Comparative Study of Hindi and Punjabi Language Scripts

[1]Vishal Goyal,   [2]Gurpreet Singh Lehal
[1]Lecturer, [2]Professor and Head
Department of Computer Science, Punjabi University, Patiala
[1]vishal.pup@gmail.com, [2]gslehal@gmail.com

Abstract:

Hindi and Punjabi languages belong to the same subgroup of the Indo-European family i.e. Indo-Aryan family of the languages. The script of Hindi language is Devanagari and the script of Punjabi Language is Gurmukhi. Not only the languages are descendant of same stock, but also their scripts are. Hindi is one of the most widely spoken languages of the world, possessing speakers of the same order of magnitude as those of English and Russian. In India it has been accorded the status of 'official language' by the central govt. for use for most administrative purposes, and Punjabi being the official language of state Punjab and has been accorded the status of 'official language' by the Punjab government for use for most administrative purposes. During the development of Machine Translation system from Hindi to Punjabi Language, there is a need to find the closeness between both the languages. It has been found that both the languages are closely related to each other. In this paper, we will explain the closeness between the Hindi and Punjabi Languages scripts.

1 Introduction:

Hindi and Punjabi are closely related languages. Both the languages are originated from Sanskrit (Masica 1991). Punjabi language is mostly used in the region of Punjab, Haryana, Delhi, Himachal Pardesh, Jammu & Kashmir and in

some areas of Pakistan namely Punjab, Sindh and Blochistan. On the other hand, Hindi is a national language of India and is spoken and used by the people all over the country. But the main regions are Haryana, Uttar Pardesh, Rajasthan, Bihar and Chattisgarh. The script of Hindi language is Devanagari and the script of Punjabi language is Gurmukhi. In this paper, closeness between Hindi and Punjabi languages is explained which includes script, consonants, vowels, conjuct consonants, numerals, punctuation, abbreviations, alphabetic orders in both the languages.

2 Writing System:

Writing System is a form of human communication by means of a set of visible marks that are related, by convention, to some particular structural level of language. The writing is in principle the representation of language rather than a direct representation of thought and the fact that spoken language has a number of levels of structure, including sentences, words, syllables, and phonemes (the smallest units of speech used to distinguish one word or morpheme from another), any one of which a writing system can "map onto" or represent.
The invention of the first writing systems is roughly contemporary with the beginning of the Bronze Age in the late Neolithic of the late 4th millennium BC. The oldest-known forms of writing were primarily logographic in nature, based on pictographic and ideographic elements. Most writing systems can be broadly divided into three categories: logographic, syllabic, and alphabetic (or segmental). A logogram is a single written character which represents a complete grammatical word. Most Chinese characters are classified as logograms. In Syllabic writing system, a syllabary is a set of written symbols that represent (or approximate) syllables, which make up words. Devanagari and Gurmukhi comes under the category of Syllabic writing

system. In alphabetic writing system, an alphabet is a small set of *letters* — basic written symbols — each of which roughly represents or represented historically a phoneme of a spoken language. The

The true origin of writing in India has not been ascertained with any certainty though scholars believe that the starting point was the Brahmi script used in the inscriptions of emperor Asoka (300 BCE). It is however known that reference to writing is seen in the ancient scriptures of India, which have also defied proper dating.

Devanagari and Gurumukhi are descendants of Brahmi. Devanagari writing system is ultimately an off-shoot of north indian variety of Brahmi Script of undertermined origin.

