI

HASAN KABIR

ABSTRACT

This paper describes an approach to syllabification that has been incorporated into Punjabi. It focuses on the syllabification rules i.e. how many syllable categories exist in Punjabi, syllabification constraints i.e. which phonemes cannot come in the start of a syllable, etc. and applicability of different syllabification algorithms on Punjabi. An algorithm for syllabification is also devised in the end. Previously most of the studies in this regard have been done on Western Languages. This paper also covers all such related details.

1. INTRODUCTION

Punjabi is a widely spoken language. Punjabi being originated from Proto Indo (Indo Aryan) languages has many distinct characteristics. The accent (pronunciation) of Punjabi varies from region to region. Punjabi being spoken in Indian Punjab is much different from the Punjabi being spoken in Pakistani Punjab as proposed by Masica (1991, 20). This paper discusses the Syllabification Rules of Punjabi spoken in Pakistan.

Syllabification is an important component of any TTS (Text to speech) system. In many languages the pronunciation of phonemes is a function of their location in the syllable relative to the syllable boundaries. Location in the syllable also has a strong effect on the duration of the phone, and is therefore a crucial piece of information for any model of segmental duration; (Kiraz, 1).
Syllabification can be achieved by writing a declarative grammar of possible locations of syllable boundaries in polysyllabic words. An extremely simplistic, constraint-based model of syllabification might state that each word in the utterance consists of one or more syllables of the structure C*VC*, i.e., of an obligatory syllable nucleus (V) optionally preceded or followed, or both, by any number of consonants (C).

2. LITERATURE REVIEW

I never encountered any authentic document on Punjabi syllables during the course of study. Therefore, this document could act as a baseline for the further research in this area.

Goldsmith (1990, 110) suggests that sounds differ in sonority. A low vowel is sonorous than a high vowel and any vowel is more sonorous than a consonant. Successive peaks and valley of sonority therefore characterize a sequence of sounds in a normal utterance. The sounds, which constitute the peaks of sonority, are called syllabic and an utterance has as many syllables as it contains syllabic word.

Different theories and rules have been established for syllabification. The ideas behind them are listed below.

2.1 Existing Algorithms

Maximum Onset Principle (MOP) prefers onset to codas. It prefers onset wherever possible; (Goldsmith, 1990, 137). In contrary, Maximum Coda Principle (MCP) is absolutely opposite to MOP.

Sonority Sequencing Principle (SSP) requires onset to rise in sonority towards the nucleus and coda to fall in sonority from the nucleus; (Kenstowicz, 1994, 254). Given this rule, the sonority of the syllable thus peaks at the nucleus and descends toward the margins.

Ambisyllabicity is the process of making intervocalic consonants member of both syllables as proposed by Hogg (1987, 53).

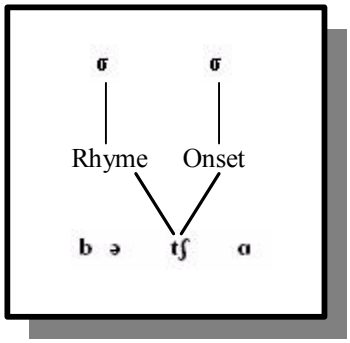

FIGURE 1 Ambisyllabication Process

Templatic syllabification algorithm suggests that a grammar defines a template that characterizes the gross prosodic shape of the maximal core syllable. If a language has several epenthesis rules and if epenthesis consists in matching to the syllable template, then the left-to-right/right-to-left direction of template mapping should be fixed for a given language and mirror its direction of syllabification as proposed by Kenstowicz (1994, 270).

Resyllabification suggests that syllable boundary changes when prefix/suffix is added or deleted as proposed by Kenstowicz (1994, 280) e.g.
(2.1)
C.V → .CV

GoldSmith (1990, 120) suggests that resyllabification sometimes change the length of a vowel coming in a syllable e.g.
(2.2)
zəbɑːn → zəbɑni

Applicability of these principles on Punjabi is discussed in detail in Results section.

3. METHODOLOGY

3.1 Subjects

For the purpose of finding syllable categories, data were obtained from two women and three men, aged 18 – 45, none of whom exhibited any evidence of speech or vocal pathology. Furthermore, to confirm data, two established dictionaries, (Bukhari, 2000) & (Punjabi, 1999), of Punjabi language were consulted.

