Phonology–orthography interface in Devanāgarī for Hindi*

Pramod Pandey
Jawaharlal Nehru University

1. Introduction

As is well-known, Devanāgarī is a derivative of Brāhmī. The parent script underwent continuous evolution and diversification in the orthographic symbols (see, e.g. Singh, 2007), but without affecting its akshara- based character (see e.g. Patel, 1996, 2005).[1] The akshara[2] is a grapheme consisting of an optional Onset (that may be simple or complex) and an obligatory Nucleus. A consonant after the Nucleus goes with the following akshara, except for a nasal.[3] An akshara may be roughly assumed to be an open syllable. As a graphemic unit, the akshara has proved productive and suitable for most languages for which it has been used. The present paper takes up the following two questions for investigation:

i   What precisely are the phonetic and phonological traits of Devanāgarī, a prominent instance of an akshara-based script?
ii. What is the relationship between the phonology of Hindi and its orthography in Devnāgarī script, and how does the relationship lead to suggestions for reforms in the orthography?

I begin, in § 2, with a discussion of the relation between speech and writing, as a basis for an account of the phonetic and phonological underpinning in Hindi orthography in § 4. The main features of the Devanāgarī script of Hindi are discussed in § 3. In § 5, I examine the nature of phonology-orthography interface in Hindi, and attempt to show that it is inconsistent with regard to the relevant phonological level (Sproat, 2000). In the light of the findings of the paper, the proposals for reform in Hindi orthography are presented in § 6. The main contributions of the paper are summarized in § 7.

2.   Speech and writing

In order to critically address the issue of reform in a script, it is useful to consider the nature of relation between speech and writing, in general, before examining the nature of the phonology-orthography interface in it. I turn to the general concern first.

Speech and writing are two different output mechanisms of language. Whereas speech is intrinsic to language learning, with a locus in the phonetic module, writing is not (Coulmas, 1997). The learning of writing is found (Sproat, 2000) to depend on various aspects of linguistic and non-linguistic awareness. Among the components of linguistic awareness, phonetics, phonology, lexis and morphology are the most crucial. Writing systems may differ on the basis of the relative order of their dependence on these components. Arabic, for example, depends more on lexical awareness than, say English and Hindi. English depends more on morphological awareness than Hindi, and Hindi depends more on phonological awareness than Arabic and English (Pandey, 2007b).

Let us now turn to the properties of speech sounds as understood in modern linguistic studies in phonetics and phonology (e.g Kenstowicz, 1994) in order to consider the phonetic and phonological bases of Devanāgarī script of Hindi.

2.1 Properties of speech sounds

Sound segments are not perceived as discrete, but rather as mingled with the context, somewhat like the 'diphones' (i.e. a sequence of two sounds stored as such) used in speech synthesis, as Ladefoged (2001:75) has suggested. They are affected by the sounds before and after them, in innumerable ways, that cannot be incorporated in a writing system, which must be conventional for communicative purposes in a society. Writing systems differ with regard to the level of phonological awareness they represent- broadly, phonetic or phonological or a combination of these. In the case of character- type writing systems, such as Chinese, phonetic information may be very limited.

Speech segments are assumed to emerge from the organization of gestures, which are phased differently for different sound sequences (see e.g. Browman and Goldstein, 1989, 1991, Clements, 1999, Sproat and Fujimura, 1993, Studdert-Kennedy, 1987, 1998). The gestures are most commonly organized within a frame called the syllable (see e.g. Blevins, 1995, Clements and Keyser, 1983, Selkirk 1982), a hierarchically structured unit, consisting of the Onset, the Nucleus and the Coda. Of these, the Nucleus is an obligatory constituent; the Onset and the Coda are optional constituents. The latter constituents are also found to be asymmetrical (see Bybee, 2001 for a recent discussion). Whereas the presence of the

Onset in world languages is seen as unmarked, the presence of a Coda is seen as marked. Syllables without the Coda are Open Syllables and syllables with the Coda are Closed syllables. There is no language without an open syllable, while there are many languages without a closed syllable.

Vowels and consonants behave differently in phonological systems, especially noticeable in Semitic languages (see McCarthy, 1981). Thus, two or three consonants give the root form of the word, while vowels give the different grammatical forms of words, as can be seen in the following output forms: /kataba/ 'wrote 3 P SG PAST', /kita:b/ 'book SG', /kutub/ 'book PL', etc.. In all these cases, the consonants **k-t-b** constitute the consonantal root, and the vowel sequences **a-a-a, i-a:** and **u-u** give the grammatical morphemes.

Speech sounds are conceived of as 'basic' and 'derived' in certain theories of segmental composition, such as Dependency Phonology (cf. Anderson and Ewen, 1987). The theory claims that human speech has three basic vowels — [a], [i], and [u], and the rest are derived from their combination, much along the lines of the organization propounded for Sanskrit in Panini's *Ashtadhyayi*.

