VISUAL WORD RECOGNITION IN HINDI-URDU: A MATTER OF ORTHOGRAPHY

A dissertation

submitted to the faculty of

National Brain Research Centre

in partial fulfilment of the requirements for

M.Sc. in Neuroscience

by

Azman Akhter

National Brain Research Centre
Manesar, Gurugram, Haryana
2017 – 2019

राष्ट्रीय मस्तिष्क अनुसन्धान केंद्र
(समविश्वविद्यालय)
(जैव एवं प्रोद्योगिक विभाग का स्वायत संस्था,
विज्ञानं एवं तकनीकी मंत्रालय, भारत सरकार)
एन० एच०-8, नैनवाल मोड़, मानेसर- 122051, (हरियाणा)

National Brain Research Centre
(Deemed University)
(An Autonomous Institute of Deptt. of Biotechnology,
Ministry of Science and Technology, Govt. of India)
NH-8, Nainwal Mode, Manesar-122051(Haryana)

CERTIFICATE

This is to certify that the dissertation entitled "Visual Word Recognition in Hindi-Urdu: A matter of orthography" is the result of work carried out by Azman Akhter in the Division of Computational and Cognitive Neuroscience, National Brain Research Centre, Manesar, Haryana, India.

The work presented herein is original and has not been submitted previously for the award of any degree or diploma to **National Brain Research Centre (Deemed University)** or to any other University. This work is completely based on the guidelines given by the **National Brain Research Centre (Deemed University)** and is a record of the candidate's own efforts.

Dr Dipanjan Roy
Supervisor

Prof. Neeraj Jain
Director, NBRC

Place: Manesar

Date : 3rd June, 2019

दूरभाष / Telephone: +91-124-2845 200 फैक्स / Fax : +91-124-2338910
ईमेल / Email : info@nbrc.ac.in Website : www.nbrc.ac.in

DECLARATION BY THE CANDIDATE

I Azman Akhter hereby declare that the work presented in this dissertation is carried out by me, under the guidance of Dr. Dipanjan Roy, National Brain Research Centre, Manesar, Haryana.

I also declare that no part of this dissertation has been previously submitted for the award of any degree or diploma at the National Brain Research Centre or any other University.

Azman Akhter

Place: NBRC, Manesar

Date: 3rd June 2019

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor Dr. Dipanjan Roy for continuous support, dedication and extreme patience throughout the dissertation work. Without his guidance, this work would not have been accomplished.

I am also extremely thankful to Dr. Arpan Banerjee, who along with Dr. Dipanjan, guided me throughout the project. His inputs have been proven to be extremely important for the completion of the project.

I would like to thank Masood Ahmad Wani and Mohd Shan who helped me with stimulus development and Mohd Aquib because of whom data collection became possible at Aligarh Muslim University.

I would also like to thank Dr.Dipanjan Ray, G. Vinodh Kumar, Priyanka Ghosh, Anagh Pathak, Neeraj Kumar for helping me in developing and running the experiments.

Also, thanks to Kirti Saluja, who provided time to time support and suggestions at times of need.

Special thanks to my parents and brothers for being understanding and supportive along the process and lows of time.

Azman Akhter

Abstract

Author: Azman Akhter
Title: Visual Word Recognition in Hindi-Urdu: a matter of orthography
Institution: National Brain Research Centre
Dissertation Advisor: Dr. Dipanjan Roy
Degree: Masters in Neuroscience
Year: 2019

Hindi and Urdu, the twin languages, jointly known as Hindustani. These languages are very similar on the spoken level but differ greatly on the written level. Differences are mainly because of orthographic depth, Hindi script is very shallow orthography while Urdu is of great depth, making it more complex and inconsistent on grapheme to phoneme mapping. Jointly these two are world's 3rd most spoken languages, but there very few studies that looked into how the brain processes these two contrastingly different orthographies of same spoken language and what is the associated brain network. The current study used a semantic lexical decision task to study the impact of orthographic differences in visual word recognition of highly frequent words, to understand how differently these two are read and to check how well or unwell the accepted reading models fit the reading in these scripts by early simultaneous Hindi-Urdu biliterates. For which, biliterates visually recognized words based on their semantic relation with some category of words. Same words were taken for both the scripts at two levels of length: one and two syllables. It was hypothesized that the recognition would be faster for transparent Hindi than opaque Urdu and would be faster for one syllable words than two syllable words for the scripts as suggested by previous studies of reading aloud. We found that the recognition was faster for Hindi than Urdu, as expected owing to the depth of orthography. But against the previous evidence and hypothesis, we found no significant difference in recognition of one syllable and that of two syllable words in both the scripts. These results suggest that the reading in Urdu, in contrast to Hindi, access different pathway or mechanism owing to its visual complexity. Also, the results from the factor of word length, suggests that skilled reading in Hindi, like in Urdu, is not in accordance with the phonological assembly theory as suggested by the previous study.

CONTENTS

1. Introduction

Reading is an ability particular to the human brain. It is an acquired skill, acquired by learning and practice, unlike spoken language which is biologically instinctive.

The naturalistic skilled reading process has two components: Visual Word Recognition which to identify and computing the meaning of the printed letter string, and Comprehension which is to make sense of an array of words together.

Visual word recognition needs decoding the orthographic code of the words to access their meaning. Many models have been proposed till date to explain how this decoding happens in the brain from connectionist triangle model to non-connectionist Dual Route Cascade (DRC) Model.

One of the most accepted is non-connectionist, DRC model, which based on the Dual Route Hypothesis of reading (Coltheart et al. 2001; Perry et al. 2007). According to this model, there are two routes to access the meaning of the printed word: a direct route which bypasses phonological recoding and the other is the indirect one, which first recodes phonological information and then accesses meaning. Which route has to be used is influenced by the depth of orthography or consistency of grapheme-phoneme mapping as proposed in Orthographic Depth Hypothesis (ODH) (Frost et al. 1987). DRC is primarily a computational realization of the dual-route theory of reading and is the only model of reading that can perform the 2 tasks most commonly used to study reading: lexical decision and reading aloud.