Author being the native speaker of Punjabi, was also a member of that group for finding out all possible syllable templates.

3.2 Procedure

All subjects were asked to read from paper a list [1] of Punjabi words, compiled by the author, (containing almost all the possible syllable templates) and point out the syllable boundaries.

3.3 Data Recording and Analysis

For few of the controversies among the native speakers, acoustic analysis of the speakers was carried out on Praat 3.9; speech-processing tools designed for Windows users. The equipment consisted of a 2K-ohm microphone attached with 32bit sound card and two high quality Panasonic 2400 Watts speakers.

4. RESULTS

This was a preliminary study designed to learn the syllabification rules in Punjabi. Several different procedures were investigated in order to gain a general impression of how sonorant and obstruents compete for a place in syllable. Last vowel in open syllable word always is a long vowel in Punjabi.

4.1 Syllable Structure

The phonetics of Punjabi allows complex consonant clusters in both the onset and the coda of syllables (might use extra syllabification as in ʃkæt̪). The maximum number of consonants in the onset is 2 whereas it allows up to 2 coda consonants. Thus, the maximum number of consecutive consonants across syllable boundaries is 4. List of all possible syllable templates along with the examples and constraints is provided in *Table 1*[2].

[1] For list of words see Appendix
[2] VV shows long vowel & diphthong and single V shows a short vowel

TABLE 1 Syllable Templates

Example	*Syllable Templates*	*Comments*
ʈə.ɽək.ɽi	CV	Not allowed at word final position
pɑ͡i	CVV	---
pək	CVC	---
drɑ	CCVV	---
tʃənd	CVCC	---
pɑn	CVVC	---
pɑnd	CVVCC	---
d̪rəxt̪	CCVCC	---
ʃkæt̪	CCVVC	---
oe	VV	Can not precede by a syllable
os	VVC	Can not precede by a close syllable

4.2 Syllable Constraints

Certain restrictions exist as to which consonants, or classes of consonants, can occur in which position within the onset or coda of a syllable. For instance, in Punjabi no phones other than obstruents can occur before an obstruent in the onset. In codas, after the first obstruent no phones other than obstruents can occur in the coda. Examples for the longest consonant clusters in Punjabi onsets and codas are given in *Table 2a & Table 2b*, respectively.

TABLE 2A Consonant Clusters in Onsets

Onset		
Class	Cluster	Example
stop+trill	d̪r	d̪rəxt̪
fricative+stop	ʃk	ʃkæt̪

TABLE 2B Consonants Clusters in Codas

Codas		
Class	Cluster	Example
fricative+stop	xt̪	d̪rəxt̪
nasal+stop	nd	tʃənd
trill+stop	rk	k^hʊrk
Retroflex+stop	ɽk	k^həɽk

Sonorants (nasals, liquids, and glides) can only occur adjacent to the syllable nucleus. This pattern is referred to as the sonority principle, which ranks phone classes according to their natural acoustic sonority, which in turn is a correlate of the degree of constriction of the vocal tract; (Kiraz, 2).

As mentioned earlier, there can be complex codas (with one or more consonants) and complex onsets in Punjabi syllables. However, there are limitations in formation of these complex onsets and codas. Firstly, Maximum Onset Principle (MOP), which prefers onset on codas, must be satisfied. Secondly, Sonority Sequencing Principle (SSP), which requires onset to rise in sonority towards the nucleus and coda to fall in sonority from the nucleus, must be satisfied. Due to this, Maximum Coda Principle (MCP) fails here. However, [sC[3]] & [ʃC] clusters violate the principles we have claimed to govern the structure of Punjabi onsets e.g. in ʃkæt̪, /ʃ/ is more sonorous than /k/ and according to SSP this word is not valid. This problem can be resolved by declaring this extra syllabification. In general, we can account for the exceptional behavior of [ʃ] & [s] if we postulate a special rule adding [s] to the onset i.e. declaring [sC] & [ʃC] as valid clusters. See GoldSmith (1990, 143), Kenstowicz (1994, 258) for further detail.

In addition, when there are two coda consonants, the first consonant in coda is limited to a voiceless fricative, trill, retroflex or nasals i.e. /f/, /s/, /ʃ/, /x/ or /r/ or /ɽ/or /n/, /m/, *provided extrasyllabification is not being done*[4], and second consonant is limited to a stop[*].