Most phonological systems make use of the economic device of underspecification of segments. The theory of Radical Underspecification (see Steriade, 1995) assumes that one of the vowels in a language is fully underspecified and surfaces as ə (**i/u** in different cases). Government Phonology (Kaye et al., 1985) also assumes that a consonant has an inherent neutral vowel, which has to be suppressed by individual languages in which it doesn't surface. In default cases, it surfaces as ə.

Phonological segments in utterances are organized into prosodic units such as the metrical Foot (a unit of stress), the Phonological Word, the Phonological Phrase (a unit with a potential pause but without a pitch change), the Intonational Phrase (a unit with a nuclear tone and pause) and the Utterance, as shown for the sentence, *The 'man in a 'red 'hat was the 'refe'ree of the' team*, as follows:

$$_U[\;_{IP}[\;_{PP}[\;_W[\text{The } _F[\text{man}]_F \text{ in a}]_{W \; W}[_F \text{ [red]}_F]_W \; [_W \; [_F[\text{hat}]_F]_W \;]_{PP \; PP}[_W[\text{was the}[$$
$$_F[\text{refe}]_F \; _F[\text{ree}]_F]_{W \; W}[\text{of the } _F[\text{team}]_F]_W]_{PP} \;]_{IP} \;]_U.$$

In the above representation, F= Foot, PP= Phonological Phrase, W= Phonological Word, IP= Intonational Phrase, U= Utterance (see e.g. Spencer, 1996, for detailed illustrations of the mapping of these units on utterances). The sentence is assigned the structure of an Utterance containing one IP, two PPs, five phonological Words and six Feet. The sentence can be produced (and thus represented) with alternative prosodic structures, such as two intonational phrases, *The man in a red hat* and *was the referee of the team*.

There is hardly any orthographic system that represents information about the prosodic structure of these units. (One of the reasons why their representation is neither necessary nor possible is that they are emergent systems, consequent on

the neuro-motor activities in which utterances are routinely produced and perceived, rather than being basic to representation.)

2.2   Properties of writing systems

Some of the main properties of writing systems that claim our attention are discussed below.

As a representational system, orthography needs to adhere to some general usability requirements that are not always compatible. Three of these are especially important: *Distinctness, Learnability,* and *Creativity. Distinctness* requires that each unit of the alphabet be distinct from the others. *Learnability*, on the other hand, requires that an alphabetic unit be similar to a phonetically 'similar' existing alphabetic unit. These requirements together lead to the *Alphabetic Paradox* (Pandey, 2003): the units of the alphabet should be both similar to and distinct from one another. Every script has to resolve the paradox in some way.

The third property, namely, *Creativity,* requires that a small set of elementary units combine and permute to yield all the existing graphemes of the orthography, and more if needed. Creativity in natural systems, it has recently been proposed, is owing to a general Particulate Principle leading to self-diversifying systems. "In such systems, discrete units from a finite set of meaningless elements (e.g. atoms, chemical bases, phonetic segments) are repeatedly sampled, permuted and combined to yield larger units (e.g. molecules, genes, words) that are higher in a hierarchy and both different and more diverse in structure and function than their constituents." (Studdert-Kennedy 1998: 161). Studdert-Kennedy has suggested that the principle can be used to investigate creativity in other natural systems, including writing systems. Pandey (2007b) has argued how a limited set of units of writing, namely, a line, a dot, and a cursive, is used to produce an open number of symbols in Brāhmi and its derivatives.

The notion of creativity being closely linked with generativity in current linguistics, it is relevant to note that the nature of generativity in orthographies is of the *representational* kind, rather than of the *derivational* kind, as noted in Pandey (2007b:245). This characterization of orthographic generativity is based on the distinction in generative systems made in Chomsky (1995:223). While a derivational generative system involves a set of successive steps before arriving at the output form, a representational generative system employs some other means to do so, such as taking one unit as the initial state and arriving at another as the final state.

Representational orthographic systems are assumed to be 'planar regular languages' recognizable using *finite state automata* (see Sproat, 2000:30–41 for a discussion of these concepts).

Two main properties of a writing system in relation to a language, as has recently been proposed in a constrained theory of writing systems (Sproat, 2000) are *regularity* and *consistency*. The claim for *regularity* assumes that the mapping between the orthography (i.e. the spelling) and the linguistic representation (called the *Orthographically Relevant Level* (*ORL*)) 'can be handled by context-sensitive rewrite rules' or 'spelling rules' and that the relation between the two is 'invertible'. The property of *consistency* lies in the ORL being the same 'across the entire vocabulary of the language'. The ORL can be 'deep' if it applies at the lexical level of phonology or some higher level, or it can be shallow if it applies at the surface phoneic or output level of phonology (see e.g. Venezky, 1970, Klima, 1972) and others cited in Sproat (2000) on an early discussion of the relation between the orthographic and the linguistic levels).

