

## Comparative Study of Standard Punjabi and Malwai Dialect with regard to Machine Translation

Harjeet Singh<sup>1</sup>, Ravinder Khanna<sup>2</sup>, Vishal Goyal<sup>3</sup>

<sup>1</sup>Research Scholar, Punjab Technical University, Jalandhar, India

<sup>2</sup>Professor, Department of Electronics, M.M. University, Ambala

<sup>3</sup> Assistant Professor, Department of Computer Science, Punjabi University, Patiala

### Abstract

Punjabi language is an modern Indo-Aryan language. It is 10th most spoken language in the world. Most of the speakers of Punjabi language lives in Punjab region of India. Punjabi has mainly four dialects viz Majhi, Doabi, Malwi and Pwadhi. Today, dialects represents the default variety of oral communication in the Punjab and Standard Punjabi is almost exclusively used for writing. The Malwai dialect is taken for the study because there are twenty two districts in Punjab state and the malwai dialect is spoken in more than half of the districts. Both Standard Punjabi and Malwai use the same script i.e. Gurukhi but the dialect is different. In this paper the difference between the Standard Punjabi and Malwai are explained.

Keywords: Machine Translation, Punjabi, Malwai Dialect

### Introduction

Punjabi is the Official language of Punjab. It is spoken by approximately 150 million people all around the world. It is 11th most spoken language in India. It is 4th most spoken language in England and 3rd most spoken language in Canada. The influence of Punjabi language as a cultural language is increasing day by day in India. Punjabi language is categorize as Modern Indo-Aryan Language. Punjabi language is differentiated from other languages due to its use of tones which differentiate the words. Punjabi language is written in Gurmukhi script.

### Gurmukhi Script

Gurmukhi is the most common script used for writing the Punjabi Language in India. Gurmukhi alphabets are derived from the Landa alphabets that has root in Brahmi alphabet. Gurmukhi script was standardized by the Second Sikh Guru, Sri Guru Angad Dev Ji in the 16th century. The name Gurmukhi is derived from the Old Punjabi term "gurumukhī", meaning "from the mouth of the Guru".

**Consonants** Modern Gurumukhi has forty-one consonants (vianjan).

Name	Pron.(IP A)	Name	Pron.(IP A)	Name	Pron.(IP A)	Name	Pron.(IP A)	Name	Pron.(IP A)					
ੳ	uɾa	–	ਅ	æɾa	ə by itself	ੲ	iɾi	–	ਸ	səs:a	sə	ੳ	ɦaɦa	ɦə
ਕ	kək: a	kə	ਖ	k <sup>h</sup> ək <sup>h</sup> : a	k <sup>h</sup> ə	ਗ	gəg:a	gə	ਘ	kəg: a	kə	ਙ	ŋaŋ: ã	ŋə
ਚ	tʃətʃ: a	tʃə	ਛ	tʃ <sup>h</sup> ətʃ <sup>h</sup> : a	tʃ <sup>h</sup> ə	ਜ	dʒədʒ: a	dʒə	ਝ	tʃədʒ: a	tʃə	ਞ	ɲəɲ: ã	ɲə
ਟ	tɛɲk a	tə	ਠ	t <sup>h</sup> ət <sup>h</sup> :a	t <sup>h</sup> ə	ਡ	dəd:d:a	də	ਢ	təd:d:a	tə	ਣ	ɳaɳa	ɳə
ਤ	t̪ət̪:a	t̪ə	ਥ	t̪ <sup>h</sup> ət̪ <sup>h</sup> :a	t̪ <sup>h</sup> ə	ਦ	d̪əd̪:a	d̪ə	ਧ	t̪əd̪:a	t̪ə	ਨ	nən: a	nə
ਪ	pəp: a	pə	ਫ	p <sup>h</sup> əp <sup>h</sup> : a	p <sup>h</sup> ə	ਬ	bəb:a	bə	ਭ	pəb: a	pə	ਮ	məm :a	mə
ਯ	jəj:a	jə	ਰ	rara	rə	ਲ	ləl:a	lə	ਵ	vav: a	və	ੜ	ɽaɽa	ɽə

There are six consonants created by placing a dot (bindi) at the foot (pair) of the consonant. These are used most often for loanwords, though not exclusively:

Name	Pron.(IPA)	
ਸ	səs:a pɛɾ bɪnd̪i	ʃə
ਖ	k <sup>h</sup> ək <sup>h</sup> :a pɛɾ bɪnd̪i	xə
ਗ	gəg:a pɛɾ bɪnd̪i	ɣə
ਜ	dʒədʒ:a pɛɾ bɪnd̪i	zə
ਘ	p <sup>h</sup> əp <sup>h</sup> :a pɛɾ bɪnd̪i	fə
ਲ	ləl:a pɛɾ bɪnd̪i	lə

