

# A COMPARATIVE PHONOLOGICAL STUDY OF THE DIALECTS OF HINDI

*Diwakar Mishra<sup>a</sup> & Kalika Bali<sup>b</sup>*

<sup>a</sup>Jawaharlal Nehru University, India; <sup>b</sup>Microsoft Research India, India  
diwakarmishra@gmail.com; kalikab@microsoft.com

## ABSTRACT

Dialectal variations provide vital cues to both synchronic and diachronic changes in sounds of a language. There has been no comparative phonological study of the dialects of Hindi in the last several decades. In this paper, we present a phonological description of seven of the major dialects of Hindi, namely, Awadhi, Bagheli, Bhojpuri, Bundeli, Haryanvi, Kanauji and Khari Boli, based on the observation and analysis of telephonic conversational data. We believe that these preliminary results will as a starting point for a more comprehensive and detailed comparison of the dialects and provide insights for language evolution as well as synchronic variations of Hindi.

**Keywords:** Hindi, comparative study, dialects

## 1. INTRODUCTION

The variations represented by the dialects of a language can be viewed as a sub-set of the diachronic variation of a language. Therefore, a comparative study of the phonetics and phonology of a language can provide valuable insight into historical sound-changes of a language. While there have been many descriptive studies of the standard dialect of Hindi (Khari Boli), for example, [1, 6, 7, 9] and relatively fewer of some of the important dialects [3, 4, 8, 10], in the recent decades there have been none that offer a comparative analysis of even the basic sound inventory of the dialects. One of the important studies that attempted to compare the sound systems of three major dialects of Hindi was done in 1958 by Gumperz [3]. Thus, though there is a lot of reference in the popular media on the current use of Hindi especially w.r.t. Hinglish-, there is no literature on the current status of the sounds of Hindi language and its dialects [5].

Further, from the point of view of speech technology applications, both for recognition and generation, there is a need for robustness to dialect variation. However, even the prominent dialects of

Hindi are neglected or dealt with in the most superficial manner for speech technology applications. The primary reason for this is the dependency on large corpus by the current statistical methods that view language as represented by a corpus. As data is more easily procured or collected for the dominant dialect, the other dialectal variations are largely ignored. This is a big challenge for Indian languages, like Hindi, where there is a scarcity of adequate databases even for the standard dialect. A comparative study of the phonological sound system of the dialects, we hope, is the first step that could help formulate transfer rules across dialects that might help redress this issue.

In this paper, we present a phonological description for seven prominent dialects of Hindi, viz., Awadhi, Bagheli, Bhojpuri, Bundeli, Haryanvi, Kanauji and Khari Boli, based on the observation and analysis of recorded telephone conversations in each dialect. The next section will provide a description of Hindi and its dialects. Section 3 will discuss the corpus analyzed and the methodology followed. In Section 4 we will list the consonant and vowel inventories of each dialect and their allophonic variations. The last section will discuss potential future work.

## 2. HINDI AND ITS DIALECTS

Hindi is the primary official language of India and the official language of ten of her states. Nearly 41% of the Indian population speaks Hindi or one of its dialects as its first language, and more than 70% Indians can understand and speak Hindi to a communicative level. Hindi is the lingua franca in many non-Hindi speaking states, such as the north-eastern Indian state of Arunachal Pradesh, and is the second most spoken language after Bangla in Andaman Islands and in north-eastern India [2].

According to the 2001 Census of the Government of India [2], there are at least 50 dialects of Hindi with speaker population varying from around 11,000 for Khairari to more than 250 million for standard Hindi. These varieties can be

roughly divided into two groups, western and eastern [6]. The standard dialect arose through a considerable borrowing of vocabulary from other dialects and languages into a grammatical skeleton based on the dialect spoken in and around the Delhi area [1, 9]

The seven dialects considered in this study are:

- Awadhi: This dialect is spoken in the middle regions of Uttar Pradesh also called Awadh, and in the nearby districts of Lucknow and Faizabad with 2.5 million speakers
- Bagheli: This dialect is spoken in the middle and southern parts of Madhya Pradesh by 2.8 million speakers
- Bhojpuri: This dialect is spoken in a wide area – eastern part of Uttar Pradesh and middle and southern part of Bihar and Jharkhand by a population of 33 million
- Bundeli: This dialect is spoken in the south-west marginal districts of Uttar Pradesh and wide northern area of Madhya Pradesh by 3 million speakers
- Haryanvi: This dialect is spoken in Haryana, Delhi and some marginal areas of Uttar Pradesh, Himachal Pradesh and Uttarakhand by a population of about 8 million
- Kanauji: This dialect is spoken in south western part of Uttar Pradesh and in the west of the Awadhi region by a population of 6 million
- Khari Boli: This dialect originally belongs to the western parts of Uttar Pradesh and Delhi and spoken in some parts of Uttarakhand, Himachal Pradesh. This dialect is considered to be the standard form of Hindi and accepted as such throughout the Hindi speaking area.