3.   Phonological aspects of the Devnāgarī script

3.1   Main features

Some of the main features of the Devanāgarī script as discussed in the literature (see e.g. Agrawala 1966, Bright, 1996, McGregor, 1977, Sproat, 2000) are as follows:

i.   The basic unit of the script is the akshara, a grapheme consisting of CV sequences, that is, Onset and Nucleus sequences, as illustrated in *Table 1.*

Table 1.   Devanāgarī CV aksharas

क	का	कि	की	कु	कू	के	कै	को	कौ
kə	ka	kɪ	ki:	kʊ	ku:	ke	kæ	ko	kəʊ

The consonant phonologically classified as the Coda of the syllable is represented in a following grapheme. The only exception is the 'anuswāra', the underlying nasal consonant surfacing as homorganic with the following stop, which is treated as a part of the grapheme. The orthographic and phonetic transcriptions of forms with the 'anusvāra' are given below.

Note that the anusvāras are alternatively represented in the general way, as part of the following grapheme, as in (1) below:

(1)   बन्दी चम्पा ठण्डा मञ्जरी सङ्घ

ii.   When not preceded by an onset C, the vowels are represented fully as independent graphemes, otherwise as diacritics, around the onset consonant(s)-

Table 2. Anusvāra forms in Hindi

बंदी	[bəndi:]	'prisoner'
चंपा	[cəmpa:]	(name of a flower)
ठंडा	[ʈʰəɳɖa:]	'cold (masc.)'
मंजरी	[məndʒəri:]	(a name)
संघ	[səngʰ]	'a group'

Table 3. Devanāgarī vowels in their full and diacritic forms, adapted from Sproat (2000: 46)

Expression	Full form		Diacritic form	
Null	अ	<ə>	क	<kə>
After	आ	<a>	का	<ka>
	ओ	<o>	को	<ko>
	औ	<əo>	कौ	<kəo>
	ई	<i:>	की	<ki:>
Above	ए	<e>	के	<ke>
	ऐ	<ɛ>	कै	<kɛ>
Below	उ	<ʊ>	कु	<kʊ>
	ऊ	<u:>	कू	<ku:>
Before	इ	<ɪ>	कि	<kɪ>

before, after, above or below in the manner given in *Table 3*, reproduced from Sproat (2000: 46)

Note that there is no diacritic within a consonant or between the consonants of the onset cluster. (The pre-Nucleus /r/ in the Onset cluster, is specified as a subscript which looks like an inscript, e.g. प्र <pr> as in प्रेस 'press'.)

iii. An underlying unreleased single consonant is represented with a subscript called 'halanta', e.g., वाक् /ʋak/ 'speech', सुप् /sup/ (name of a declensional suffix in Sanskrit), खट् /kʰəʈ/

iv. An underlying unreleased consonant in a cluster is represented in alternative ways, either with the 'halanta' subscript, or with the consonant letter with a part deleted, usually the rightmost element, e.g., उद्धार /uddʰa:r/ emancipation, विस्तार /ʋista:r/ 'expanse', etc.

v. The vowel commonly known as 'schwa' or the neutral vowel is assumed to be without a diacritic. A fully represented underlying consonant without a *halanta* is assumed to have the vowel inherent in it. However, considering the convention to represent underlying unreleased consonants as either with a *halanta* or incompletely, it may be assumed that the vertical line in the consonant cluster of the grapheme represents the neutral vowel for a majority of

consonants. The exceptions are क [k], फ [pʰ] and the retroflex plosives ट [ʈ], ड [ɖ].

vi. A grapheme ऋ <ri>, with a CV structure, is a carry over from Sanskrit segmental system where it represents a syllabic sound, with only a V. The added vowel in Hindi is /i/, as in other languages of northern India. The vowel is /u/ in the languages of southern India. Post-consonantally, the syllabic element is assumed to be represented diacritically as a subscript in certain forms such as कृष्ण /kr̩ʂɳə/, 'Lord Krishna', पृष्ठ /pr̩ʂʈʰə/ 'page', etc.

vii. Onset clusters are treated as a single constituent around which the vowel diacritics are marked. Onset formation involves the following

3.2 Two-consonant clusters

i. Generally, half the letter of the first consonant precedes the full letter of the second consonant: e.g., स्क<sk>, प्त <pt>, क्ल<kl> etc. Alternatively, the practice of specifying the diacritic for unreleased consonants, known as 'halanta', is used for the first consonant, e.g., द्भ<dbʰ> उद्भव /udbʰəʋ/

ii. For C+r cluster, as noted above, the /r/ is specified as a subscript that looks like an inscript: क्र <kr>, ख्र <kʰr>, फ्र <pʰr>.

iii. For r+C clusters, the the /r/ is specified as a superscript above the grapheme, e.g., र्म <rm>, र्त <rt>

iv. In the case of the following two-consonant clusters, a new ligatured group is formed. These are: त्र <tr>, क्ष <kʃ>, ज्ञ <gɲ>, श्र <ʃr>, क्त <kt>.

