Evaluating the Effect of Phrase Set in Hindi Text Entry

Mohit Jain, Khushboo Tekchandani, Khai N. Truong

HAL Id: hal-01510544
https://inria.hal.science/hal-01510544

Submitted on 19 Apr 2017

Evaluating the Effect of Phrase Set in Hindi Text Entry

Mohit Jain[1], Khushboo Tekchandani[2], and Khai N. Truong[3]

[1] IBM Research, Bangalore, India
`mohitjain@in.ibm.com`
[2] Dhirubhai Ambani Institute of ICT, Gandhinagar, India
[1, 3] University of Toronto, Computer Science, ON, Canada

Abstract. Recently, many different Indic text entry mechanisms have been proposed and evaluated. Whereas the use of a common phrase set across text-entry research may help to produce generalizable results across studies, previous Indic Text entry evaluations have used a variety of different text entry phrases. In this paper, we develop and evaluate three different types of Hindi phrase sets that have been previously used in the literature – Hindi films, a grade VII textbook and a translated version of MacKenzie and Soukoreff's phrases – to study effects of their characteristics on performance. No statistical difference was found in novice user performance due to the different phrase sets. However, based on participant feedback, we report that consideration should be taken with regards to phrase length, frequency, understandability, and memorability in the design and selection of text-entry phrases.

Keywords: Hindi, Text Input, Phrase Set.

1   Introduction

Researchers are increasingly investigating how to best support text-entry for non-English languages, such as Indic languages [3], [5], [6], [7], [8], [9]. Indic languages such as Hindi, Marathi, Nepali, use Devanagari script. Hindi has 53 base letters – 34 consonants, 11 vowels and 8 diacritic marks. Unlike English, Hindi is not disconnected and mostly requires two or more letters to be combined together to form a *character* [6]. The combination of letters changes a character's visual appearance (*e.g.,* typing क(k) and ई(i) produces की(ki)).

The many challenges involved with entering text in Indic languages, including Hindi, have been discussed at length by others; for more detail, consult [5], [6], [7]. One of the key challenges is that for some characters, there are differences between the phonetic sequence of letters and the visual sequence of writing the letters (*e.g.*, character पि(pi) in the word पिता(pitaa) requires writing the vowel इ(i) before the consonant प(p), but pronounce the इ(i) after the प(p)) [6]. Furthermore, multiple base letters share a key button, often resulting in the use of multiple keystrokes to enter a

letter (*e.g.,* typing Shift + L for थ). Many novel hardware and alphabetical layouts have been proposed to support effective Indic text entry on computers [3], [6], [8], [15] and mobile devices [5], [6], [9], [14], [17]. For evaluation, participants enter phrases of text using a technique while data is logged to calculate text entry performance. Due to the unavailability of a standard Hindi phrase set, researchers have used a variety of alternatives, including random words [9], random phrases [3], [6], paragraphs from a news portal [8], phrases from textbooks [5], and Hindi films dialogues [7]. Different phrase sets may have different properties, which can affect the text-entry task [21] and can potentially bias performance. Read [16] argues that when constructing text phrases, language and context must be taken into consideration.

In this research, we test three different Hindi phrase sets representing those that have been commonly used in previous studies. The phrases were taken from Hindi films (similar to [7]), grade VII Hindi textbook (similar to [5]), and a Hindi translated version of the MacKenzie and Soukoreff's phrase set [10] (state-of-the-art for English text entry) [http://www.dgp.toronto.edu/~mjain/HindiTextEntry.zip]. We show that the phrase sets differed in a variety of ways, such as the frequency of the words used, the length of the phrases, their correlation with a Hindi corpus, and the readability of the phrases. We conducted a study to examine whether these properties can bias performance of novice Hindi text entry users. We did not find any statistical difference exists in user performance, in terms of input speed and error rate, due to the different phrase sets. However, Mackenzie and Soukoreff [10] previously argue that use of a standard phrase set for English will help in attaining more generalizable results across studies. The same argument should hold true for a standard Hindi phrase set, and hence using these standard phrase sets for future Hindi text entry research will help in generalizing results. Additionally, based on participant feedback, consideration should be taken with regards to phrase length, frequency, understandability and memorability in the design of the phrases. In particular, our post-study questionnaire shows that ~8 words per phrase is an effective phrase length and familiar phrases minimize learnability and help motivate the participants.

2   Phrase Sets

We developed three different types of phrase sets previously used in the literature. Similar to [10], none of the phrase sets include any punctuation.