In contrast to non-connectionist DRC model, connectionist Triangle model is another famous model of the reading network(Rayner and Reichle 2010; Joshi and McCardle 2018). According to this model, orthography-phonology mapping develops according to the frequency of exposure spelling-sound correspondence. Pronunciation of word is mediated by semantics, Words with consistent grapheme-phoneme mapping can be pronounced with semantics and inconsistent word’s pronunciation is mediated by semantic content(Seidenberg and McClelland 1989a; Harm and Seidenberg 1999a, 2004a).

If a word is in a script which is orthographically shallow like Spanish and Italian i.e. there is consistency in grapheme-phoneme mapping, then the recognition and naming of the word will be faster than when the script is with deep orthography like English, Arabic, Hebrew. There has been extensive research addressing the question of the brain's reading network and role of orthography. But most of them have attempted this with two different languages in bilingual conditions such as Arabic and Hebrew, Spanish and English, Finnish and French(Ibrahim et al. 2013; Oliver et al. 2017). Studies with vowelized as shallow orthography and non-vowelized Arabic as deep orthography(Taha and Azaizah-Seh 2017) can be misleading as well because vowelized Arabic has supplementary visual features that are not unlike the majority of the text, making it unfamiliar.

Given that Spanish differs from English or Arabic differ from Hebrew not just in orthography but also in morphophonology other linguistic properties like grammar. Also, most of them have used lexical decision tasks involving word and pseudo-and non-words to gain access to different parts of the reading network. Pseudo-and non-words do not represent natural reading since there is no representation of these alien letter strings in the mental lexicon of words. Such tasks are useful and can be errorless with the studies of reading acquisition and learning to read, and development of brain's reading network since during such processes there is exposure to new letter strings that are learned to associate with their meanings later. In contrast for skilled naturalistic reading, every word has to be recognized by mapping to the acquired representation of it in the semantic lexicon to access its meaning.

The 3rd most spoken language of the world according to very recent estimates("Hindi | Ethnologue" 2019) and language dominant in Northern Indian subcontinent is Hindustani, a group of two twin languages: Hindi and Urdu(Pulsipher et al. 2006). It also being argued and debated that these are not two different languages but essentially one language written in different orthographic forms (Shackle and Snell 1990; Jain and Cardona 2007). Mutually intelligibly spoken, they share a large set of vocabulary originated from both Sanskrit and Persian. This combination of language provides a unique window to examine the contribution of orthographic properties since they are practically similar in linguistic aspects: phonology, orthography, grammar, and lexicon. Hindi written in Devanagari script is highly transparent and very consistent in grapheme-phoneme mapping. On the other hand, Urdu written in Persian script is known to be of very deep orthography with high inconsistency in orthography-phonology coding. Thus, can be a great tool to advance the understanding role of orthography in the reading process and impact of orthography on development of language network of the brain.

Also, can be very helpful, to understand more accurately both the structural and functional aspects of the reading network.

There are very few studies which attempted to understand the mechanism of reading by the brain in relation to Hindustani and the role of orthography with the help of Hindi-Urdu. One of them examined the role of orthography with the help of Urdu-Hindi language in relation to the DRC model (Rao et al. 2011a) with a reading-aloud task, found slow naming in Urdu than Hindi. In other, divided visual hemifield study role of orthography and morphology has been examined in hemispheric lateralization in reading aloud process(Rao and Vaid 2017). They have found involvement of right hemisphere for naming in Urdu but not in Hindi. One neuroimaging study reported about higher activation in language areas of the brain for Urdu than for Hindi(Kumar 2014).

In this fMRI study by Kumar et al a perceptual level task (to identified italicized letters within words) was used and the behavioral studies by Rao et. al. employed reading aloud task. The limitation of the first approach is that it does not enforce the brain to read the word at all and in the reading aloud task, there is no enforcement on accessing the meaning of the word, so no account for semantic processing. Hense, both these approaches do not necessarily address the word identification during a natural skilled comprehensive reading.

So, to study structural and functional properties of the brain's reading network with the help of the Hindustani language group, there was a requirement to develop a task that can address the above-mentioned limitations.

Current study was aimed to address these limitations so as to develop a complete framework for further neuroimaging studies. This framework would be used along with sub-lexical and perceptual-level task in neuroimaging experiments to advance the understanding of reading and role of orthography in structural and functional properties of developing and skilled readers.

To achieve our aim, we designed a lexical decision task which mimics visual word recognition during natural skilled comprehensive reading. We designed a semantics dependent lexical decision task to examine visual word recognition as a function of the depth of orthography

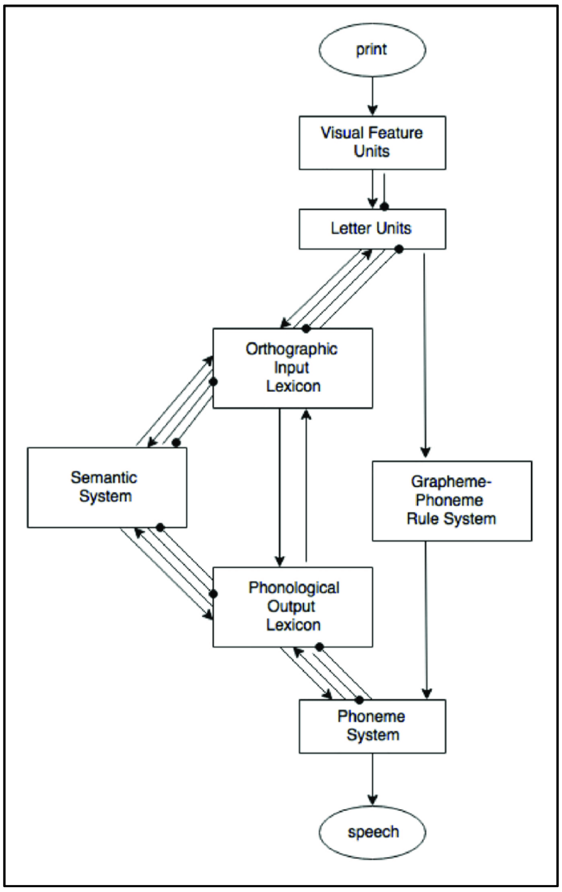
2. Models of reading

2.1 Dual Route Cascade model

According to this model, there are two routes to access the phonetic information of a word to read aloud: a direct route or lexical route which bypasses phonological recoding and the other is indirect or non-lexical route, which first recodes phonological information and then accesses meaning(Coltheart et al. 2001; Perry et al. 2007). Which route has to be used is influenced by the depth of orthography or consistency of grapheme-phoneme mapping as proposed in Orthographic Depth Hypothesis (ODH) (Frost et al. 1987).