3 Script and Language:

3.1 Gurmukhi Script:

Gurmukhi Script derived from the Sharada script and standarized by Guru Angad Dev in the 16th century, was designed to write the Punjabi Language(Gill,Gleason,1963). The word Gurmukhi is commonly translated as "from the mouth of Guru". However, the term used for the Punjabi script has somewhat different connotations. The opinion given by traditional scholars is that as the Sikh holy writings, before they were scribed, were uttered by the Gurus, they came to be known as Gurmukhi or the "Utterence of the Guru". And consequently, the script that was used for scribing the utterence was also given the same name. However, the prevalent view among Punjabi linguists is that asin the early stages the Gurmukhi letters were primarily used by Gurmukhs, or the Sikhs devoted to the Guruy, the script came to be associated with them. Another view is that as the Gurmukhs, in accordance with the Sikh belief, used to mediate on the letter ਵ , ਹ , ਗ , ਰ which jointly forms

ਵਾਹਿਗੁਰੂ or God in Sikhism, these letters were called Gurmukhi or the "Speech of the Gurmukhs". Subsequently, the whole script came to be known as Gurmukhi.

Like most of the north indian writing systems, the Gurmukhi script is a descendent of the Brahmi script. It is believed that Gurmukhi script was invented by the second Sikh Guru, Guru Angad Dev, However, it would be correct to say that script was standarized rather than invented, by the Sikh Gurus. E.P. Newton (Panjabi Grammar, 1898) writes that at least 21 Gurmukhi characters are found in ancient manuscripts : 6 from 10th century, 12 from 3rd century BC and 3 from 5th century BC. Apparently, the first Sikh Guru, Guru Nanak Dev also used the Gurmukhi script for his writings. The usage of Gurmukhi letters in Guru Granth Sahib meant that the script developed its own orthographical rules. In the following epochs, Gurmukhi became the prime script applied for literary writings of the Sikhs. Later in the 20th century, the script was given the authority as the official script of the Eastern Punjabi Lanuage. Meanwhile, in western Punjab a form of the Urdu script, known as Shahmukhi is still in use.

Gurmukhi is a form writing system called an abugida, as each consonant has an inherent vowel (a) that can be changed using vowel signs. Modern Gurmukhi has fortyone consonants (Vianjans), nine vowel symbols (Laga Matra), two symbols for nasal sounds (Bindi and Tippi) and one symbol which duplicates the sound of any consonant (Adddak). In addition, four conjuncts are used : three subjoined forms of the consonants Rara, Haha and Vava and one half-form of Yayya. Use of conjuct forms of Vava and Yayya is increasingly scarce in modern contexts. Gurmukhi has been adapted to write languages, such as Sanskrit, Hindi and Braj Bhasha.

3.2 Devanagari script:

The Devanagari script, used for writing Sanskrit and other Indian languages had evolved over a period of more than two thousand years. Devanagari emerged around 1200 AD out of the Siddham script, gradually replacing the earlier, closely related Sharada script (which remained in parallel use in Kashmir). Both are immediate descendants of the Gupta script, ultimately deriving from the Brāhmī script attested from the 3rd century BC; Nagari appeared in approx. the 8th century as an eastern variant of the Gupta script, contemporary to Sharada, its western variant. The descendants of Brahmi form the Brahmic family, including the alphabets employed for many other South and South-East Asian languages.

Nāgarī is in Sanskrit the feminine of *nāgara*. The feminine form is used because of its original application to qualify the feminine noun *lipi* "script". There were several varieties in use, one of which was distinguished by affixing *deva* "divine, deity" to form a tatpurusha compound meaning the "divine urban(e) [script]". However, the widespread use of "Devanagari" is a relatively recent phenomenon; well into the twentieth century, and even today, simply "Nagari" was (and is) also in use for this same script. The rapid spread of the usage of "Devanagari" seems also to be connected with the almost exclusive use of this script in colonial times (particularly by European scholars) to publish works in Sanskrit (held by many to be the language of the gods), even though traditionally nearly all indigenous scripts have actually been employed for this language. This has led to the establishment of such a close connection between the script and Sanskrit that it is erroneously widely regarded as "the Sanskrit script" today.

4 Consonants:

There are thirty three basic consonantal signs in Devanagari script and thirty five in Gurmukhi script which are as follows.