Phrases from Hindi Films (FP).
Previous Hindi text entry research has used phrases from Hindi films [7], as they are very memorable and capture the attention of the audience. We generated a phrase set comprising of 60 famous phrases from Hindi films. The phrases were randomly selected from online forums and blogs, citing famous Hindi film dialogues. Example phrases include: हम जहाँ खड़े हो जाते हैं लाइन वहीं से शुरू होती है (*where I stand, the queue begins from there*), डान को पकड़ पाना मुश्किल ही नहीं नामुमकिन है (*to catch Don is not only hard but impossible*).

Phrases from Textbooks (TP).

Grade VII textbook phrases have been used previously for evaluating Hindi text entry [5]. We generated a phrase set consisting of 50 phrases from a grade VII standard Hindi textbook [12]. We extracted phrases from a paragraph from 6 different stories in the book, thus maintaining some relationship between consecutive phrases. Example phrases include: दिव्या अनिल कि छोटी बहन है (*Divya is Anil's younger sister)*, यों तो वह शुरू से ही कमज़ोर है लेकिन इधर कुछ दिनों से उसे हर समय थकान महसूस होती रहती है (*we know she is weak from the start but the past few days she seems tired all the time*).

Translated MacKenzie and Soukoreff's Phrase Set (MSP).

The phrase set provided by MacKenzie and Soukoreff [10] is considered as standard and been used extensively for the evaluation of English-based text entry techniques. One of the authors, who is a native Hindi speaker, translated the phrase set into Hindi. In some instances, more context-appropriate words were used in place of English proper nouns; for example, *spaghetti*, *racketball*, and *Sam* were replaced with words with Indian context such as खिचड़ी (*khichdi* – an Indian food dish), क्रिकेट (*cricket*), and राम (*Ram* – an Indian name) respectively. Additionally, some English adjectives also do not have a direct translation in Hindi. We used the closest matching adjective with a high frequency in the Hindi language. Example phrases include: हाथी चूहों से डरते हैं (*elephants are afraid of mice*), प्यार के कई मतलब है (*love means many things*).

To verify the validity of the translated phrase set, two native Hindi speakers who are fluent in English were hired to rate the translation in three different parameters. The raters were provided with the English phrase set [10] of 500 phrases along with their Hindi translations. First, the raters were asked to rate the Hindi translated phrase according to this rating scheme – 1: *Directly Translatable* such that the English and the Hindi phrase mean exactly the same. 2: *Indirectly Translatable* such that synonyms must be used to keep the intended meaning (for example, "ginormous" to "very large"). 3: *Indirectly & Partially Translatable* such that words required changes to gain similar meaning (for example, "Sam" to "Ram", "spaghetti" to "khichdi"). 4: *Not Translatable* such that meaning is lost because the phrase contains idioms, colloquial expressions, etc. Second, these raters were asked to rate a Hindi phrase as "Y" if it is a valid Hindi phrase, "N" otherwise. A high-correlation was found between the two raters (95.45%). The rater rated 255 phrases as Directly Translatable, 173 as Indirectly Translatable, 12 as Indirectly Partial Translatable, and 60 as Not Translatable. Raters agreed that 444 phrases were valid Hindi phrases. For the study, we chose the first 150 valid phrases, which were rated as 1 by both the raters.

Linguistic Analysis of the Phrase Sets.

An important characteristic that MacKenzie and Soukoreff [10] have established about their phrase set is its high frequency-based correlation with the English corpus. MacKenzie and Soukoreff's phrases vary from 16 to 43 characters (m=28.61). There are 2712 words (1163 unique) varying from 1 to 13 characters (m=4.46). It has a 0.945 *single-letter correlation* with the English letter frequencies reported by Mayz-

Table 1. Linguistic Analysis of the Phrase Sets and Corpus

Metrics	EMILLE/CIIL Corpus	FP	TP	MSP	MS English Set
Number of phrases/sentences	737528	60	50	150	500
Number of words	12295677	490	673	881	2712
Number of unique words	202042	267	382	464	1163
Minimum word length	2	2	2	2	1
Maximum word length	33	10	13	14	13
Min. phrase length (# words)	1	4	3	3	3
Max. phrase length (# words)	888	14	39	11	9
Min. phrase length (# letters)	1	16	10	12	16
Max. phrase length (# letters)	4752	58	167	49	43
Single-letter correlation	-	0.97	0.98	0.98	0.95
Word-based correlation	-	0.70	0.68	0.75	0.85
Readability	m=10.34 sd=6.76	m=5.36 sd=2.4	m=8.0 sd=3.82	m=5.68 sd=2.46	m=4.17 sd=3.88
Words per phrase	m=16.67 sd=13.27	m=8.16 sd=2.4	m=13.46 sd=7.45	m=5.87 sd=1.6	m=5.4 sd=1.1
Letters per phrase	m=83.34 sd=67.4	m=35.45 sd=10.15	m=61.44 sd=34.63	m=26.82 sd=7.08	m=28.61 sd=5.02
Letters per word	m=4.06 sd=2.16	m=3.46 sd=1.44	m=3.63 sd=1.65	m=3.73 sd=1.72	m=4.46 sd=2.4