3.3 Three-consonant clusters

i. Generally, the first two consonants are specified for half their letters, and the third is fully specified, e.g., स्प्ल <spl>. This convention is usually followed for borrowed words.

ii. For C+C+r clusters, and for r+C+C clusters, which are highly restricted, the convention for two-consonant clusters applies, e.g., स्व्ल <spr> स्व्ल <r.

4. Phonological insights in Devnāgarī script

As a descendent of Brahmi, Devnāgarī embodies in itself some of the basic phonological insights of ancient Indian grammarians. The most noticeable and significant of these are discussed below:

1. The akshara is the minimal articulatory unit.

2nd proofs

2. A vowel can be an independent unit of akshara word-initially or post-vocalically. A consonant can also function as an independent unit in contexts where the following /ə/ has been deleted on the surface.
3. Vowels and consonants are assumed to be different types of units and are so represented in the grapheme.
4. The long vowels are systematically represented as related to the short vowel counterparts and being derived from them, for example अ /ə/ and आ /a:/, इ /i/ and ई /i:/, उ /u/ and ऊ /u:/. In a similar manner, the long mid-open vowels, औ /æ əo/ are represented as derived from the long mid-close vowels, ओ /e: o:/.
5. For consonant letters, the neutral vowel /ə/ is assumed to be inherent in them and is pronounced as such word initially and medially in certain contexts, for example, the first grapheme in पल <pəl>. The inherent neutral vowel is not pronounced word-finally or medially in certain contexts (see sec. 6.2.1).
6. A majority of conjunct onset clusters are represented as single constituents, e.g., क्ष <kʃ> and ज्ञ <gɲ>
7. Some of the conjunct letters, e.g. त्र <tr>, क्ष <kʃ>, ज्ञ <gj)> or <gɲ> and श्र <ʃr>, are assumed to constitute not only a single constituent, but also a single onset, like an affricate. These are the historical remnants of what might have been single consonants of Sanskrit.
8. Some of the graphemes represent historically distinct phonetic units that are neutralized in Hindi, e.g. श [ʃ] and ष [ʂ].

Most of the properties of phonology underpinning Hindi orthography as described above relate to the discussion of phonological properties assumed in the recent literature, as discussed in § 2.

5. Hindi phonology and orthography

A consideration of the correspondence between the topography of Devnāgarī and Hindi speech gives the impression that it is regular and systematic. For a section of the vocabulary which includes forms from Middle Indo-Aryan and Modern Indo-Aryan, known as 'tadbhava' forms, this is largely true, as also for recent borrowings from English and other languages. For the forms which are direct borrowings from Sanskrit, however, known as 'tatsama' forms, the correspondence is not consistent, as these forms are at variance with contemporary Hindi speech. We discuss below the main aspects of phonetic and phonological facts that the orthography represents.

5.1 Phonological facts of representation

5.1.1   Input forms
Hindi orthography excludes the effects of the phonological processes of Schwa Deletion, Nasal Assimilation (optionally), Consonant Gemination, and Word-final Lengthening, and represents forms that are inputs to these processes. The processes are illustrated with the help of relevant data below.

a.   Schwa Deletion
Consider the data in *Table 4*.

Table 4. Schwa deletion in Hindi

Stem	Devanāgarī	Gloss	Words	Devanāgarī	Gloss
nəkəl	नकल	copy	nəkli:	नकली	Counterfeit
sərəl	सरल	easy	sərla:	सरला	(a name)
ba:hər	बाह	outside	ba:hri:	बाहरी	External
dərʃən	वर्शन	view	da:rʃnik	वार्शनिक	Philosopher
ɟəŋgəl	जंगल	jungle	ɟəŋli:	जंगली	Wild

The deletion of the schwa in Hindi is subject to several conditions discussed in detail in Ohala (1983) and Pandey (1990). Some of the conditions are : (a) occurrence of the schwa in a weak position in the metrical foot, as for the word [sərla:], which has the metrical structure $_{Wd}[_{\Sigma s}[_{\sigma s}[sə] _{\sigma w}[rə]] _{\Sigma w}[la:]]]$, (b) permissible clusters resulting from the deletion (e.g. *पलड़ा [pələɽa:] 'side of a balance', *सम्बन्धी [səmbəndʰi:] 'relative' etc.), absence of morphological boundary preceding the syllable with the deleted schwa, presence of an onset before the schwa, etc. As is clear from the orthographic representations of the forms, there is no clue regarding the deletion of the schwa in it.

b.   Nasal Assimilation (anusvāra)
Hindi shares the feature of nasal assimilation along with many other Indian languages, as well as Sanskrit. Some examples to illustrate the phenomenon are given in *Table 5* below.