In the following table, the correspondence between basic consonants in both the scripts are presented in Table 1 below:

Table 1 : Basic Consonants

Devanagari	क	ख	ग	घ	ङ
Gurmukhi	ਕ	ਖ	ਗ	ਘ	ਙ
Devanagari	च	छ	ज	झ	ञ
Gurmukhi	ਚ	ਛ	ਜ	ਝ	ਞ
Devanagari	त	थ	द	ध	न
Gurmukhi	ਤ	ਥ	ਦ	ਧ	ਨ
Devanagari	ट	ठ	ड	ढ	ण
Gurmukhi	ਟ	ਠ	ਡ	ਢ	ਣ
Devanagari	प	फ	ब	भ	म
Gurmukhi	ਪ	ਫ	ਬ	ਭ	ਮ
Devanagari	य	र	ल	व	
Gurmukhi	ਯ	ਰ	ਲ	ਵ	
Devanagari	श	ष	स	ह	
Gurmukhi	*ਸ਼		ਸ	ਹ	
Devanagari	**ੜ	**ਠ	**अ	**इ – ए	
Gurmukhi	ੜ	ੳ	ਅ	ੲ	

The characters marked with * symbol in the above Table 1 are not a basic consonant. Basic alphabet ਸ is supplemented with a dot diacritic. In this chart, to show the corresponding

transliteration for Devanagari basic consonant श and प, has been shown.
Similarly the characters marked with **symbol in the Table 1 above are not basic consonant. Basic alphabet ड is supplemented with a dot diacritic.

4.1 Dead and Live Consonants:

Devanagari employs a sign known in Sanskrit as the *virama* or vowel omission sign. In Hindi it is called *hal* or *halant*, and that term is used in referring to the virama or to a consonant with its vowel suppressed by the virama. The virama sign (◌्) nominally serves to cancel (or kill) the inherent vowel of the consonant to which it is applied. When a consonant has lost its inherent vowel by the application of virama, it is known as a *dead consonant*; in contrast, a *live consonant* is one that retains its inherent vowel or is written with an explicit dependent vowel sign. It is to noted that there are no dead consonants in Gurmukhi Script.

4.2 Consonant Conjuncts:

The Indic scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms (Michael, 1986). This abbreviation takes place only in the context of a *consonant cluster*. An orthographic consonant cluster is defined as a sequence of characters that represents one or more dead consonants followed by a normal, *live* consonant letter.

Table 2: Conjuct Consonants

Devanagari	ज्ञ	क्ष	श्र	त्र
	(ज् + ञ)	(क् + श)	(श् + र)	(त् + र)
***Gurmukhi	ਗਿਅ	ਕਸ਼	ਸ੍ਰ	ਤਰ

Corresponding to the consonants mentioned in devnagri mentioned in Table 2, no corresponding consonant conjuncts in Gurmukhi script are present. But the corresponding transliterations have been shown in the Table 2.

In Table 1, to show the corresponding transliteration for Gurmukhi basic consonant ੲ, has been shown. Similarly, ਅ , ੲ , ਏ, ੳ are not considered consonants and just to show Devanagari equivalents of Gurmukhi ੳ, ਅ, ੲ has been mentioned in this chart.

Some of the basic alphabet is supplemented by the following letters with a dot diacritic:

Table 3

Devanagari	क़	ख़	ग़	ज़	फ़	ढ	श/ष	———
Gurmukhi	———	ਖ	ਗ਼	ਜ਼	ਫ਼	ੜੁ (ੜ+ ਹ)	ਸ਼	ਲ਼

In Gurmukhi, only three types of conjunct consonants are used. In all bases, a modified form of the second consonant is subjoined to the unaltered form of the first.