## Vowels

Gurmukhi is similar to Brahmi scripts in that all consonants are followed by an inherent ‘a’ sound (unless at the end of a word when the ‘a’ is usually dropped). This inherent vowel sound can be changed by using dependent vowel signs which attach to a bearing consonant. In some cases, dependent vowel signs cannot be used – at the beginning of a word or syllable for instance – and so an independent vowel character is used instead. Independent vowels are constructed using three bearer characters: Ura (ੳ), Aira (ਅ) and Iri (ੲ). With the exception of Aira (which represents the vowel 'a') they are never used without additional vowel signs.

Vowel			Transcription		IPA	Closest English equivalent
Ind.	Dep.	with /k/	Name	Usage		
ਅ	(none)	ਕ	Muktā	a	[ə]	like <i>a</i> in <i>about</i>
ਆ	ਾ	ਕਾ	Kannā	ā	[ɑ] , [ã]	like <i>a</i> in <i>car</i>
ਇ	ਿ	ਕਿ	Sihārī	i	[ɪ]	like <i>i</i> in <i>it</i>
ਈ	ੀ	ਕੀ	Bihārī	ī	[i]	like <i>i</i> in <i>litre</i>
ਉ	ੁ	ਕੁ	Onkaṛ	u	[ʊ]	like <i>u</i> in <i>put</i>
ਊ	ੂ	ਕੂ	Dulankaṛ	ū	[u]	like <i>oo</i> as in <i>food</i>
ਏ	ੇ	ਕੇ	Lāvāñ	ē	[e]	like <i>e</i> in <i>Chile</i>
ਐ	ੈ	ਕੈ	Dulāvāñ	e	[ɛ]	like <i>e</i> in <i>sell</i>
ਓ	ੋ	ਕੋ	Hōṛā	ō	[o]	like <i>o</i> in Spanish <i>amor</i>
ਔ	ੌ	ਕੌ	Kanōṛā	o	[ɔ]	like <i>o</i> in <i>off</i>

### Vowel examples

Word	IPA	Meaning
ਆਲੂ	[ɑlu]	potato
ਦਿਲ	[d̪il]	heart
ਗਾਂ	[gã]	cow

### Nasalisation: tippi and bindi

Ṭippi ( ੱ ) and bindi ( ੰ ) are used for producing a nasal phoneme depending on the following obstruent or a nasal vowel at the end of a word. All short vowels utilise ṭippi and all long vowels are paired with bindi except for Dulankar ( ੁ ) which uses ṭippi instead . Older texts may not follow these conventions. The aforementioned bindi ( ੰ ) is also used for nasalisation.

### Gemination: addak

The use of addak ( ੱ ) indicates that the following consonant is geminate. This means that the subsequent consonant is doubled or reinforced.

### Halant

The halant ( ੜ ) character is not used when writing Punjabi in Gurmukhi. However, it may occasionally be used in Sanskritised text or in dictionaries for extra phonetic information. When it is used, it represents the suppression of the inherent vowel.

The effect of this is shown below:

ਕ – kə

ਕ̣ – k

## Visarg

The *visarg* symbol (◌ᳵ U+0A03) is used very occasionally in Gurmukhi. It can either represent an abbreviation (like period is used in English) or it can act like a Sanskrit Visarga where a voiceless ‘h’ sound is pronounced after the vowel.

## Udaat

The *udaat* symbol (◌ᳶ U+0A51) occurs in older texts and indicates a high tone.

## Numerals

Gurmukhi has its own set of numerals that behave exactly as Hindu-Arabic numerals do. These are used extensively in older texts. In modern contexts, they have been replaced by standard Latin numerals.

Numeral	Name	Number
੦	ਸਿਫਰ [sifər]	zero
੧	ਇੱਕ [ik:]	one
੨	ਦੋ [d̪o:]	two
੩	ਤਿੰਨ [t̪in:]	three
੪	ਚਾਰ [tʃa:r]	four
੫	ਪੰਜ [pəndʒ]	five
੬	ਛੇ [tʃʰe:]	six
੭	ਸੱਤ [sət:]	seven
੮	ਅੱਠ [ətʰ:]	eight
੯	ਨੌ [nɔ̃]	nine
੧੦	ਦਸ [d̪əs]	Ten

## Unicode

Gurmukhī script was added to the Unicode Standard in October, 1991 with the release of version 1.0. Many sites still use proprietary fonts that convert Latin ASCII codes to Gurmukhi glyphs.