ner and Tresselt [11]. *Word-based correlation* shows that 85.08% of the words in MacKenzie and Soukoreff's phrase set are from 1% of the total words from the COCA corpus (an English corpus that is recent to 2012: http://corpus.byu.edu/coca/).

The EMILLE/CIIL corpus [20] consists of 2387 Hindi files, totaling 737528 sentences and 12295677 words (202042 unique). Overall, all three phrase sets have a high word-based correlation with the EMILLE/CIIL corpus (0.70, 0.68 and 0.75 of the words in FP, TP and MSP, respectively, are from 1% of the total words from the corpus). This shows that the words used to build the phrase set are very frequently used words. The three phrase sets were similar in terms of single-letter correlation with the EMILLE/CIIL corpus (FP: 0.97, TP: 0.98, MSP: 0.98). Even the number of letters forming a word in each phrase set was not significantly different: FP: 3.46, TP: 3.63, MSP: 3.73. However, the phrase sets differed in terms of the number of words forming a phrase: FP: 8.16, TP: 13.46, MSP: 5.87 (Table 1).

An important characteristic of written text is readability, the ease with which text can be read. The lower the readability score of a phrase, the easier it is to read that phrase. The US Government Department of Defense uses Flesch-Kincaid Grade Level formula, which relies on the fact that the difficulty of a word is directly proportional to the number of syllables. Past Hindi research [1] has used this measure to evaluate readability. The readability attribute takes the effect of conjuncts into account. To

identify the number of syllables in a Hindi word, we used the syllabification algorithm proposed by [2] which has been reported to be more than 96% accurate. Our phrase sets readability level varies: FP: 5.36, TP: 8.0, MSP: 5.68. Phrases from grade VII textbook are found to be of grade 8 (sd=3.82), showing that the readability measure being used is efficient (Table 1).

Hypothesis.

We hypothesize that (**H1**) use of MSP and FP will result in faster text entry and a lower error rate than TP. This is based on the fact that MSP and FP have lower readability, higher word correlation, and lower words per phrase (Table 1) compared to TP. Additionally, MSP's high word-based correlation to the EMILLE/CIIL corpus shows that it is very similar to what people read and write on a day-to-day basis; hence, we hypothesize that (**H2**) participants would prefer MSP over FP and TP.

3   Evaluation Method: Participants, Apparatus and Procedure

We recruited 18 individuals (mean age=21.8, 12 males, 4 females) to participate in a user study comparing the phrase sets. All were undergraduate Computer Science students, with an average 10.16 years of QWERTY keyboard usage. The criteria for selecting participants for the study was that they must know how to read, write, and speak in Hindi, but have never used an Inscript (Indian Script) keyboard [19] before. The mother tongue of all the participants, except two (who also know Gujarati), was Hindi. All participants, except one, received their primary education in English. Each participant was paid Rs 100 (~$2) per session. To motivate the participants, prize money of Rs 1000 and Rs 500 was given to the fastest and second fastest participants respectively.

   We used the standard Inscript keyboard layout [19], attached to a 15.4 inches laptop screen. We developed custom software in C# which presented a test phrase at the top of the screen and asked the user to type the same phrase into a text box below it. The software logs all keystrokes for later analysis.

   A within-subject multiple session study was conducted. Each participant took part in three 45-minute sessions. In each session, the participant was presented with phrases from one of the phrase set, and was asked to enter the visible text as quickly and as accurately as possible. A session consisted of two 20-minute typing blocks with a break of 3-5 minutes between the blocks. A minimum gap of 2 hours and a maximum gap of 2 days was maintained between two consecutive sessions. We counter-balanced the ordering of the phrase sets among the 18 participants, to allow learning to equally affect all conditions. Phrases within a set were always in the same order. After each session, participants were required to complete a questionnaire, rating the phrase set in terms of memorability, understandability, phrase length, and frequency of usage (phrase set measures were taken from previous research [10], [21]), on a 5-point Likert scale. For instance, for understandability 1 was very difficult to understand and 5 was very easy to understand.

4 Study Results

All participants completed the 3 session study. In total, together they entered 1340 phrases (388 FP, 310 TP and 642 MSP).