### 3. METHODOLOGY

#### 3.1. Data

The speech corpus used for this study was collected by Appen Technologies in 2006. This is a proprietary database not available in public domain. The corpus was collected from seven different locations of Hindi speaking area each representative to one dialect. The map in Figure 1 shows the areas where these dialects are spoken. The database collection was done through recorded telephone conversations. The target was to collect speech samples of 1000 speakers, out of which 700 are males, and 300 are females, from 700 mobile and 300 are landline phone conversations. The per-dialect target was 143 speakers. The actual

collection (996 speakers) is almost near the target and constitutes over 60 hours of recorded speech. The speech data is transcribed both in Devanagari and Roman scripts.

On a closer observation of the data it was apparent that though the speech was supposed to represent different dialects, a number of speakers actually used the standard forms. This was true across all dialects but was more prevalent in certain dialects. Chhattisgarhi, for instance, was almost always replaced by the standard form and hence left out of current analysis. A total of 70 conversations (approximately 10 per dialect) were analyzed constituting nearly 4 hours of speech.

**Figure 1:** The map of the Hindi speaking regions of India indicating the areas where the selected dialects are spoken.



#### 3.2. Methodology

The overarching goal of the project is to conduct an in depth comparison of the different dialects of Hindi. For this purpose, the following linguistic patterns are to be considered for comparison:

1. Basic vowel inventory
2. Diphthong formation
3. Allophonic variation
4. Basic consonant inventory
5. Aspiration of consonants
6. Morphological inflections and their variations
7. Verb base formation
8. Auxiliary verbs
9. Tense markings
10. Participle/gerund/infinitive formations
11. Tones or word accents
12. Intonation at the phrasal level (for clause and sentence level)

The first phase of the project looked at 1-5, whereas the second phase considered 6-10. The third phase which will consider the prosodic patterns across the dialects is yet to begin. In this paper, we present the pilot study of the first phase of the project.

Trained linguists heard the selected conversations across the dialects and re-transcribed them. Analysis was done based on auditory observations as well as spectrographic analysis.

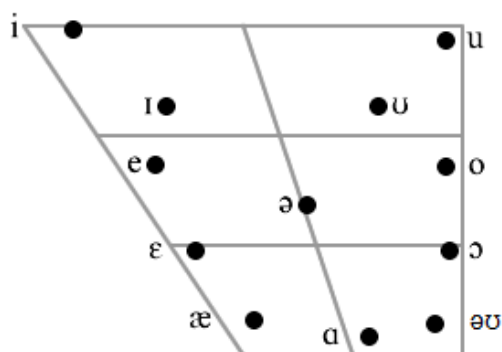
#### 4. A COMPARISON OF HINDI DIALECTS

The comparison of the seven dialects of Hindi was done on the following parameters:

##### 4.1. Basic vowel inventory

Standard Hindi consists of a 10 vowel inventory of native vowels and two borrowed vowels to represent English sounds [7]. Each of these 10 vowels also exists in phonologically distinct nasalized form. Figure 2 gives the vowel chart for standard Hindi.

Figure2: Vowel Inventory for Hindi.



All the seven dialects of Hindi studied here retain the same basic vowel inventory. However, a number of dialects show distinct allophonic variations.

##### 4.1.1. Awadhi

Awadhi shows a number of allophonic free variations for the following vowels:

1. /e:/, /o:/ are in free variation with /ja:/, /wa:/. E.g., /d ja:khau/ ~ /de:khau/.
2. /e:/, /o:/ can shorten as the following consonant lengthens. e.g., /ek: au/ ~ /e:kau/.

##### 4.1.2. Bagheli

In Bagheli, /e/, /o/ occur in complementary distribution with the more common /i/, /u/. e.g., /dustana/ vs /dostana/ “friendship”.

##### 4.1.3. Bundeli

In general, the high vowels in Bundeli tend to be lower than the other dialects. e.g. /bahota den ð se/ vs /bahot ðn ð se/ “from many days”. Otherwise, Bundeli vowels tend to be very similar to the std Hindi inventory.

##### 4.1.4. Bhojpuri

Bhojpuri has the most allophonic variations in vowels:

1. The vowels are generally more open and lower than the other dialects.
2. /a/ tends to be in variation with the rounded and more back /ɔ/
3. /o/ occurs in free variation with /u/. e.g. /du/ vs /do/ “two”
4. Nasalized lax vowels are absent
5. /a/ at the end of a word is often not pronounced in spontaneous speech.

##### 4.1.5. Haryanvi

In general, vowels tend to be more open than other dialects. Thus, /a/ occurs in free variation with /e/, and /i/ with /e/. e.g. /kəha/ vs /kəhe/ “said”

##### 4.1.6. Kanauji

Kanauji vowels show the same distribution as that of Awadhi with the addition of an extra diphthong /aɪ/. E.g., /gaɪ/

#### 4.2. Diphthongs

In standard Hindi, historically two diphthongs /əi/ and /əu/ have transformed to monophthongs /ɛ/ and /ɔ/ respectively. However, in some instances the diphthongs do occur in allophonic variation, for example, when followed by a semi-vowel [7]. The dialects of Hindi show a large diphthong inventory as compared to Std Hindi. Table 1 lists the diphthongs observed in the different dialects of Hindi.