If the orthography is deep or there is inconsistency in grapheme-phoneme mapping then the lexical route would be chosen and orthographic and semantic lexicons will be accessed first before knowing the sound of the word. On the other hand, for shallow orthography, words with consistent letter to sound mapping, the non-lexical route will be chosen and the word will be read aloud without reaching its semantic representation.

Fig 1: A schematic of the DRC model(Coltheart et al. 2001)



2.2 Triangle model

This model has two fundamental assumptions: First, a word's pronunciation is generated by activation of processing inputs from units with orthographic representation along with connections to other units with phonological output. Second, lexical information is presented in a distributed manner, instead of located in discrete units this model assumes that lexical and semantic information is in the connections that mediate between orthographic input and phonological output(Seidenberg and McClelland 1989b; Plaut and Shallice 1993; Harm and Seidenberg 1999b, 2004b). Since the strength of these connections is influenced by repeated experience, this model predicts that frequent words would be pronounced faster than less frequent words. Likewise, connections being more consistent for consistent orthography than for inconsistent, consistent words would be pronounced faster than inconsistent.

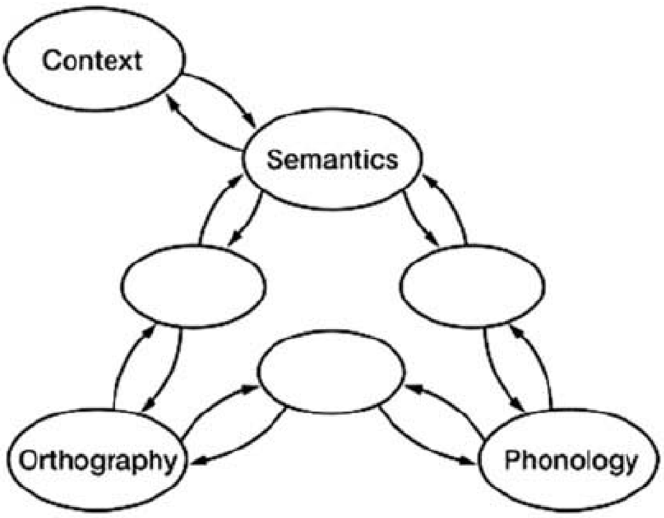


Fig 3: A schematic of the Connectionist Triangle Model of Reading(Martin-Chang and Levesque 2015)

3. Hindi-Urdu Orthographies

Although both Hindi and Urdu are identical in morphophonological and linguistic aspects, the written forms are completely different. They have different origin and are structurally very distinct. There are differences in reading/writing direction (Hindi is written from left to right and Urdu from right to left) and their orthographic transparency and complexities are also very distinct.

Hindi, an Abudiga writing system(Share and Daniels 2016), is written in Devanagari script. Devanagari is highly consistent and consistent in its grapheme to phoneme mapping. The primary orthographic unit in Hindi is *Akshara*, which corresponds to a consonant-vowel syllable. Consonants have graphemic representation with a full form and an abbreviated, or ligatured form used to represent consonant clusters as words(Vaid and Gupta n.d.; Rao et al. 2011b).

Urdu is written in right to left direction and is a Persio-Arabic script(Mirdehghan 2010; Mirdehghan and Moradkhani 2010). Urdu script holds higher complexity and less consistency in grapheme to phoneme mapping. Similar to the Arabic script, some sounds in Urdu have more than one letter representation. Prior knowledge of the word is necessary to pronounce it. Moreover, letters in Urdu take on different shapes depending on whether they occur in the beginning, middle or end of the word, the shape also depends on how the letter is joined with other letters within a word. Some letters can be joined in more ways than others, some must be written cursively and others not. Thus, a reader of Urdu has to recognize different forms of the letter. Also, visual complexity is added by omitted vowels, namely small vowels those which are represented by diacritics, in standard script. This omission adds up the ambiguity to the words, some are spelled same but spoken differently, causing many such words to have multiple potentials of it may sound and what it may mean. Contrastingly, in Hindi there is no such ambiguity, there is very clear one to one mapping of symbols and sound, and letters are discrete, not fused or amalgamated with other as in Urdu.

4. Methods

4.1 Participants

40 healthy and neurologically normal, right-handed native Hindustani speakers from participated in the experiment. All participants had normal or corrected to normal vision and had no history of reading or any other kind of neurological disorder. Participants are recruited from three centers: Jamia Milia Islamia, Aligarh Muslim University and National Brain Research Centre. The sample comprised of two groups of participants.

Group 1: Named here as Hindi-Urdu Biliterates, includes 20 Multilinguals, proficient in both Hindi and Urdu Reading. These were simultaneous early learners of Hindi and Urdu, started to learn to read and write before the age of 5.

Group 2: Named as Hindi Mono-literates, 20 proficient literates of Hindi and with no learning of Urdu reading/writing at any stage of life.

All participants acquired English as their second language(L2). And were highly proficient in their L1: Hindi-Urdu or Hindi and L2: English, with minimal exposure to regional languages of North India.

Language proficiency in the reading domain was measured by an indigenously developed reading test. They also filled in a language background questionnaire.

There were no differences between the two groups in terms of Age of Acquisition of Hindi/Urdu and Hindi, percent average daily exposure to Hindi, proficiency in reading scores and participants from both groups also self-rated themselves 4 higher on a scale of 1-5 in proficiency in reading Hindi. Within the Hindi-Urdu biliterates group, there was no difference in Age of learning reading/writing Urdu and Hindi, duration of study as a subject, self-rated, and objective proficiency in Urdu and Hindi and percent average daily exposure to Hindi and to Urdu script.