There may be more restrictions on constriction of these complex onsets and codas. More research needs to be done to

[3] C denotes consonant having week sonority then s & ʃ

[4] e.g. nʊks is a valid word but here s is extra syllable as it is more sonorous than k. Therefore, we would not take them as two consonants in coda.

[*] Referred to Hussain(1997, 42), who did this with Urdu perspective.

determine complete phonotactic constraints on syllable constriction in Punjabi.

4.3 Templatic Syllabification

Punjabi is a right-to-left language. It is clear from the data in *Figure 2* & *Figure 3* by considering their feet. Punjabi always makes degenerate foot on the left side[5]. This shows that templates would always be matched from the right hand side to check validity of a word or to find number of syllables in a word.

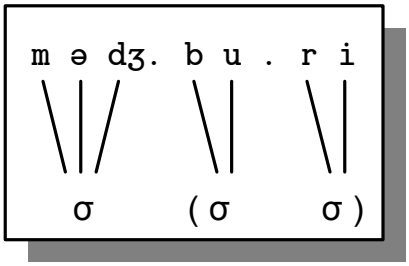

FIGURE 2 Feet Formation Process in Punjabi (مجبوری)

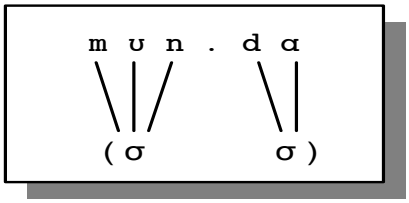

FIGURE 3 Feet Formation Process in Punjabi (منعا)

Matching it with templates checks validity of a word. If at any stage word do not match with a template then that is an invalid word of a language.

Again to find syllables in a word, templates are matched with a word and the templates found at the end cumulatively conforms the total syllables e.g. ʈəvɑɖe has three syllables because of three templates participation.

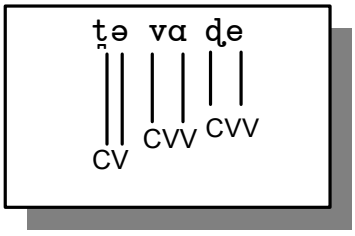

FIGURE 4 Template Matching Process from Right to left

[5] Only two examples are shown here. But this is true for all the data. It can be seen from words given in the appendix by doing foot formation.

Template syllabification is applicable to Punjabi but with few restrictions. An algorithm is provided for this purpose.

4.4 Templatic Syllabification Algorithm

1. Match syllable template from right to left.
2. In case of tie among two templates, the one with maximum onset would be preferred, provided SSP is being satisfied e.g. while matching from right to left, we have three possible templates in *Figure 5* i.e. VV, CVV & CCVV but CVV would be applied because of MOP and SSP. In this example, CVV would be preferred over VV because it is satisfying both MOP and SSP. However, CCVV would be rejected because it is satisfying MOP but refusing SSP.

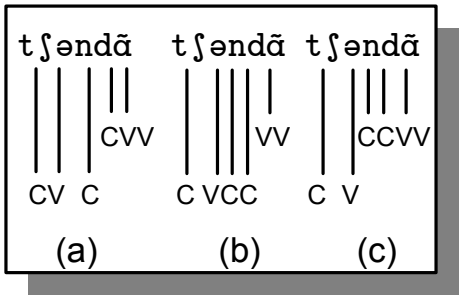

FIGURE 5 Syllabification Possibilities

Therefore, *Figure 5(a)* shows the valid syllable boundaries.

3. Stop matching templates when word is either finished or no more template matches with the syllable in word.
4. If templates cumulatively satisfy a word then that is a valid word of a language.
5. Total number of templates used to satisfy a word is the actual number of syllables in the word.

4.5 Ambisyllabicity

Gemination is applicable whenever consonant is preceded by a short vowel and followed by a vowel. However, ɽ is an exception e.g. in ʈə.ɽək.ɽɪ gemination rule is applicable but Punjabi do not allow two consecutive retro flexes. Also, in case of an aspirated consonant, gemination would be applicable but consonant in first syllable

would become unaspirated e.g. lətʃʰən would become lətʃ.tʃʰən.

The basic reason of gemination is that short vowel always requires a consonant to follow it and if that consonant is an onset of the next syllable then it causes gemination. Therefore, open syllable with short vowel in word final position is not allowed in Punjabi. However, short vowel could come at start of a syllable as well as of word.

