

# Design and Development of an Online Computational Framework to Facilitate Language Comprehension Research on Indian Languages

Manjira Sinha, Tirthankar Dasgupta, Anupam Basu

Indian Institute of Technology Kharagpur

Kharagpur, India

{manjira87, iamtirthankar, anupambas}@gmail.com

## Abstract

In this paper we have developed an open-source online computational framework that can be used by different research groups to conduct reading researches on Indian language texts. The framework can be used to develop a large annotated Indian language text comprehension data from different user based experiments. The novelty in this framework lies in the fact that it brings different empirical data-collection techniques for text comprehension under one roof. The framework has been customized specifically to address language particularities for Indian languages. It will also offer many types of automatic analysis on the data at different levels such as full text, sentence and word level. To address the subjectivity of text difficulty perception, the framework allows to capture user background against multiple factors. The assimilated data can be automatically cross referenced against varying strata of readers.

**Keywords:** language comprehension, text readability, Indian languages, empirical data collection, online computational framework

## 1. Introduction

Language comprehension or text comprehension generally refers to the ability to understand a particular piece of text by a specific group of people (DuBay, 2004). Comprehending a text is a complex cognitive phenomenon. The first step towards reading a piece of text is to understand the words or phrases in it. A word has a four dimensional measure of complexity: orthography, phonology, morphology and semantic. The next step is to comprehend the full sentence. This involves two things: the reader has to first process the syntactic structure of the sentence to understand the relation between different parts and second, to get the semantic meaning by integrating the semantics of the different components. Finally, the whole text is considered. Therefore, it can be stated that the cognitive load associated with reading a text broadly depends on five factors: **lexical complexity** that is the complexity of the different words or phrases used in the text; **syntactic complexity**: it depends on the syntax and structure of the sentences; **semantic complexity**: that is the difficulty faced due to unknown or non-familiar word or phrase structure; **discourse level complexity**: it depends on text properties like cohesion, coherence, rhetorical structure of text. Moreover, the background of the reader: this defines the cognitive ability and perception of the text difficulty by a target reader. We will use the terms *text comprehensibility* and *text readability* interchangeably.

There is a rich literature on the automatic identification of text readability where research have been done not only for English but also French, German, Dutch, Italian and many other languages (Rabin et al., 1988). These approaches can be broadly classified in three categories (Benjamin, 2012). The classical or traditional methods are those, which were designed and developed during the beginning phase of readability research and are based on shallow text characteristics. Typically, these yield either an absolute score (Flesch, 1948) or a grade level for which a text is considered appropriate (Dale and Chall, 1948; Kincaid and others, 1975;

McLaughlin, 1969). The next in line is the readability metrics motivated by the progress in the field of cognitive science. This class of methods uses more detail text parameters like cohesion, organization and cognitive aspects of the reader. Proposition and inference model (Kintsch and Van Dijk, 1978), prototype theory (Rosch, 1978), latent semantic analysis (Landauer et al., 1998), text leveling methods (Britton and Gülgöz, 1991) are examples of this category. One distinguished instance of this class is Cohmetrix (Graesser et al., 2004). Owing to the advances in the fields of natural language processing (NLP) and machine learning (ML), readability research has seen a revival in the past decade or so. In recent studies, readability has been linked with more complex lexical and syntactic and discourse text characteristics (Petersen and Ostendorf, 2009; Collins-Thompson and Callan, 2005; Si and Callan, 2003; Heilman et al., 2008; Schwarm and Ostendorf, 2005). Several experiments have already established that existing readability measures in English cannot directly be used to compute readability of other languages like, Bangla and Hindi (Sinha et al., 2012b). As compared to the mountainous work done in readability research in foreign languages, very few attempts have been taken to develop readability matrices for Indian language texts, which are quite different from many of their Indo-European cousins (refer to (Sinha et al., 2012b) for a detail account). This makes it even more important to develop similar readability measures for different languages.

Apart from addressing readability from a full-text comprehension point of view, a plethora of research has also been done to model the complexity of sentence comprehension. The syntactic complexity of a sentence can be measured by approaches like T-unit analysis (Bardovi-Harlig, 1992), graph-based approach such as average dependency distance (Oya, 2011). It has been observed that increasing the number of nesting makes a sentence difficult to process (Gordon et al., 2004; Gibson and Fedorenko, 2013). In Hindi, self-paced reading experiments (Vasishth and Lewis, 2006) pro-

vided evidence for facilitation in verbal structures decrease in activation due to the presence of multiple possible candidates during an integration event. A handful of researches have been performed on sentence complexity and word order preference in sentence comprehension (SWINNEY, 1998; Kaiser and Trueswell, 2004). The authors have provided novel methods to compute variation of the relative cognitive load associated with sentence comprehension in Bangla depending on the surface forms (Sinha et al., 2013). Recent studies are focusing on how different languages and different cognitive systems have adopted preference for different word orderings (Gibson et al., 2013; Hall et al., 2013).

Another dimension of complexity in text comprehension is visual word recognition. During visual word recognition process, orthography and phonology play crucial roles (Grainger and Dufau, 2012). One important aspect of an effective read-out model is to associate the orthographic information of a word to its phonological representation. The literature contains many studies provide different architectures to deal with this issue (Perry et al., 2010; Grainger and Ziegler, 2011; Seidenberg and Plaut, 2006). Architectures like BIAM (bi-modal interactive-activation model)(Diependaele et al., 2010), MTMM (multi-trace memory model)(Ans et al., 1998), junction model (Kello and Andrews, 2006) have a phonological layer and an orthographic layer and a mapping functions between them.