Table 1: Participants demographic details and linguistic characteristics by group for Hindi only.

	Hindi mono-literate (n = 20)	Hindi-Urdu biliterates (n= 20)	p-values
Age	23.3(2.9)	22.9(2.1)	0.37
Gender (% female)	25	60	
AOA	0	0	
Age of acquisition of reading	3.75(0.7)	4.1(1.2)	0.43
Duration of study	13.1(1.5)	13.7(3.2)	0.58
Proficiency (self-rated)	4.3(0.7)	4.6(0.5)	0.11
Proficiency (Rubric Fluency score (%)	92(7)	93(10)	0.49
Average Daily exposure with script (%)	18.4(17.7)	13.5(9.4)	0.35

Table 2: Linguistic characteristics of Hindi-Urdu Biliterates for both Hindi and Urdu script

	Hindi	Urdu	p values
AOA	0	0	
Age of acquisition of reading	4.1(1.2)	3.8(0.9)	0.42
Duration of study	13.7(3.2)	12.8(3.9)	0.48
Proficiency (self-rated)	4.6(0.5)	4.3(0.4)	0.52
Proficiency (Rubric Fluency score %)	93(10)	87(12)	0.09
Average Daily exposure with script (%)	13.5(9.4)	23(17.9)	0.05

4.2 Ethics Statement

Participants gave informed written consent before the experiment. The experimental procedure was approved by the ethical committee of the institute.

4.3 Material

Language Background Questionnaire: It was used to collect information about languages known, learned, in use and frequency of use. See **appendix 2** for the questionnaire.

Indigenous test of Reading proficiency: Reading fluency was assessed by using Fluency Rubric (Zutell and Rasinski 1991). This task assessed the fluency of participants to read while reading a written passage in each script. Each passage was followed by 4 questions to enforce the participants for a comprehensive reading. **Appendix 1** presents the scoring process.

4.3.1 Stimulus and Design

Total of 245 high-frequency Hindustani words was selected based on the subjected frequencies from separate 21 raters, exempted from the main experiment. Among these 245 words, 120 were one syllable and 125 were two syllables. Same stimuli were used in both the scripts, Hindi and Urdu. Majority of the words were nouns.

All the words were categorized among 7 categories according to their semantic relationship**.** Half of the words in each category do have semantic relation and belong to the home category and half of the words don't.

A Go/No-Go task is designed. Go for each trial with a word belonging to the category and No Go for each trial with a word not belonging to the category.

Fully balanced within-subject 2 X 2 factorial design manipulated (i) Script of the word: Hindi and Urdu (ii) Syllabic length: One and Two. This resulted in 4 sets of conditions:

One syllable word in Urdu	Two syllable word in Urdu
One syllable word in Hindi	Two syllable word in Hindi

Each set makes up a block of Script X Syllabic length. Blocks of one syllable words consisted of 120 trials each and that of two syllable words consisted of 125 trials each. Each category within a set forms a subset of trials within a block. Every block had 7 subsets of words belonging to each category of words.

The categories were: Animals, Body Parts, Action, Place, Food Item, Number, Non-Living Objects

4.4 Procedure

The participants carried out a semantic Go/No-Go task. To the Hindi-Urdu Biliterates, all four blocks and to Hindi mono-literates only 2 blocks that of Hindi script were presented. The order of blocks was counterbalanced across subjects.

In each set of stimuli, first, the category name was presented in the center of the screen. Participants were asked to press a key to enter into the category. Once entered, words from that category presented sequentially in the center of the screen with a short inter-trial interval of 300ms. Participants were instructed to make a lexical decision on the presented word, whether it belongs to the aforementioned category or not with Left arrow key for 'No-Go' and Right for 'Go'. Each word was presented for 1500ms or till the end response, whichever is first and participants were asked to respond to best of their ability of speed and accuracy within that duration when the word is on screen. Each word was preceded by a fixation cross for 300ms. With the end of all the words of a category, other category starts. Categories within each set and words within each category were randomized.

All stimulus and instructions were presented in white color on a black background. Hindi stimuli were presented in 'Mangal' font and 'Urdu' in Fajer Noori Nastaliq. All instructions were presented in the English language. All words were presented in the center of the screen.

Each set makes up a block of Script X Syllabic length. Blocks of one syllable words consisted of 120 trials each and that of two syllable words consisted of 125 trials each. Each category within a set forms a subset of trials within a block. Every block had 7 subsets belonging to each category of words.

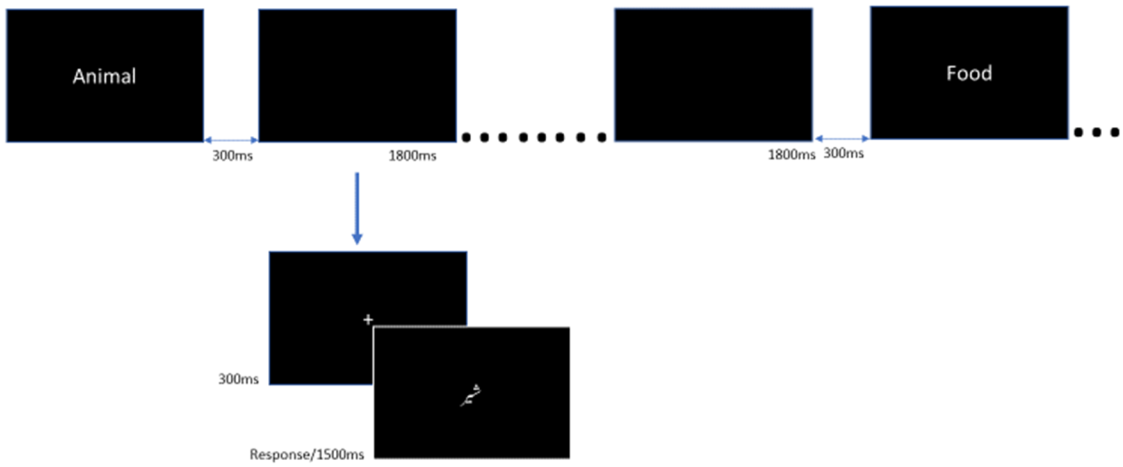


Figure 2: The Paradigm design. Each set starts with the name of category followed by trials of that category. After the last trial of each category, next category starts.