In the first type, a form of ਹ is subjoined

Table 4

Base	Form	Devanagari Equivalent	Example
ੜ	ੜੑ	ਢ	ਪੜੑ
ਨ	ਨੑ	ਨੑਂ	ਨੑੇਰ
ਲ	ਲੑ	ਲੑ	ਲੑੁ
ਮ	ਮੑ	ਮੑ	ਮੑੇਸ

In second type of conjunct, a form of ਵ is subjoined to certain consonants, most commonly stops. These occur only in tatsamas (Those words that are directly borrowed from Sanskrit with little or no phonetic alteration) like ਪ੍ਰ, ਵ੍ਰ, ਸ੍ਰ etc.

In Devanagari, when र is served as the second member of a cluster, it is indicated by a small diagonal slash (going in the opposite direction from that of the virama) written under the sign for the first member of a conjunct : क्र प्र द्र ग्र

When र is served as the first member of a conjunct, the sound is indicated by a small
hook placed on the top of the rekha for the second consonant: र्क र्हे र्श र्म . This hook is deferred until after any matra written to the right side of the conjunct like र्थी र्मो
In third type of conjunct, a form of ੍ਰ is subjoined. For

example: ਸ੍ਰ in Gurmukhi is written as स्व in Devanagari ,

Similarly ਸ੍ਰਰ in Gurmukhi is written as स्वर in Devanagari.

Several Devanagari conjunct are so irregular as to prelude the immediate recognition of their components. The most important of these are क्क क्ष ज्ञ क्त द्द द्ध द्म . The consonant श has

a special combining form श्र that is often used in place of श्

in some clusters.(e.g. श्र, श्य) . Slightly irregular conjuncts

exists in which ह stand as the first element (e.g. ह्र ह्न ह्य ह्ल ह्व).

4.3 Geminate (Doubled) consonants :

In Gurmukhi, gemination is written by the sign ੱ (addak)

above and before the consonant to be doubled. In Devanagri, doubled consonant cluster, gemination is written by writing the first component of the consonant cluster as the truncate form of the consonant (which is frequently built from the independent version of the latter consonant by the deletion of the vertical bar that appears on the right side of many Devanagri character and the second component of the consonant cluster is, the unaltered full symbol for the second

consonant. E.g. पक्की(ਪੱਕੀ in Gurmukhi), कच्चा (ਕੱਚਾ in

Gurmukhi).   Similarly, in Gurmukhi, clusters of unaspirated stop plus homorganic aspirate stop are written by use of ੱ (addak) before the letter for the aspirate. In Devanagari, this

cluster is written with the short form of unaspirated stop plus

full form of homorganic aspirate stop. For example : अच्छा

(ਅੱਛਾ in Gurmukhi), पक्खी (ਪੱਖੀ in Gurmukhi).

In a small number of cases, the components of a consonants are sting out in a horizontal line (e.g. न्न) , arranged vertically

or juxtaposed in some less regular manner (द्र, द्द) . Similarly in the Gurmukhi two geminates /nn/ and /mm/ are written with /tippi/ (ੰ) . e.g. ਪੰਨਾ (पन्ना in Devanagari) , ਪੰਮਾ (पम्मा in Devanagari)

It must be noted that there are no short forms like in Devanagri for consonant. So, while transliterating the short from of Hindi consonant , it is transliterated into full form of that consonant in Gurmukhi like मग्न in Devanagari will be transliterated into ਮਗਨ.

5 Vowels:

Both the Scripts possess two different forms for each of the vowels- Full form and short form.

5.1 Full form :

In Devanagari, a full form is employed for a vowel that does not immediately follow a consonant or consonant cluster, i.e. in word-initial position or when the second of a sequence of vowels. Whereas in Gurmukhi, when a vowel is not proceeded by a consonant, it is written with one of the three vowel bearers - consonant like sign – ੳ , ਅ , ੲ indicating the absence of consonant.

5.2 Short form (or *matra*):

In Devanagari, short form is used when the vowel immediately follows a consonant or consonant cluster. These short forms consist of lines, hooks or combination of both above, below or to the side of the consonantal characters.