4.1 Speed and Accuracy

Text entry speed is measured in terms of words per minute (wpm), calculated using (letters per second)*60/5, with the definition that a word consists of 5 letters [18]. This definition was used to make the results more generalizable with the previous studies, and also to provide context for a larger audience to be able to relate the obtained results with English text entry speed; this has been done for Korean text input as well [4]. The average text entry rate over all the sessions across all the participants was 6.91 wpm (sd=2.43) (FP: m=6.22, sd=2.16; TP: m=7.28, sd=2.62; MSP: m=7.22, sd=2.48). Contrary to our hypothesis *H1*, no significant difference ($F_{2,34}$=2.5, *p*=0.1) was found between the three phrase sets.

   Two metrics were used to measure text entry accuracy: (a) Keystrokes per Letter (KSPL) [18] – the number of keystrokes required to input a letter in Hindi, (b) Minimum String Distance (MSD) [18] between the presented and transcribed phrase. MSD accounts for the uncorrected errors in the final transcribed text; while KSPL measures the corrected errors as every correction adds multiple keystrokes, i.e., delete letter, re-enter letter. Note that for English, ideal KSPL value is 1 [18]. For Hindi, several letters require multiple keystrokes; we calculated and found the ideal KSPL for Inscript keyboard to be 1.12. For the entire study, a low average KSPL of 1.41 (sd=0.16) was found (FP: m=1.41, sd=0.13; TP: m=1.40, sd=0.1; MSP: m=1.43, sd=0.22). The average MSD value was 0.035 (sd=0.02) (FP: m=0.028, sd=0.01; TP: m=0.046, sd=0.03; MSP: m=0.03, sd=0.01). The higher MSD value for TP could be because the textbook phrases are not only longer than FP and MSP, but also have a higher readability value. However, no significant accuracy difference (KSPL: $F_{2,34}$=1.3, *p*=0.3, MSD: $F_{2,34}$=2.4, *p*=0.1) was found between the phrase sets, disproving our hypothesis *H1*.

4.2 Post-Session Questionnaire

Because MSP preserves many of the positive qualities found in MacKenzie and Soukoreff's English phrase set (i.e., short, easy to understand and memorable phrases), participants preferred it the most (Friedman χ^2(2)=14.7, *p*<0.01) (*H2*). The preference is derived cumulatively from post-questionnaire rating of phrases in terms of memorability, understandability, phrase length, and frequency of usage (FP: m=3.87, sd=0.1; TP: m=2.78, sd=0.1; MSP: m=4.41, sd=0.08). However, we did not find any significant difference between FP and MSP in terms of participants rating for understandability and memorability.

Understandability. Although participants perceived that the words used in the phrases were easy to understand across all the phrase sets, participants strongly agreed that FP and MSP contained phrases that are very easy to understand, while being neu-

tral towards TP. Fisher's Exact test showed significant differences for understandability between FP (m=4.6, sd=0.8) and TP (m=3.1, sd=0.9) with $p<0.0001$, and between TP and MSP (m=4.17, sd=1) with $p=0.01$. This could be because TP has the smallest word-based correlation of the three phrase sets, but the highest readability level (where a high readability score means that the text is hard to read).

Length. Whereas participants agreed that phrases in TP were too long to type, they were neutral towards FP and somewhat disagreed that phrases from MSP were too long. Thus, around 8 words per phrase seems an acceptable length for phrases, but this might be influenced due to high understandability and memorability of FP.

We found a strong correlation between frequency of phrases and phrase length, with Pearson's r(52) = 0.57, $p<0.0001$. Participants thought that MSP contained very frequently used phrases but somewhat disagreed that FP and TP contained very frequently used phrases. Thus, seemingly longer phrases may not represent what participants would associate as frequently used phrases.

Memorability. Recently, Vertanen and Kristensson [21] argue that phrases should be memorable to help participants type them after only reading once. Participants strongly agreed that FP and MSP contained phrases that are very easy to memorize, while being neutral towards TP. We found that participants were able to quickly recognize the movies the phrases from FP were from and enjoyed typing them: *"Phrases should be interesting, so that we enjoy typing."* Many would often read the phrases aloud and laugh while typing. Fisher's Exact test showed significant differences for memorability between FP (m=4.2, sd=0.2) and TP (m=2.5, sd=0.1), and between TP and MSP (m=3.9, sd=0.2) with $p<0.0001$. Also a strong correlation was found between memorability and understandability, with Pearson's r(52)=0.64, $p<0.0001$.