#### 4.3. Basic consonant inventory

Standard Hindi has a set of 37 consonants including those whose distribution is dependent on different cultural and literary registers. All consonants show phonemic length distinction, (exceptions are /b<sup>h</sup>, t, t<sup>h</sup>, f/). These geminates are always medial and preceded by one of the lax vowels (that is, /ə/, /ɪ/, or /ʊ/) [7]. The stop consonants of Hindi show a four way contrast in

terms of voicing and aspirations. However, it may be noted that what are traditionally termed as “voiced aspirated” are phonetically quite distinct in their voicing and release patterns from voiced, and voiceless aspirates [9].

**Table 1:** Diphthong inventory for Hindi dialects. One in the parentheses is the allophonic variant of the preceding vowel.

Dialect	Diphthongs
Awadhi	/əɪ/(əe), /əʊ/(əo), /aɪ/(ae), /ɪʊ/ /ʊɪ/, /eʊ/, /oɪ/, /oʊ/
Bagheli	/əe/(əɪ), /aɪ/(ae), /aʊ/(ao), /ɪʊ/, /ʊɪ/, /eɪ/, /eʊ/, /oɪ/, /oʊ/
Bundeli	/əe/(əɪ), /ɪ/(ae), /aʊ/(ao), /ɪʊ/, /ʊɪ/, /eɪ/, /eʊ/, /oɪ/, /oʊ/
Bhojpuri	/əɪ/(əe,ɛ), /əʊ/(əo,ɔ), /aɪ/, /aʊ/, /ae/, /ɪʊ/, /ʊə/, /ʊɪ/, /ʊe/, /eə/, /ei/, /eʊ/(əʊ), /oʊ/
Haryanavi	/əɪ:/, /əe/(əɪ,ɛ), /əʊ/(əʊ,ɔ), /ae/, /ao/, /ea/(ja:), /eɪ/, /eʊ/
Kannauji	/əɪ/(əe), /əʊ/(əo), /aɪ/(ae), /aʊ/(ao), /ɪʊ/, /ʊɪ/, /eɪ/, /eʊ/(eʊ), /ɔɪ/(oɪ), /ɔʊ/(oʊ), /ɛ/(jɛ), /ɔ/(wɔ)

From the perspective of dialectal variations, almost all the dialects retain the consonant inventory. The few exceptions are:

1. The consonants /f/ and /z/ are not found in almost all the dialects other than Khari-Boli. These two sounds are Perso-Arabic in origin and are considered by some as borrowings.
2. In Awadhi, Bundeli and Bagheli, /w/ is often replaced by /b/. Moreover, /j/ in all of the above, except Awadhi is also in variation with /dʒ/.
3. Haryanvi tends to replace /r/, /n/ and /l/ with retroflex forms, that is, the retroflexion is a marked feature of this dialect.
4. Almost all the dialects, with the exception of Khari Boli do not distinguish consistently between the alveolar fricative /s/ and palatal /ʃ/, the former being the preferred form.

5. In spontaneous conversational speech, there seems to be a case for some h-elision in almost all dialects. This can vary a lot from dialect to dialect with Haryanvi showing the maximum instances of this phenomenon. However, this needs further investigation especially w.r.t. the tonal properties of the adjacent vowels.

## 5. FUTURE WORK

In this paper we have presented a comparative study of the basic vowel and consonants of seven major dialects of Hindi based on the analysis of telephonic conversational speech. This is a preliminary report of an ongoing project with the aim of a more detailed analysis of the dialectal variations in Hindi, including acoustic studies of some of the prominent phenomena and the prosodic patterns. We hope that such a study will provide a more contemporary insight into both diachronic and synchronic variations in Hindi.

## 6. REFERENCES

- [1] Bhatia, T.K. 1996. *Colloquial Hindi – A Complete Course for Beginner*. Routledge, London.
- [2] Census of India website – Census 2001, Statement 5, [http://censusindia.gov.in/Census\\_Data\\_2001/Census\\_Data\\_Online/Language/Statement5.html](http://censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement5.html)
- [3] Gumperz, J.J. 1958. Phonological differences in three Hindi dialects. *Language* 34(2), 212-224.
- [4] Jaiswal, M.P. 1962. *A Linguistic Study of Bundeli (A Dialect of Madhyadēśa)*. Leiden.
- [5] Kachru, Y. 2006. Mixers lyricizing in Hinglish: Blending and fusion in Indian pop culture. *World Englishes* 25, 222-233.
- [6] Kaul, O.N. 2008. *Modern Hindi Grammar*. Dunwoody Press.
- [7] Ohala, M. 1983. *Aspects of Hindi Phonology*. Motilal Banarsidass, Delhi.
- [8] Prasad, S. 2008. *Magahi phonology: A Descriptive Study*. New Delhi: Concept Publishing Company.
- [9] Shapiro, M.E. 1989. *A Premier of Modern Standard Hindi*. Motilal Banarsidass, Delhi.
- [10] Snell, R. 1991. *The Hindi Classical Tradition: A Braj Bhāṣā Reader*. The School of Oriental and African Studies, University of London.