5. Result

Participants responded to the go and no-go trials during lexical decision task with more than 75% accuracy in each set of stimuli, indicating that they paid attention to the task. Both groups responded to the same trials. We performed two separate two-way ANOVAs, one for within-subject factors (script x syllabic length) within Hindi-Urdu biliterate group and other with Group as between-subject Factor (Hindi-Urdu and Hindi-only) and syllabic length as within-subject factor across both groups, with RTs and accuracy as dependent factors.

5.1 The reaction time of lexical decision for Hindi-Urdu biliterates:

A two-way repeated measure ANOVA was conducted on response latencies as a function of Script (Hindi and Urdu) and Syllabic length (one and two syllables). Outliers (RT<200ms) were eliminated before analysis.

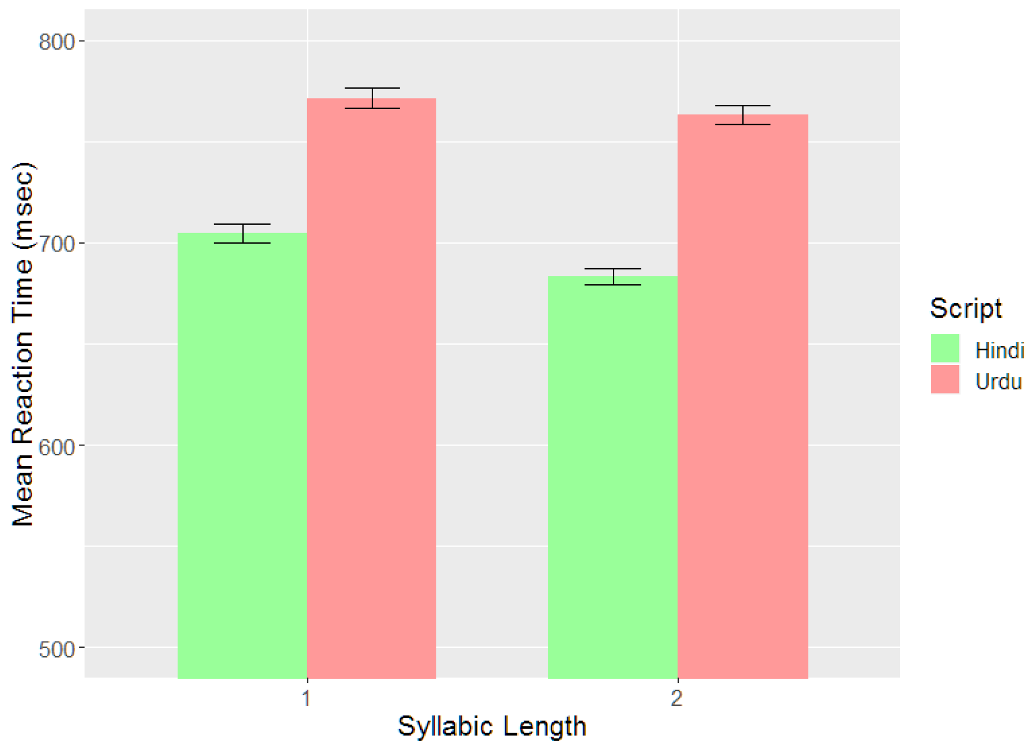
The analysis revealed the main effects of one variable: Script [$F(1,18)=13.55$, $p < 0.01$], but no significant effect of Syllabic length[$F(1,18)=0.81$, $p=0.38$]. That is, RT was shorter for Hindi than Urdu, and no overall significant difference was found in RTs for one syllable word from that of two syllables. We also didn't find any interaction of Script x Syllabic length [$F(1,18)=1.53$, $p=0.23$]. Post hoc analysis revealed that effect arose from both scripts in both conditions of word length: Hindi one-syllable words were significantly faster recognized than Urdu one syllable words [$p<0.05$] and similarly faster RT for Hindi two syllable words than Urdu two syllable words[$p<0.05$].

5.2 The reaction time of lexical decision for Hindi-Urdu biliterates vs Hindi-mono-literates with Hindi Script only:

The analysis revealed the no main effects of within-subject variable: Syllabic length [$F(1,37)=8.54$, $p>0.05$], also for between-subject variable which was group of participants: Hindi-Urdu biliterates and Hindi mono-literates [$F(1,37)=0.74$, $p=0.39$]. That means RT for word recognition in Hindi by Hindi mono-literates was not significantly different from that of by Hindi-Urdu mono-literates. Also, within each group, there was no difference in RT for both 1 syllable and two syllable words. So, overall Reaction time for word recognition in Hindi was similar by both the groups.

Fig 2: (A) Mean reaction time and (B) Accuracy for visual word recognition by Hindi-Urdu Biliterates.

(A)



(B)