Table 5. Homorganic nasals in Hindi

Words	Gloss	Devanāgarī
kəmbəl	blanket	कंबल
gənda:	dirty	गंदा
ɟəŋgəl	jungle	जंगल
tʰəɳɖa:	cold	ठंडा
rəɲɟ	unhappy	रंज

Hindi orthography has two ways of representing the hormorganic nasals, one by means of an archiphonemic device of 'anusvāra' marked as a superscript on the preceding vowel, as mentioned in Section 3.2, and the other by representing the nasal as a part of a conjunct akshara. It is in the 'anusvāra' representation that the homorganic nasals are not manifest in the orthography.

It should be noted that on account of schwa deletion, nasal phonemes can occur adjacent to a heterorganic consonant in surface pronunciation. The orthography however assigns a different akshara to the heterorganic nasal phoneme, as in *Table 6*.

Table 6. Heterorganic nasals in Hindi

Words	Gloss	Devanāgarī
səmtəl	plain surface	समतल
sənki:	eccentric	सनकी

The retroflex, palatal and velar nasals are allophones of the alveolar nasal /n/, and as such they do not occur as isolate aksharas.

c.   **Consonant Gemination**

Morpheme-internally, consonants are geminated when they precede a glide or a liquid, i.e. /ʋ j r l/, for example:

Table 7. Consonant gemination in Hindi

Input Forms	Devanāgarī	Gloss	Output Form
/əvəʃjə/	अवश्य	certainly	[əvəʃʃj]
/ʃəklə/	शक्ल	face	[ʃəkkl]
/sətʋə/	सत्व	essence	[səttʋ]
/mitrə/	मित्र	friend	[mittr]

Hindi orthography represents the non-geminated input forms.

Note that the process is found to be carried over into Hindi English or perhaps general Indian English, e.g. [sekjoːr] 'secure', [ælɹebraː] 'algebra', etc.

d.   **Word-Final Lengthening**

A general phonological process of final vowel lengthening (of the short vowels [i u], but not [ə]) is found in all varieties of Hindi, except the very formal one, as can be seen in *Table 8* below.

In all these cases, Hindi orthography represents the input forms with final short vowels. It should be noted that the difference in the very formal and other varieties of Hindi with regard to final vowel lengthening has given rise to two patterns of word-stress in Hindi, noted in Pandey (1989), on account of stress applying to different input forms for stress assignment: [əˈditi] ~ [ˈəditiː].

Table 8. Word-final Lengthening in Hindi

Input Forms	Devanāgarī	Surface form	Gloss
/əditi/		[əditi:]	
/ʃəni/	शनि	[ʃəni:]	'Saturn'
/əpitu/	अपितु	[əpitu:]	'but'
/mənu/	मनु	[mənu:]	(a name)

5.1.2   *Output forms*

Hindi orthography represents the output forms of the processes of Flapping and Nasal Assimilation (optionally).

a.   **Flapping**

The voiced retroflex plosives /ɖ and ɖʰ/ in Hindi have flap allophones in native vocabulary. The flaps occur inter-vocalically and word-finally on the surface, as in *Table 9*.

Table 9. Flapping in Hindi

Surface form	Devnāgarī	Gloss
[ga:ɽi:]	गाड़ी	train
[pəɽʰo:]	पढ़ो	read (Imp.)
[pe:ɽ]	पेड़	tree

Srivastava (1968) and Pandey (2002) have argued in favour of there being a final schwa in the input forms for a uniform analysis of the flapping data. Both word-internally and finally a flapped [ɽ ɽʰ] is followed by a vowel, in the latter case a schwa. Hindi orthography represents the surface forms of flapped retroflexes.

Notice, however, that Flapping does not apply to borrowed vocabulary. Thus, for instance, in the borrowed words from English, such as रोड /ro:ɖ/ 'road' and सोडा /so:ɖa:/, the retroflex is a stop, not a flap.

b.   **Nasal Assimilation (Conjunct aksharas)**

The phenomenon of nasal assimilation has the alternative representation in Devnagari of having the homorganic nasals as part of conjunct aksharas, as discussed above. The forms in (1) thus have the alternative representations in (2):

(2)   कम्बल   गन्दा जङ्गल   ठण्डा रञ्ज

As the palatal and velar nasals are allophones in Hindi, the representation in (9) clearly is at the allophonic or surface level.

5.2   Historical phonological facts

Hindi orthography also represents facts of historical phonology of Hindi (see Salomon 2003). Essentially, these are a few segments of Old Indo-Aryan. They are listed below.

i.   **Voiceless retroflex fricative /ʂ/**

The segment is represented both independently (ष) as well as homorganic with the following retroflex plosive, e.g.