The Unicode block for Gurmukhī is U+0A00–U+0A7F:

Gurmukhi Official Unicode Consortium code chart																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+0A0x		ੱ	ੰ	ੰ		ਅ	ਆ	ਇ	ਈ	ਉ	ਊ					ਏ
U+0A1x	ਐ			ੳ	ਐ	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	ਟ
U+0A2x	ਠ	ਡ	ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ		ਪ	ਫ	ਬ	ਭ	ਮ	ਯ
U+0A3x	ਰ		ਲ	ਲ਼		ਵ	ਸ਼		ਸ਼	ਹ			੍ਰ		ਾ	ਿ
U+0A4x	ੀ	ੂ	ੂ					ੇ	ੈ			ੋ	ੌ	੍ਰ		
U+0A5x		੍ਰ								ਖ਼	ਗ਼	ਜ਼	ੜ		ਫ਼	
U+0A6x							ੰ	ੴ	ੲ	ੳ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ
U+0A7x	ੰ	ੱ	ੲ	ੳ	ੴ	ੴ										

**Notes**

1. [▲](#) As of Unicode version 7.0
2. [▲](#) Grey areas indicate non-assigned code points

## Peculiarities of Malwai Dialect:

### Grammatical Peculiarities:

In Malwai dialect there are some difference at word level and at grammar level

a) Pronouns: Malwai Pronouns are different from Standard Punjabi as

ਪੰਜਾਬੀ	ਮਲਵਈ
ਕੂੰ	ਕੈਂ
ਤੁਹਾਡਾ	ਥੋਡਾ
ਤੁਹਾਨੂੰ	ਥੋਨੂੰ
ਆਪਣਾ	ਆਵਦਾ

b) Conjunctive words: There are two categories of conjunctive words in Malwai dialect : ਕਾ, ਕੇ, ਕੀ and ਦਾ, ਦੇ, ਦੀ. The conjunctive words ਕਾ, ਕੇ, ਕੀ are used to show the relationship while ਦਾ, ਦੇ, ਦੀ are non-relationship conjunctive words.

c) Verbs: In Malwai dialect Future Verbs are treated in different way. For Example

ਪੰਜਾਬੀ	ਮਲਵਈ
ਖਾਵਾਂਗਾ	ਖਾਊਂਗਾ
ਲਵਾਂਗਾ	ਲਊਂਗਾ

d) Helping Verbs: In Malwai dialect the helping verbs are used as below

ਪੰਜਾਬੀ	ਮਲਵਈ
ਹੈ	ਐ
ਹਨ	ਆ
ਸੀ	ਸਾ

### Phonetic Peculiarities:

a) There are 28 consonants and 10 vowels in Malwai dialect. ਵ and ਙ are not pronounced in the Malwai dialect.

b) In Malwai the words like ਟ, ਠ, ਡ, and ਣ have more dominance like

ਪੰਜਾਬੀ	ਮਲਵਈ
ਦੇਰ	ਡੇਰ
ਤੋਰ	ਟੋਰ

c) In Malwai the 'ਸ਼' and 'ਛ' are used interchangeably.

ਪੰਜਾਬੀ	ਮਲਵਈ
ਸ਼ੇਰ	ਛੇਰ
ਛੇਤੀ	ਸ਼ੇਤੀ

d) The 'ਵ' is replaced with 'ਬ' and 'ਮ' in Malwai dialect.

ਪੰਜਾਬੀ	ਮਲਵਈ
ਵੀਰ	ਬੀਰ
ਵੀਹ	ਬੀਹ

e) In Malwai the vowel 'ਅ' and 'ੲ' is used less frequently.

ਤੀਵੀਂ	ਤੀਮੀਂ
-------	-------

<b>ਪੰਜਾਬੀ</b>	<b>ਮਲਵਈ</b>
ਅਨਾਜ	ਨਾਜ
ਅਨੰਦ	ਨੰਦ
ਇਕੱਠਾ	ਕੱਠਾ
ਇਕੱਲਾ	ਕੱਲਾ

f) In the following words, the “ਦਿ” sound is missing from the conjunctive verb “ਦਿੱਤਾ”

<b>ਪੰਜਾਬੀ</b>	<b>ਮਲਵਈ</b>
ਦੇ ਦਿਤਾ	ਦੇ ਤਾ
ਦੇ ਦਿਤੀ	ਦੇ ਤੀ
ਦੇ ਦਿਤੀਆਂ	ਦੇ ਤੀਆਂ

g) When we use conjunctive verb “ਲੈਣਾ” then the difference in Standard Punjabi and Malwai dialect is as follows.