Gemination process is applicable after templatic syllabification has been done because templatic syllabification gives the syllable boundaries. Ambisyllabicity process considers adjacent syllables for gemination.

4.6 Resyllabification

Resyllabification is also possible in Punjabi e.g. tʃənd is a single syllable but when ã is added, it becomes tʃən.dã; that is the syllable boundary changes. The reason is that Punjabi follows MOP. Syllables might change their boundaries whenever an affix is added/deleted in a word in Punjabi.

5. DISCUSSION

Currently, there are two competing hypothesis which attempt to explain the validity of a word in Punjabi. Also, they help us to find the possible number of syllables in a word. SSP along with the MOP does the same job as templatic syllabification. But the flaw in SSP and MOP is that they would validate the words which are not the member of Punjabi but satisfy the MOP and SSP principles e.g. CCV is not a valid template of Punjabi syllable but SSP & MOP will validate a syllable, if it satisfies them. Therefore, templatic syllabification is much easier to cope with i.e. we would have to write fewer rules for it. Templatic syllabification algorithm is provided in Results section.

During the course of study, various controversies existed among the native speakers. These were mostly about gemination and extrasyllabification. These controversies were resolved by recording and then acoustic analysis of these words. I recorded three persons speaking ʃkæt̪ and found out that vowel does not exist between /ʃ/ and /k/ showing that this is a consonant cluster. Earlier, few native speakers suggested ʃə.kæt̪.

kʰəɽk was also recorded to analyze that whether it is kʰəɽk or kʰəɽək and found out kʰəɽk was the actual word depicting that /ɽ/ can come at second last place in a syllable.

6. REFERENCES

Bukhari, T. 2000. *Punjabi Urdu Lughat*; Lahore: Urdu Science Board

Comric, B. 1990. *The major languages of South Asia, The Middle Eash and Africa*; London: Routledge. (26)

Goldsmith, J. A. 1990. *Autosegmental and Metrical Phonology*; Cornwall: T. J. Press (Padstow Ltd). (103-168)

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

Hussain, S. 1997. “Phonetic correlates of lexical stress in Urdu”; Unpublished Ph.D. dissertation, North Western University, IL, USA.

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

Kiraz, G. A. & Mobius, B. “Multilingual syllabification using weighted finite-state transducers”; Bell labs, Lucent Technologies, Murray Hill, NJ.

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

Punjabi, Irshad Ahmed. 1999. *Punjabi Lughat*; Lahore: Punjabi Adbi Lehar.

Radford, A. 1999. *Linguistics: An introduction*; Cambridge: Cambridge University Press.