Table 10.   Retroflex fricatives in Hindi orthography

Devnāgarī	Surface form	Gloss
शेष	[ʃeːʃ]	remainder
वर्ष	[vərʃ]	year
कष्ट	[kəʂʈ]	pain
कृष्ण	[kriʂṇ]	'Lord Krishna"

ii.   **Syllabic approximant /ɹ̩/**

The segment is represented as a full akshara in Devnagari as ऋ, and as a diacritic subscript (ृ), when it occurs as the second member of a conjunct akshara:

Table 11.   Nonsyllabic /r/ in Hindi from syllabic /ɹ̩ / in Sanskrit

Devnāgarī	Sanskrit form	Output form	Gloss
ऋषि	/ɹ̩ʃi/	[riʂiː]	Sage
पृष्ठ	/pɹ̩ʂʈʰ/	[priʂʈʰ]	Page

The syllabic approximant has two different manifestations in modern Indian languages: as /ri/ in north Indian languages and as /ru/ in sout-west and southern languages. Thus in languages such as Marathi, Gujarati and all the Dravidian languages, the equivalent forms for the words in *Table11* are /ruʃi/ and /pruʂʈʰ(ə)/. None of the present Indian languages have the syllabic approximant in them.

iii.   **Graphemes for consonant clusters**

There are certain graphemes which are pronounced as consonant clusters but which are represented as single graphemes, following the practice for Sanskrit, e.g. ज्ञ <gnə>, as in ज्ञान [gna:n] 'knowledge', and क्ष <kʃ>, as in क्षति [kʃəti] 'loss'. The phonetic correspondence of the graphemes in present-day Hindi is quite regular as the consonant clusters, rather than as a single sound unit.

5.3   Orthographic and phonological levels

The various aspects of the phonological facts of Hindi that the orthography represents are inconsistent with regard to the level of representation in phonology, the notion of 'level' being assumed to be the same as in derivational phonology. It appears, however, that this situation is not peculiar to Hindi, but can be found to be true for other orthographies that have early origin and continuity, such as English.

The notion of 'inconsistency' of the orthography, however, applies to the forms in relation to the phonological and phonetic representations, but not in relation to historical facts, as can be seen from the discussion above. The forms relating to history can be explained in terms of a correspondence with the regular facts of orthography. It can thus be assumed that users establish correspondence of the historical symbols ऋ ष etc. with the values represented by the regular symbols.

6.   **Reform in Hindi orthography**

Any proposal for a reform of the existing orthography must be based on a consideration of the fundamental facts of writing systems, and their relation with linguistic and non-linguistic aspects of cognition. On a careful consideration, it can be shown that the problems of orthography must be divided into those of production and perception. While the latter are much less tractable and more open and intriguing, with many contending models of explanation, the former are more easily amenable to examination, especially from the educational perspective. I submit the following proposals in awareness of the complexity of the issue involved:

1.   The basic unit of the orthography, akshara, should be kept as such, and need not be modified as a fully syllabic type, such as Hankul. It has been recently argued (see Pandey, 2007a, Patel, 2007) that the akshara may well be the minimum articulatory unit of speech.

2.   For a regular relation between spelling and speech, it is desirable that the relation between spelling and speech in Hindi is consistent with regard to a linguistic level. As is obvious from the discussion above, that relation is inconsistent. Whereas forms in § 5.2.1 relate to the underlying level, those in § 5.2.2 relate to the output level. The forms in § 5.2.3 relate to neither, but rather to an earlier, historical stage. Towards the goal of reducing the inconsistency in the relation between sound and spelling in Hindi, we suggest the following.

   a.   For the numerous Sankritic loanwords in Hindi in § 5.3, the letters representing the retroflex fricative ष /ʂ/, the syllabic approximant ऋ /ɹ̩/ and the

2nd proofs

conjunct letters श्र /ʃr/ ज्ञ /gɲ)/ क्ष /kʃ/, should be dropped from the Hindi script, thereby leading to a more economic and more easily learnable alphabet for synchronic Hindi.

b. For the forms representing the underlying level of sounds, they can be kept as they are, except for the word-final length neutralization. My reasons for allowing a change for the latter but not for the others in this category are as follows:

  i. The full letter representation of surface consonants not followed by vowels should be kept as such for various reasons. One, in cases such as समतल [səmtəl] and सनकी [sənki:], the lack of nasal place assimilation, i.e. * [səntəl] and *[səŋki:], is accountable, compared with forms in which there is nasal assimilation, such as संथाली/सान्थाली [sa:ntʰa:li:] शंका / शङ्का [ʃəŋka:]. Two, the morphophonemic relatedness of forms can be preserved, as for example in तबला /təbla:/ 'tabla (a percussion instrument)' and तबलची /təbəlci:/ 'one who plays the tabla'

  ii. Gemination of consonants preceding approximants on account of a phonological process, as pointed out in §5.2.1.c, need not be represented as such, since they are entirely predictable, and the representation of double consonants will be cumbersome in certain cases, as in अवश्य [əvəʃʃj] and शक्ल [ʃəkkl]. Gemination in these forms is obligatory and a commonly shared phenomenon with other Indian languages, such as Bangla, Marathi and Telugu. (However, lack of representation of gemination in these forms is a problem for foreign language learners, such as speakers of English or French, languages in which such a process is not found.)