These vowels are written around (that is, below, above, to the right, and to the left) the consonant signs.
Following table shows the both the above form of vowels for both the scripts and correspondence of the vowels between both the scripts are also shown.

Table 5

Devanagari		Gurmukhi	
Short Form	Full Form	Short Form	Full Form
No Sign	अ	No Sign	ਅ
◌ा	आ	◌ਾ	ਆ
ि◌	इ	ि◌	ਇ
◌ी	ई	◌ੀ	ਈ
◌ु	उ	◌ੁ	ੳ
◌ू	ऊ	◌ੂ	ੳੁ
◌े	ए	◌ੇ	ਏ
◌ै	ऐ	◌ੈ	ਐ
◌ो	ओ	◌ੋ	ੳ
◌ौ	औ	◌ੌ	ਔ
◌ं	---	◌ं / ◌ਂ	---
◌ँ	---	◌ं / ◌ਂ	---
Conjunct	---	◌ੱ	---
◌ृ	---	ऋ	----

5.3 Inherent a :

One vowel, 'a' has no special short form. The absence of a *matra* adjacent to a consonant suffices to indicate the presence of this vowel. At the end of a word, the inherent 'a' is not normally vocalized.

5.4 Nasalized vowels:

The two signs are used for nasalization. In Devanagari, *anusvara* (ॊ) and *anunasika* (ँ) also called *candrabindu*. Indian grammarians have formulated rules for their usage. The first of these, *anusvara* is always used when the vowel marking protrudes above the rekha (e.g. ई, ऐ, ओ, आ). With other vowel signs, both *anusvara* and *anunasika* can be used. Whereas In Gurmukhi, tippi (ੰ) is used with (ਅ, ੁ , ਿ) and with (ੁ) when final.

x
6 Punctuation:

Only viraama (।) or a double vertical line (॥) was used in traditional writing for marking end of sentence and the end of a verse respectively. In modern writings, period, comma, hyphen, semicolon, exclamation sign, question mark and dash has also been used. Similarly, in Gurmukhi Script (।), viram is used for end of the sentence.

7 Abbreviation :

Abbreviationa are formed in Hindi by the use of either a small circle (०)or a dot after the first syllable of the word to be abbreviation: प्रो॰, डा॰, ई॰, पू॰ whereas in Gurmukhi Script, sign (:) is used to mark abbreviation like ਪ੍ਰੋ: , ਡਾ:

8 Numerals:

Following chart shows the correspondence between the numerals of both the scripts:

Table 6

Devanagari	Gurmukhi
०	੦
१	੧
२	੨
३	੩
४	੪
५	੫
६	੬
७	੭
८	੮
९	੯

9 Alphabetic Order:

The alphabetic order of Devanagari is a model of logic and rational design, reflecting a keen understanding of the phonetic properties of the sounds designated by the various characters in the system. In Devanagari, Vowels precede consonants with the latter divided up into groups containing stops and nasals, semi vowels, sibilants, and h respectively.

The full alphabetic order of Devanagari as used for Hindi is as follows:

अ आ इ ई उ ऊ ऋ ए ऐ ओ औ क (क़) ख (ख़) ग (ग़) घ ङ च छ ज ञ झ ञ ट ठ ड ढ ण त त्र थ द ध न प फ ब भ म य र ल व श ष स ह

The full alphabetic order of Gurmukhi as used for Punjabi is as follows:

ੳ ਅ ੲ ਸ ਹ ਕ ਖ ਗ ਘ ਙ ਚ ਛ ਜ ਝ ਞ ਟ ਠ ਡ ਢ ਣ ਤ ਥ ਦ ਧ ਨ ਪ ਫ ਬ ਭ ਮ ਯ ਰ ਲ ਵ ੜ

In Hindi, sequence under each consonants is the letter without any symbol, then followed by vowel symbols ा , ि ,ी, ु , ू ,े, ै, ो, ौ