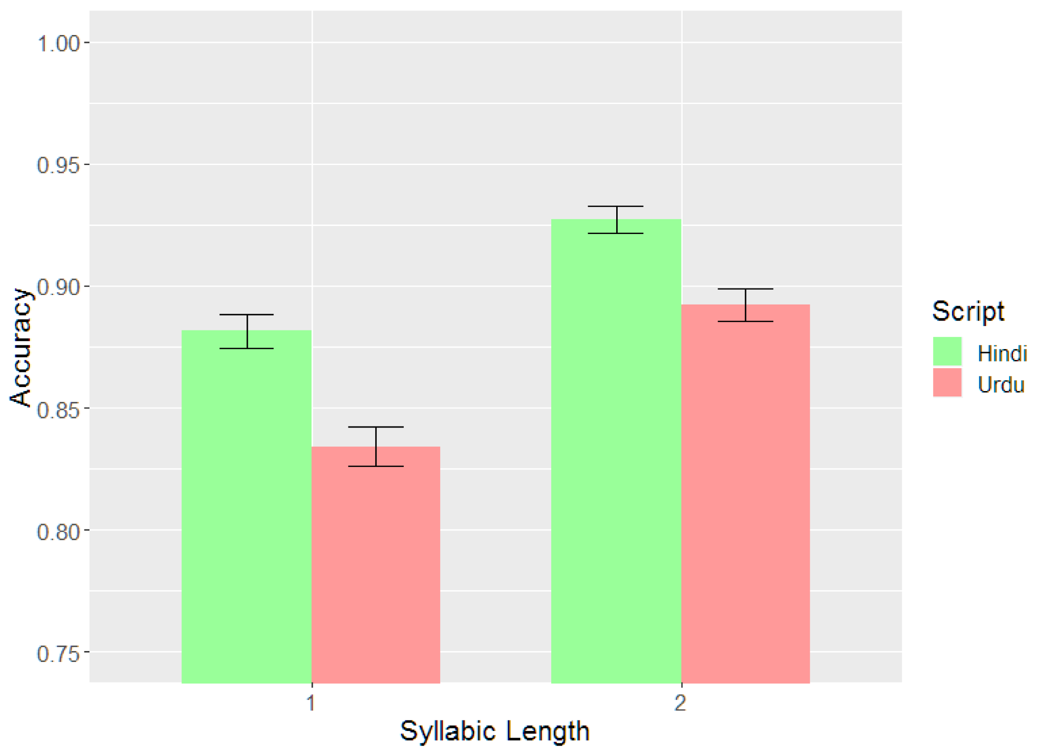
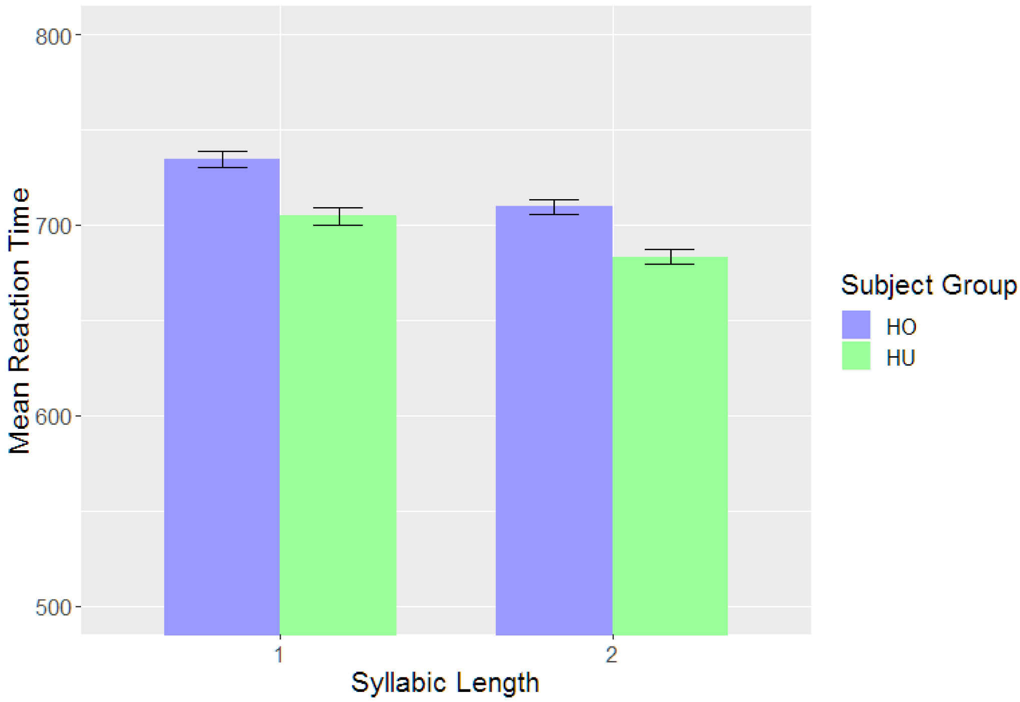
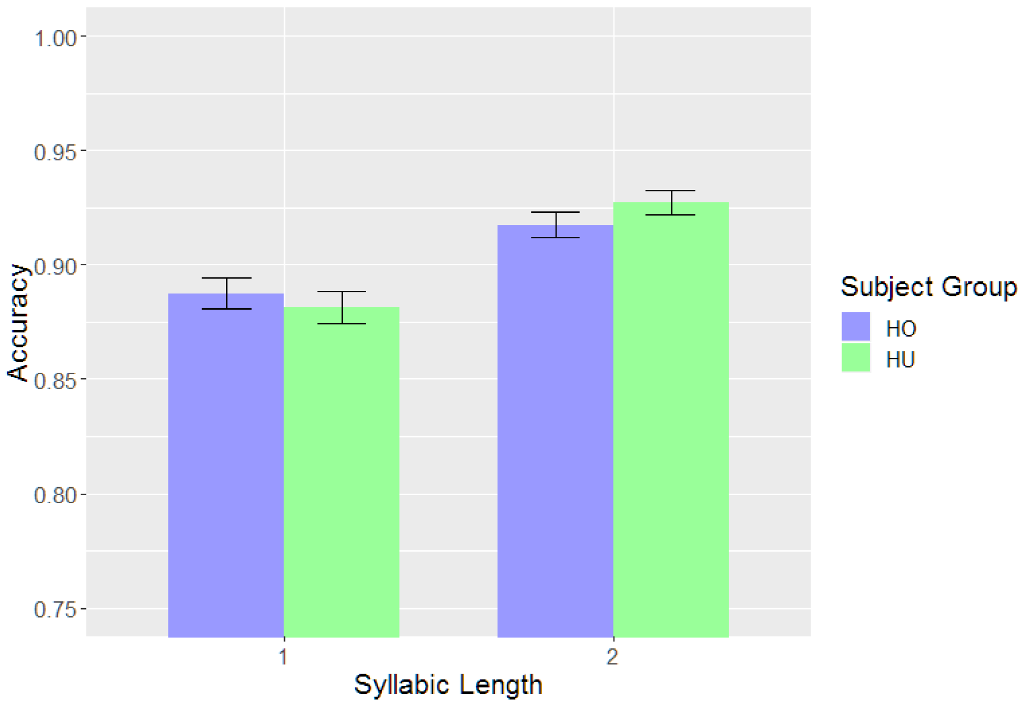


Fig 2: (A) Mean reaction time and (B) Accuracy for visual word recognition by Hindi-Urdu Biliterates (HU) and Hindi mono-literates (HO) in Hindi Only.