  iii. The short and long vowel distinctions in Hindi spelling can be neutralized for the word-final position. (Note that the formal varieties of Hindi which maintain the distinction in that context are highly restricted.). Word-medially, the the orthographic representation of the neutralization of vowel length distinction is not fruitful on the grounds that medially, the neutralization is a consequence of shortening (not lengthening, as in the final position). For four of the ten vowel letters, namely, the vowels /e:/ ऐ /æ/ ओ /o:/ औ /əo/, there are no short vowel letter counterparts, so the shortened vowels in those cases cannot be represented there. Note that this would not be a problem for a language such as Tamil or Telugu, which has pairs of short and long vowel letters in the alphabet.

c. For forms showing variability in spelling on account of allowing for both underlying and output level representations, as in the case of anusvāra and anunāsika representations for homorganic nasals in consonant clus-

ters (§ 5.1.1.b and § 5.1.2.b), the variability in spelling should be avoided. The homorganic nasals in input forms can best be represented using the anusvāra, as in §5.1.1b., rather than as parts of conjunct aksharas, as in §5.1.2b. Apart from avoiding unnecessary ambivalence in graphemic choice, this will help jettison at least two nasal graphemes, [ɲ] and [ŋ], which have very limited distribution elsewhere in the orthography.

d. The only remaining output level link between phonology and script in Hindi, namely the retroflex flaps ड़ /ɽ/ and ढ़ /ɽʰ/ (§ 5.2.2.a), can be kept as such, even though they represent a regular process, Flapping. As discussed above, the process only applies to native words. Loan-words, especially from English, do not undergo Flapping. The flapped and the non-flapped distinction in orthography helps distinguish the pronunciation of retroflex stops in native and borrowed words.

7. Conclusions

I have been concerned in this paper with the description and anlysis of the phonetic/ phonological bases of the Devanāgarī script as it is used for the witing of modern standard Hindi. In presenting my analysis I have tried to bring together the received insights of modern phonetic and phonological theories into the organization of speech sounds in human language. I have also tried to take into account current research on writing systems that has contributed to the understanding of writing as a representational system. The discussion of the insights from the advances in phonetics, phonology and writing systems forms the basis of the account of the main features of the akshara based character of Devanāgarī script.

I have attempted to demonstrate in detail the close relationship between phonology and orthography in Devanāgarī. I gave concrete examples to indicate the exact nature of that relationship with a view to making linguistically valid proposals for reform in present-day Hindi orthography. The proposed reforms centre around the goal of reducing the inconsistencies between the written and the spoken standard synchronic Hindi to the extent possible.

Notes

* A version of the paper was presented at the International Symposium on "Indic Scripts: Past and Present", December 15–18, 2003, IULC, Tokyo.

1. The only script derived from Brāhmī that has moved away from this trait is the Hankul script of Korea (Coulmas, 1997).

2. The term 'akshara' has been used in various senses. For a summary of some of its uses, see Kapoor (2007). I use the term in the sense of a unit of orthography in Brāhmī and its derivatives. The structure of the akshara in relation to the syllable

3. Sanskrit phoneticians (for instance, the writer of the phonetic text vājɹəsəneyī *Pratishākhya*) consider the post-vocalic nasal to have greater vocalicness than consonantality in it, and thus represent it as part of the preceding akshara.

References

Agrawala, V. S. (1966). The Devnagari script. In *Indian Systems of Writing*, (pp. 12–16). Delhi: Publications Division.

Anderson, J. M. & Ewen, C. J. (1987). *Principles of dependency phonology*. Cambridge: CUP.

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (ed.), *The handbook of phonological theory*, (pp. 206–244). Oxford: Blackwell.

Bright, W. (1996). The Devanagari script. In P. Daniels & W. Bright (Eds.), *The world's writing systems,* (pp. 384–390). New York, NY: OUP.

Browman, C. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*: 201–251.

Browman, C. & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes, and historical change. In I. Mattingley & M. Studdart-Kennedy (Eds.), *Modularity and the motor theory of speech perception,* (pp. 313–38). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bybee, J. (2001). *Phonology and language use*. Cambridge: CUP.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: The MIT Press.

Clements, G.N. (1985). The geometry of phonological features. *Phonology Yearbook, 2*: 225–53.