<b>ਪੰਜਾਬੀ</b>	<b>ਮਲਵਈ</b>
ਖਾ ਲਈ	ਖਾ ਲੀ
ਖਾ ਲਵਾਂ	ਖਾ ਲਾਂ
ਖਾ ਲਈਆਂ	ਖਾ ਲੀਆਂ
ਖਾ ਲਊਂ	ਖਾ ਲੂੰ

h) The usage of Conjunctive verb “ਦੇਣਾ” in Standard Punjabi and Malwai

<b>ਪੰਜਾਬੀ</b>	<b>ਮਲਵਈ</b>
ਦੇ ਦਈਂ	ਦੇ ਦੀਂ
ਦੇ ਦੇਵਾਂ	ਦੇ ਦਾਂ
ਦੇ ਦਊਂ	ਦੇ ਦੂੰ

i) The usage of Conjunctive verb “ਪੈਣਾ” in Standard Punjabi and Malwai

<b>ਪੰਜਾਬੀ</b>	<b>ਮਲਵਈ</b>
ਭੱਜ ਪਵਾਂ	ਭੱਜ ਪਮਾਂ
ਭੱਜ ਪਈਏ	ਭੱਜ ਪੀਏ
ਭੱਜ ਪਇਓ	ਭੱਜ ਪਿਓ
ਭੱਜ ਪਈ	ਭੱਜ ਪੀ

j) The usage of Conjunctive verb “ਜਾਣਾ” in Standard Punjabi and Malwai

<b>ਪੰਜਾਬੀ</b>	<b>ਮਲਵਈ</b>
ਆ ਜਾਵਾਂ	ਆ ਜਾਂ
ਆ ਜਾਈਏ	ਆ ਜੀਏ
ਆ ਜਾਵੀਂ	ਆ ਜੀਂ
ਆ ਜਾਓ	ਆ ਜੋ



## Conclusion

As malwai is one of the dialect of Standard Punjabi Language so Standard Punjabi and Malwai are closely related to each other. There are significant difference at the various parts that need to be taken care while developing Machine translation system from Standard Punjabi to Malwai dialect and vice-versa. The comparative study can be used for the development of machine translation system from standard Punjabi to Malwai dialect. Since Standard Punjabi and Malwai are closely related languages so Direct approach of Machine Translation shall be used. The Challenge for the development of Standard Punjabi-Malwai Machine Translation system is lack of language resources. The Malwai dialect is either used in the spoken communication or used by some of the Punjabi writers in their novels, Natak's. So it's a challenging task to collect the data and prepare a machine readable bilingual dictionary.

## References:

1. [http://simple.wikipedia.org/wiki/Punjabi\\_language](http://simple.wikipedia.org/wiki/Punjabi_language) extracted on 16-12-2014.
2. [http://en.wikipedia.org/wiki/Gurmukh%C4%AB\\_alphabet](http://en.wikipedia.org/wiki/Gurmukh%C4%AB_alphabet) extracted on 12-01-2015
3. Tarlochan Singh Bedi, Dr. *Gurmukhi Lipi da Janam te Vikas*. Patiala: Punjabi University, 1999.
4. Kala Singh Bedi, Dr. *Lipi da Vikas*. Patiala: Punjabi University, 1995.
5. Manminder Singh, *Malwai Shabad Kosh*, Publication Bureau, Punjabi University, Patiala
6. Vishal Goyal, Gurpreet Singh Lehal, “*Comparative Study of Hindi and Punjabi Language Scripts*”, p.p 67-82, Napalese Linguistics, Journal of the Linguistics Society of Nepal, Volume 23, November, 2008
7. <http://patch.com/connecticut/trumbull/history-of-punjabi-language--gurmukhi-alphabet> extracted on 12-01-2015.
8. Harjeet Singh Gill and Henry A. Gleason, Jr. *A Reference Grammar of Punjabi*. Publication Bureau, Punjabi University, Patiala, India. 1986.
9. Joginder Singh Puar. *The Punjabi verb form and function*. Publication Bureau, Punjabi University, Patiala, India. 1990
10. Duni Chandar. *Punjabi Bhasha Da Vyakaran (Punjabi)*. Punjab University Publication Bureau, Chandigarh, India. 2002

11. Shiv Sharma Joshi (ed.). *Punjabi Bhasha: Viyakaran Ate Bantar (Punjabi)*. Publication Bureau, Punjabi University, Patiala, India. 2000.