(A)



(B)



6. Discussion

Our study examined the impact of depth of orthography in visual word recognition by skilled simultaneous biliterate readers of languages Hindi and Urdu that share core morphophonological and grammatical properties but differ markedly in their written form, particularly in reading direction and orthography. The study accessed the response time and accuracy of responses while participants recognized words based on their semantic information.

As expected on the basis of differences of orthographies of two scripts, recognition was significantly faster for words in Hindi script than for words in Urdu script, for both two syllable and one syllable words. It is consistent with prior findings on non-lateralized and lateralized reading aloud tasks with these languages(Rao et al. 2011b; Rao and Vaid 2017). Also, the accuracy was higher for Hindi than Urdu. Confirming that reaction time results are not a consequence of reaction time-accuracy tradeoff.

The reaction time in recognition of one syllable words in both the scripts was not significantly different from the reaction time in recognition of two syllable words. This result is unexpectedly not in accordance with the previous study of this language set with reading aloud task(Rao and Vaid 2017), in which reading was faster for shorter one syllable words than for two syllable words. One possible explanation of this could be that words of both lengths have been read with whole word as a single grain as the Grain Size theory explains (Ziegler and Goswami 2005) and the higher latency in naming two-syllable words in the previous study with reading aloud task, may be only due to articulation of two phoneme as compared to one for one syllable words, to produce sound.

Word recognition in Hindi script across the groups: Hindi mono-literates and Hindi-Urdu biliterates, found to be insignificantly different in terms of reaction time and accuracy. Indicating that there is no impact on shallow orthography processing by learning a deep orthography simultaneously and may be in both groups the network employed for reading shallow orthography is identical. The take off for this is that the Hindi mono-literate group can be taken as a control group for neuroimaging studies when studying the impact of different orthographies in development in-and skilled reading by Hindi-Urdu biliterates.

The results of within-subject word recognition by biliterates of Hindi and Urdu indicates that these two orthographies are being processed via two different ways. DRC model can be thought to explain this two-way processing but the non-lexical route, as can be thought for words in Hindi script, assumes to have no interaction with semantic information. That's why for semantics dependent word recognition as the task demanded, even in shallow Hindi script, interaction with units of semantic representation is necessary. We suggest that the triangle model will be better to explain reading in Hindi and Urdu, as it assumes interaction across phonology, orthography and semantic units.

Also, no differences in responses in recognition of one syllable and two syllable words indicate that there is no letter assembly to phoneme recoding happening but might be first the orthographic lexicon being accessed for the whole word then its output goes to phonological and semantic units.

Though there were a certain limitation, one of which was that this experiment couldn't account for orthographic familiarity because of lack of standardized corpus for both languages. There are evidences of the influence of orthographic familiarity on the lexical decision(Proverbio and Adorni 2008). Further study with controlled orthographic familiarity and with more variable length of words is needful to find the exact influence of word's length on word recognition to be certain about the grain size in both the scripts.

We conclude that our study disagrees on the prior suggestions by Rao et al that Hindi and Urdu reading is in accordance with the DRC model. Also, this study provides a considerable experimental design for further studies of functional network dynamics and structural dynamics in regards to the development of biliterates of Hindi and Urdu.

7. References

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychol Rev. 108:204–256.

Frost R, Katz L, Bentin S. 1987. Strategies for visual word recognition and orthographical depth: A multilingual comparison. J Exp Psychol Hum Percept Perform. 13:104–115.

Harm MW, Seidenberg MS. 1999a. Phonology, reading acquisition, and dyslexia: insights from connectionist models. Psychol Rev. 106:491–528.

Harm MW, Seidenberg MS. 1999b. Phonology, reading acquisition, and dyslexia: insights from connectionist models. Psychol Rev. 106:491–528.

Harm MW, Seidenberg MS. 2004a. Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. Psychol Rev. 111:662–720.

Harm MW, Seidenberg MS. 2004b. Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. Psychol Rev. 111:662–720.

Hindi | Ethnologue [WWW Document]. 2019. . URL https://www.ethnologue.com/language/hin

Ibrahim R, Khateb A, Taha H, Characteristics HO. 2013. How Does Type of Orthography Affect Reading in Arabic and Hebrew as First and Second Languages ? 3:40–46.

Jain D, Cardona G. 2007. The Indo-Aryan languages. Routledge.

Joshi RM, McCardle P. 2018. Models of Reading in Different Orthographies: An Introduction. J Learn Disabil. 51:419–421.

Kumar U. 2014. Effect of orthography over neural regions in bilinguals: A view from neuroimaging. Neurosci Lett. 580:94–99.

Martin-Chang S, Levesque K. 2015. Reading Words In and Out of Connected Text: The Impact of Context on Semantic and Orthographic Processing. Sci Stud Read.

19:392–408.

Mirdehghan M. 2010. Persian, Urdu, and Pashto: A comparative orthographic analysis. Writ Syst Res. 2:9–23.

Mirdehghan M, Moradkhani S. 2010. Personal Pronouns in the Kakavandi Laki Dialect of Harsin (Kermanshah, Iran). Iran Stud. 43:513–531.

Oliver M, Carreiras M, Paz-Alonso PM. 2017. Functional dynamics of dorsal and ventral reading networks in bilinguals. Cereb Cortex. 27:5431–5443.

Perry C, Ziegler JC, Zorzi M. 2007. Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. Psychol Rev. 114:273–315.

Plaut DC, Shallice T. 1993. Deep dyslexia: A case study of connectionist neuropsychology. Cogn Neuropsychol. 10:377–500.

Proverbio AM, Adorni R. 2008. Behavioral and Brain Functions Orthographic familiarity , phonological legality and number of orthographic neighbours affect the onset of ERP lexical effects. 13:1–13.

Pulsipher LM (Lydia M, Pulsipher A, Hapke HM. 2006. World regional geography : global patterns, local lives. W.H. Freeman.