In Punjabi, Sequence under each consonants is the letter without any symbol, then followed by vowel symbols ਾ , ਿ , ੀ , ੁ , ੂ , ੇ , ੈ , ੋ , ੌ

10 Conclusion:

Hindi and Punjabi belongs to same sub group of the Indo European family and hence are closely related languages. Not only the languages are descendant of same stock, but also their scripts are. The writing system of both the scripts is similar but there are some interesting and significant differences. These are not simply in the form of the letters, but in the structure of the writing system. Most interesting differences are those in the writing of initial vowels, geminate clusters, other clusters, and of course writing of tone. This comparative study for writing system is very beneficial and important during the development of Machine translation system from Hindi to Punjabi and vice versa.

11 Future Works

Researchers working in this field can use this comparison in machine translation system development. Moreover this work can be extended to comparison of Hindi and Punjabi word inflections. Even this work can be taken as basis for comparing the speech system of Hindi and Punjabi and other similar languages.

References:

Bharati A, Moona R., Singh S., Sangal R., Sharma D.S., "MTeval: An evaluation methodology for Machine Translation Systems", Proceedings of the SIMPLE-Symposium on Indian Morphology, Phonology and Language Engineering 2004, IIT Kharagpur, India

Bhatia, Tej K. 1996. *Colloquial Hindi*. New York : Routledge.

Günther Hartmut, Ludwig Otto. 1996. *Writing and its use, an Interdisciplinary Handbook of international Research*. Walter de Gruyter.

Gill Harjeet , Gleason Henry A. 1963. *A Reference Grammar of Punjabi*. Patiala: Punjabi University Publication Bureau.

Hajic J, Hric J, Kubon V., "Machine Translation of Very Close Languages", proceedings of the 6th Applied Natural Language Processing Conference, April 29th - May 4, 2000, Seattle, Washington, USA. pp 7-12

Joshi S.S. 1978. *Punjabi-English Dictionary*, Patiala: Punjabi University Publication Bureau.

Marrafa, Palmira and Ribeiro A., "Quantitative Evaluation of Machine Translation Systems: Sentence level", Proceedings of the MT Summit VIII Fourth ISLE workshop 2001, spain, pp. 39-43.

Masica, Coline P. 1991. *Indo-Aryan languages*. Cambridge: Cambridge University Press.

Newton, E.P., 1898. *Panjabi Grammar*. Ludhiana mission press

Shapiro Michael C. 1986. *A primer of modern standard hindi*. New Delhi: Moti Lal Banarsi Dass Publishers.

Singh Jodh, Badan Baldev Singh, Singh Maninder, Joshi Ramsharan, Singh Rajinder. 1990. *Hindi to Punjabi Dictionary*, New Delhi: National Book Shop.

Singh H. 1991. *Saadian Bhashawan,* Ed. 1, New Delhi: Punjabi Academy.

Singh H. Singh L. 1986. *College Punjabi Viakaran* ,Chandigarh: Punjab State University text Book Board.

Slype V. 1979. "Critical Methods for Evaluating the Quality of Machine Translation," Prepared for the European Commission Directorate General Scientific and Technical Information and Information Management. Report BR-19142. Bureau Marcel van Dijk

Tomas J, Mas J. A., Casacuberta F., "A Quantitative Method for Machine Translation Evaluation", presented in workshop of 11th Conference of the European Chapter of the Association for Computational Linguistics April 12-17, 2003 Agro Hotel, Budapest, Hungary.

Web References:

http://www.hindisociety.com/ArticleHindiHistory.htm
http://www.omniglot.com/writing/devanagari.htm
http://www.knowledgerush.com/kr/encyclopedia/Gurmukhi_script/
http://www.punjabilok.com/heritage/punjabi_language.htm
http://www.omniglot.com/writing/definition.htm
http://en.wikipedia.org/wiki/Devanagari
http://www.experiencefestival.com/a/Devangar_-_Origins/id/1288685
http://sikhiwiki.com/index.php/Punjabi_Lessons