7. APPENDIX

TABLE A Words starting with consonant

Example	*Possible Syllable Clusters*	*Punjabi*
dra	CCVV#	ڑ را
lə̃g	CVC#	لنگ
mədʒ	CVC#	مَج
nĩ	CVV#	نیں
rəb	CVC#	رب
lur	CVVC#	لُور
nal	CVVC#	نال
næʃ	CVVC#	نیش
tʃənd	CVCC#	چنڈ
pand	CVVCC#	پا نڈ
nʊks	CVCC#	نُقص
mʊr.ɖa	CVC.CVV#	مُردا
mər.ɖa	CVC.CVV#	مَردا
bən.ɖa	CVC.CVV#	بندا
bənɖ.ja	CVCC.CVV#	بَند یا
vək^{h}.ra	CVC.CVV#	وَکھرا
ka.ki	CVV.CVV#	کا کی
kʊ.ɽi	E-CV.CVV#	کڑی
vĩga	G-CVC.CVV#	وِنگا
ma.ɽa	CVV.CVV#	ماڑ ا
dran.ɖa	CCVVC.CVV#	ڑ راندا
pran.ɖa	CCVVC.CVV#	پراندا
tʃə̃ga	G-CVC.CVV#	چنگا
ʈʊsi	G-CVC.CVV#	ٹُسی
p^{h}ɪka	G-CVC.CVV#	پِھکا
pəvẽ	G-CVC.CVV#	پویں
dʒəp^{h}i	G-CVC.CVV#	جپھی
gʊdi	G-CVC.CVV#	گُڈی
pət^{h}e	G-CVC.CVV#	ھٹ
dʒəvai	G-CVC.CVV#	جوائی
pu.ʃəl	CVV.CVC#	پوشل
nəsəl	G-CVC.CVC#	نسل
mə̃g.ʈa	CVC.CVV#	منگتا
tʃɪɽ.j ã	CVC.CVV#	چڑ یاں
kə̃g.ɽi	CVVC.CVV#	کا نگڑی
mam.la	CVVC.CVV#	ما
mən.dʒi	CVC.CVV#	منجی
kæn.tʃi	CVVC.CVV#	قینچی
mʊn.da	CVC.CVV#	منڈا
nal.ɖa	CVVC.CVV#	نالدا
vər.ga	CVC.CVV#	ورگا
poʈ.ra	CVVC.CVV#	پوترا
vək^{h}.ni	CVVC.CVV#	ویکھنی
lɪʈər	G-CVC.CVC#	لِتر
lətʃ^{h}ən	G-CVC.CVC#	لچھن
lɪtʃəʈ	G-CVC.CVC#	لیچڑ
tʃəper	G-CVC.CVVC#	چپیڑ
tʃəvəl	G-CVC.CVC#	چّول
kɪɖər	G-CVC.CVC#	کِد ر
də̃gəl	G-CVC.CVC#	د نگل
ɖea.ɽi	CVV.CVV#	د یاڑی
pʊʈər	G-CVC.CVC#	پُتر
kʊtʃ^{h}əʈ	G-CVC.CVC#	کُچھڑ
pək.ɽoʈ	CVC.CVVC#	پکڑ وڑ
kət.vəd^{h}	CVC.CVC#	کٹ وڑھ
lədʒ.pal	CVC.CVVC#	لج پال
nin.ɖər	CVVC.CVC#	نیند ر
məm.nun	CVC.CVVC#	ممنون
vəndʒ.li	CVCC.CVV#	ونجلی
ʈəva.ɖe	G-CVC.CVV.CVV#	تواڑے
kəra. ra	G-CVC.CVV.CVV#	کرارا
fɪte.nək	G-CVC.CVV.CVC#	ٹ نک
gəvan.di	G-CVC.CVVC.CVV#	گوانڈی
kəbaʈ.ja	G-CVC.CVVC.CVV#	کباڑ یا
mədʒ.bu.ri	CVC.CVV.CVV#	مجبوری
məz.du.ri	CVC.CVV.CVV#	مزد وری
pətʃ.vɪn.dʒa	CVC.CVC.CVV#	پَچونجا
səʈ.ja.nas	CVC.CVV.CVVC#	ستیاناس
vɪtʃ.kar.li	CVC.CVVC.CVV#	وچکارلی
bəɣæ.rəʈ	G-CVC.CVV.CVC#	بغیرت
ma.tʃəs	CVV.CVC #	ما چس
kʊru.li	G-CVC.CVV.CVV#	کُرولی
ʈə.ɽək.ɽi	E-CV.CVC.CVV#	ٹڑکڑی
bɪp^{h}ər.ja	G-CVC.CVC.CVV#	بپھریا
gəvatʃ.ja	G-CVC.CVVC.CVV#	گواچیا
pa.kəs.ʈan	CVV.CVC.CVVC#	پاکستان

TABLE B: Words starting with vowel

Example	*Possible Syllable Clusters*	*Punjabi*
oe	VV#	اوے
os	VVC#	اوس
o.ḍɑ	VV.CV#	اودا
əsi	G-VC.CVV#	ا سی
o.k^{h}ɑ	VV.CVV#	اوکھا
əmɑ̃	G-VC.CVV#	ا ماں
æ.ḍər	VV.CVC#	اےد ر
ɑk^{h}.jɑ	VVC.CVV#	آکھیا
ɑ.lə.nɑ	VV.CV.CVV#	آ لنا
ɑ.p^{h}ər.nɑ	VV.CVC.CVV#	آپھرنا
u.pər.lɑ	VV.CVC.CVV#	اُ وپر
ʊbɑk.nɑ	G-VC.CVVC.CVV#	اُباکنا
ɑ.kɑ.bɑ.kɑ	VV.CVV.CVV.CVV#	آکاباکا
ɑ.ʈəʃ.bɑ.zi	VV.CVC.CVV.CVV#	آتش‌بازی

G: Gemination
E: Exception for Gemination