Rao C, Vaid J. 2017. Morphology, orthography, and the two hemispheres: A divided visual field study with Hindi/Urdu biliterates. Neuropsychologia. 98:46–55.

Rao C, Vaid J, Srinivasan N, Chen H-C. 2011a. Orthographic characteristics speed Hindi word naming but slow Urdu naming: evidence from Hindi/Urdu biliterates. Read Writ. 24:679–695.

Rao C, Vaid J, Srinivasan N, Chen HC. 2011b. Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. Read Writ. 24:679–695.

Rayner K, Reichle ED. 2010. Models of the reading process. Wiley Interdiscip Rev Cogn Sci. 1:787–799.

Seidenberg MS, McClelland JL. 1989a. A distributed, developmental model of word recognition and naming. Psychol Rev. 96:523–568.

Seidenberg MS, McClelland JL. 1989b. A distributed, developmental model of word recognition and naming. Psychol Rev. 96:523–568.

Shackle C, Snell R. 1990. Hindi and Urdu since 1800 : a common reader. Heritage Publishers.

Share DL, Daniels PT. 2016. Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). Writ Syst Res. 8:17–31.

Taha H, Azaizah-Seh H. 2017. Visual word recognition and vowelization in Arabic: new evidence from lexical decision task performances. Cogn Process. 18:521–527.

Vaid J, Gupta A. n.d. Exploring word recognition in a semi-alphabetic script: the case of Devanagari. Brain Lang. 81:679–690.

Ziegler JC, Goswami U. 2005. Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. Psychol Bull. 131:3–29.

Zutell J, Rasinski T V. 1991. Training teachers to attend to their students' oral reading fluency. Theory Pract. 30:211–217.

Appendix 1:

Rubric for fluency measure: For reading fluency

Scores	Smoothness/pacing	Confidence	Accuracy	expression
4	Reader reads all of the familiar text smoothly and continuously. The reader pays attention to punctuation marks and understands how to break text up into meaning groups of words.	Reader appears relaxed/confident and recovers quickly if a mistake is made.	Reader self corrects, or does not make errors when reading the familiar text.	Reader reads familiar test with appropriate changes in voice pitch/expression that reflect comprehension of the text and add dramatic emphasis to the text.
3	Reader reads most of the familiar text smoothly and pays some attention to punctuation marks.	Reader appears relaxed/confident, but is slightly agitated/confused by mistakes.	Reader makes occasional error that do not affect the content of the text (e.g. mispronouncing character names).	Reader reads familiar text with appropriate changes in voice pitch/expression that reflect comprehension of the text.
2	Reader reads familiar text either too quickly or with awkward pauses.	Reader appears somewhat nervous and is confused/agitated by mistakes.	Reader makes occasional errors that affect the content of the text.	Reader reads familiar text with changes in voice pitch/expression that may not match the text meaning.
1	Reader reads familiar text with long extended pauses or by slowing sounding out each word.	Reader appears nervous and cannot concentrate to read.	Reader makes frequent errors when reading familiar text and text appears to be above students comfortable reading level.	Reader reads familiar test in a monotone voice.

Appendix – 2

Demographic and Linguistic characteristics

(To be filled with details of every language known)

		Native Language	2nd Language	3rd Language	4th Language
Language					
Age of Acquisition					
Age of learning at school					
Age of Acquisition of Reading and Writing Skills					
Age at which became fluent in Reading and Writing					
Highest level of formal study and duration (years)					
Daily Exposure (hr:min)	Phonology				
	Orthography				
Proficiency rating on scale of 1-5					
Listening					
Speaking					
Reading					
Writing					

*Age is in years

Appendix 3(a)

Urdu paragraph for reading proficiency test:

ان عورتوں کے لیے جو علاقہ منتخب کیا گیا ہے ، وہ شہر سے چھ کوس دور تھا اور اس کے آگے کوس بھر کچّا راستہ تھا. کسی زمانے میں وہاں کوئی بستی ہوگی مگر اب تو کھنڈروں کے سوا کچھ نہ رہا تھا. جس میں سانپوں اور چمگاڈروں کے مسکن تھے اور دن دھاڑے الو بولتے تھے. اس علاقے کے نواح میں کچھ گھروندوں والے چھوٹے چھوٹے گاؤں تھے. مگر کسی کا فاصلہ بھی یہاں سے دو دہائی میل سے کم نہ تھا. ان گاؤں کے بسنے والے کسان دن کے وقت کھیتی باڑی کرتے یا پھرتے پھراتے ادھر نکل آتے. ورنہ ام طور پر اس شہر خموشاں میں آدم زاد کی صراط نظر نہ آتی تھی. بعض اوقات روز روشن ہی میں گیدڑ اس علاقے مے پھرتے دیکھ گے تھے

Appendix 3(b)

Hindi paragraph for reading proficiency test:

वैसे तो मेरे मामा के गाँव का होने के कारण मुझे बदलू को 'बदलू मामा' कहना चाहिए था परंतु मैं उसे 'बदलू मामा' न कहकर बदलू काका कहा करता था जैसा कि गाँव के सभी बच्चे उसे कहा करते थे। बदलू का मकान कुछ ऊँचे पर बना था। मकान के सामने बड़ा-सा सहन था जिसमें एक पुराना नीम का वृक्ष लगा था। उसी के नीचे बैठकर बदलू अपना काम किया करता था। बगल में भठी दहकती रहती जिसमें वह लाख पिघलाया करता। सामने एक लकड़ी की चौखट पड़ी रहती जिस पर लाख के मुलायम होने पर वह उसे सलाख के समान पतला करके चूड़ी का आकार देता। पास में चार-छह विभिन्न आकार की बेलननुमा मुंगेरियाँ रखी रहतीं जो आगे से कुछ पतली और पीछे से मोटी होतीं। लाख की चूड़ी का आकार देकर वह उन्हें मुँगेरियों पर चढ़ाकर गोल और चिकना बनाता और तब एक-एक कर पूरे हाथ की चूड़ियाँ बना चुकने के पश्चात वह उन पर रंग करता।