Clements, G.N. (1999). Affricates as non-contoured stops . In O. Fujimura, B.D. Joseph, & B. Palek (Eds.), *Proceedings of LP '98: Item order in language and speech,* (pp. 271–299). Prague: The Karolinum Press.

Clements, G.N. & Keyser, J. (1983). *C V phonology: A generative theory of the syllable*. Cambridge, MA: The MIT Press.

Coulmas, F. (ed) (1997). *The blackwell encyclopedia of writing systems*. Oxford: Blackwell.

Kapoor, K. (2007). Akshara in Indian Thought. In P.G. Patel, P. Pandey & D. Rajgor (Eds.), *The Indic scripts: Palaeographic and linguistic perspectives,* (pp. 1–8). New Delhi: DK Printworld.

Kaye, J, Lowenstamm, J. & Vernaud, J-R. (1985). The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook 2*: 305–28.

Kenstowicz, M. (1994). *Phonology in generative grammar. London:* Blackwell.

Klima, E. (1972). How alphabets might reflect language. In J. Kavanagh & I. Mattingley (eds.), *Language by ear and by eye: The relationships between speech and reading*, (pp. 57–80). Cambridge, MA: The MIT Press.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. London: Blackwell.

McCarthy, J. (1981). A prosodic theory of non-concatenative phonology. *Linguistic Inquiry 12*: 443–66.

McGregor, R. S. (1977). *Outline of Hindi grammar*. 2nd edn. Delhi: OUP.

Ohala, M. (1983). *Aspects of Hindi phonology*. New Delhi: Motilal Banarasidass.

Pandey, P.K. (1989). Word accentuation in Hindi. *Lingua, 77*: 37–73.

Pandey, P.K. (1990). Hindi schwa deletion. *Lingua, 82*: 277–311.

Pandey, P.K. (2002). Hindi me shabdant ə. (Word final ə in Hindi). *GaveshaNa*. Central Institute of Hindi, Agra.

Pandey, P.K. (2007a). Akshara as the minimum articulatory unit. In P.G. Patel, P. Pandey & D. Rajgor (Eds.), *The Indic scripts: Palaeographic and linguistic perspectives*, (pp. 167–232). New Delhi: DK Printworld.

Pandey, P.K. (2007b). Phonological and generative aspects of Brāhmī and its derivatives. In P.G. Patel, P. Pandey & D. Rajgor (Eds.), *The Indic scripts: Palaeographic and linguistic perspectives*, (pp. 233–248). New Delhi: DK Printworld.

Patel, P.G. (1996). Linguistic and cognitive aspects of the Orality-Literacy complex in Ancient India. *Language and Communication*, *16*: 315–29.

Patel, P.G. (2005). *Reading acquisition in India*. New Delhi: Sage.

Patel, P.G. (2007). Akshara as a linguistic unit in Brāhmī scripts. In: P.G. Patel, P. Pandey & D. Rajgor (Eds.), *The Indic scripts: Paleographic and linguistic perspectives*, (pp. 167–215). New Delhi: DK Printworld.

Prince, A. & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar* [Technical Report 2 of the Rutgers Centre for Cognitive Science]. New Brunswick, NJ: Rutgers University.

Salomon, R. (2003). Writing systems of the Indo-Aryan languages. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages*, (pp.67–103). London: Routledge.

Selkirk, E. (1982). Syllables. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations*, Vol II., (pp. 337–383). Dordrecht: Foris.

Singh, A. K. (2007). Progress of modification of Brāhmī alphabet as revealed by the inscriptions of sixth-eighth centuries. In P.G. Patel, P. Pandey & D. Rajgor (Eds.), *The Indic scripts: Paleographic and linguistic perspectives*, (pp. 85–107). New Delhi: DK Printworld.

Spencer, A. (1996). *Phonology: Theory and practice*. London: Blackwell.

Sproat, R. (2000). *A computational theory of writing systems*. Cambridge: CUP.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21*: 291–311.

Srivastava, R.,N. (1968). Theory of morphonematics and aspirated phonemes of Hindi. *Acta Linguistica, 18*: 363–73.

Steriade, D.. (1995). Markedness and underspecification. In J.Goldsmith (ed.), *Handbook of phonological theory,* (pp. 114–174). Oxford: Blackwell

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. Mackay, W. Prinz & E. Scheerer (Eds.), *Language, perception and production*, (pp. 67–84). New York, NY: Academic Press.

Studdert-Kennedy, M. (1998). The particulate origins of language generativity: From syllable to gesture. In J. Hurford, M. Studdart-Kennedy & C. Knight (Eds.), *Approaches to the evolution of language*, (pp. 202–221). Cambridge: CUP.

Venezky, R. (1970). *The structure of English orthography* [No. 82 in Janua Linguarum]. The Hague: Mouton.

2nd proofs