5   Conclusion

In this paper, we develop and evaluate three types of phrase sets for Hindi text entry: phrases taken from Hindi films, a grade VII Hindi textbook, and a Hindi translated version of the MacKenzie and Soukoreff's phrase set. Although these phrase sets have different characteristics (Table 1), we did not find any text entry performance difference resulting amongst them. Our results suggest the use of different types of phrases in previous Hindi text entry studies potentially did not significantly affect performance results. However, it has yet to be shown that there is perhaps no significant difference between any sets of phrases. Thus, use of these common phrase sets can help in attaining more generalizable results across future research papers and hence facilitate comparison of different text entry techniques. Evaluating additional types of phrase sets (*e.g.*, [13]) and different characteristics of the phrase sets is important future work.

Additionally, we learned from participant feedback that MSP was preferred, and the acceptable Hindi phrase length should be ~8 words/phrase, but ~13 is too long. Readability contributed to the acceptability of FP, as does memorability. Familiar

phrases help motivate the participants and make the whole typing task fun; this can be important in studies which ask participants to complete a large number of typing sessions. Thus, factors like readability, memorability, and phrase length, should all be taken into consideration together while choosing a phrase set.

References

1. Agnihotri, R.K., Khanna, A.L.: Evaluating the Readability of School Textbooks: An Indian Study. In: Journal of Reading, vol. 35, pp. 282--288, (1991).
2. Choudhury, M., Basu, A.: A Rule Based Schwa Deletion Algorithm for Hindi. In: Proceedings of KBCS, pp. 343--353. Vikas Publications (2003).
3. Dey, P., et al.: Voice Key Board: Multimodal Indic Text Input. In: Proceedings of ICML-MLMI, pp. 313--318. ACM Press (2009).
4. Ilinkin, I., Kim, S.: Evaluation of Text Entry Methods for Korean Mobile Phones, a User Study. In: Proceedings of CHI, pp. 2023--2026. ACM Press (2010).
5. Joshi, A., et al.: Design and Evaluation of Devanagari Virtual Keyboards for Touch Screen Mobile Phones. In: Proceedings of MobileHCI, pp. 323--332. ACM Press (2011).
6. Joshi, A., et al.: Keylekh: A Keyboard for Text Entry in Indic Scripts. In: Extended Abstracts of CHI, pp. 928--942. ACM Press (2004).
7. Jung, Y., Joshi, D., Narayana-Saroja, V., Desai, D.P.: Solving the Great Indian Text Input Puzzle: Touch Screen-based Mobile Text Input Design. In: Proceedings of MobileHCI, pp. 313--322. ACM Press (2011).
8. Krishna, A.: Gesture Keyboard – User centered Design of a Unique Input Device for Indic Scripts. HCI International (2005).
9. Kumar, A., Shah, H., Joshi, A.: Saral: Devanagari Text Input System in Mobile Phones. In: Proceedings of UbiComp, pp. 1--4. ACM Press (2009).
10. MacKenzie, I.S., Soukoreff, R.W.: Phrase Sets for Evaluating Text Entry Techniques. In: Proceedings of CHI, pp. 754--755. ACM Press (2003).
11. Mayzner, M.S., Tresselt, M.E.: Table of single-letter and digram frequency counts for various word-length and letter-position combinations. In: Psychonomic Monograph Supplements, vol. 1, pp. 13--32 (1965).
12. NCERT Books, http://ncertbooks.prashanthellina.com
13. Paek, T., Hsu, P.: Sampling representative phrase sets for text entry experiments: A procedure and public resource. In: Proceedings of CHI, pp. 2477--2480. ACM Press (2011).
14. Panini Keypad, http://www.paninikeypad.com
15. Rathod, A., Joshi, A.: A Dynamic Text Input scheme for phonetic scripts like Devanagari. In: Proceedings of Development by Design. ThinkCycle Press (2002).
16. Read, J.C.: Deceit, Distortion and Decision – Choosing Phrase Sets for Text Entry Research, http://www.chici.org/references/deceit_distortion_and_decision.pdf
17. Shanbhag, S., et al.: An Intelligent Multi-layered Input Scheme for Phonetic Scripts. In: Proceedings of SmartGraph, pp. 35--38. ACM Press (2002).
18. Soukoreff, R.W., MacKenzie, I.S.: Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In: Proceedings of CHI, pp. 113--120. ACM Press (2003).
19. TDIL Inscript Layout, http://tdil.mit.gov.in/pdf/keyboard/Keyboard%20Standard.pdf
20. The EMILLE/CIIL Corpus, http://catalog.elra.info/product_info.php?products_id=696
21. Vertanen, K., Kristensson, P.O.: A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In: Proceedings of MobileHCI, pp. 295--298. ACM Press (2011).