Studying text readability has a long-term impact in the field of education and literacy. Easy to read texts improve comprehension, retention and reading persistence (DuBay, 2004). Despite being attached to such a huge population in India, little has been studied in Indian language text readability. The need for specifically focusing on native lies in the fact that people can interpret better, when the documents are in their own languages or mother tongue, i.e. their L1 language (Oakland and Lane, 2004). However, even a native language instruction has to be comprehensible by the target reader; many welfare programs fail, as they require people to have a higher reading level than the present. Therefore, texts have to design, customized and presented in a manner that suits the cognitive capacity of target population. In the context of a country like India, where the literacy rate is well below the world average this is the need of the hour in every level of formal or informal addressing. One of the primary bottlenecks in readability research in any Indian languages are the scarcity of sufficiently large readability annotated data sets with satisfactory level of inter-annotator agreement. Apart from textbooks, there is no annotated corpus of generic texts in Indian languages, that can be used to develop readability prediction models. Moreover, readability is not only a function of textual characteristics. Literacy skills of readers, their cognitive abilities, background knowledge, and other socio-economic factors also play an important role in determining whether a text is readable for a particular group of reader. Therefore, such annotated data sets should also look into the respective readers background level. However, to the best of our knowledge, no such publicly available Indian language datasets with readability annotation is available. Therefore, based on the above discussion, the objective of

this paper is to develop an open-source online computational framework that can be used by different research groups to conduct reading researches on Indian language texts, to measure different attributes that will contribute to its comprehensibility. The framework will also be used to develop a large annotated Indian language text comprehension corpus from user feedback surveys, which will be suitable for automatic NLP applications. The novelty in this framework lies in the fact that it brings different empirical data-collection techniques for text comprehension under one roof. The framework has been customized specifically to address language particularities for Indian languages. It will also offer many types of automatic analysis on the data at different levels such as full text, sentence and word level. To address the subjectivity of text difficulty perception, the framework allows to capture user background against multiple factors. The assimilated data can be automatically cross referenced against varying strata of readers. The measures and predictions will incorporate elegant Natural Language Processing and psycholinguistic methods.

## 2. Design and Development of the Framework

The architecture of the proposed framework is displayed in Figure 1. The framework is primarily classified into the following modules:

- a) Author module
- b) User module
- c) Experimental Data Descriptor and Back-end Resources
- d) Experiment Set
- e) Analysis and Recording of results

In the following subsections, each of these modules is discussed in details.

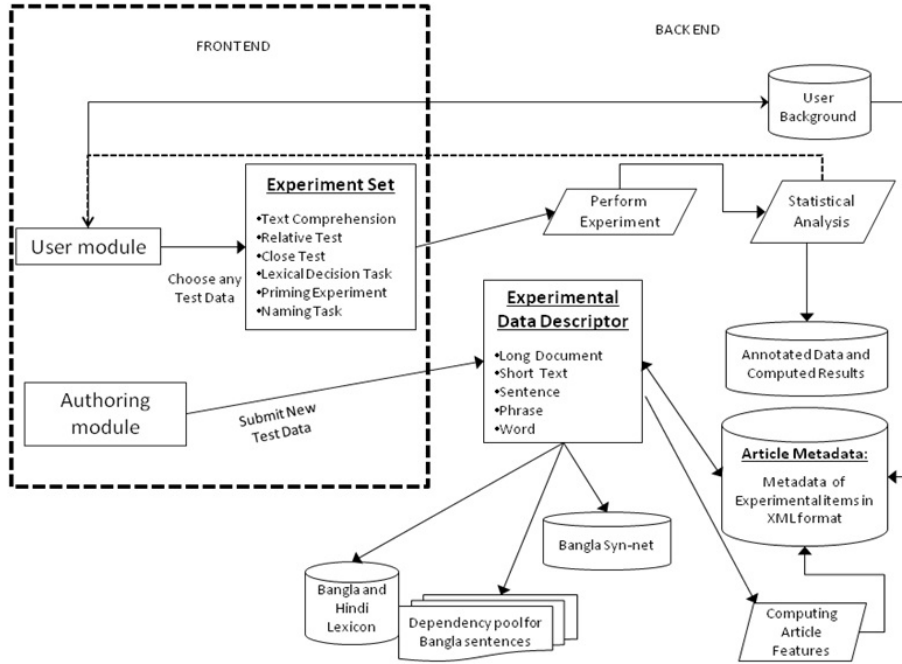
### 2.1. The Author Module

In the author module, an author can submit new experimental data for annotation through the article descriptor interface (see EDD for details). Corresponding to each author, a separate account will be maintained where authors can view, submit, edit, and delete their own dataset and test results.

### 2.2. The User Module

In the user module, a separate interface will be provided which will first perform a detailed survey on the respective user's background. User attributes are reported in Table 1. After the completion of the user survey, the user can choose any of the existing experiment on her preferred language. Presently only three languages, namely, English, Hindi and Bangla, are added into our proposed framework. After the experiment, the user can view her detailed results and the general results about the test such as average rating provided by other participants. The user can also view the records of the previous experiments he/she has performed.

Figure 1: The architecture of the proposed framework



1. Age	6. Home town	11. Hobbies
2. Gender	7. Second language proficiency	12. Parent’s educational qualification
3. School	8. Educational qualification	13. Parent’s occupation
4. Mother tongue	9. Confidence in 1st and 2nd language	14. Parent’s yearly income
5. Medium of study	10. Socio-Economic Class (SEC as per MRSI guidelines)	15. Subject of interest

Table 1: Details of the user attributes

### 2.3. The Experimental Data Descriptor

The experimental data descriptor (EDD) contains different attributes, as shown in Table 3, corresponding to each type of data. We define a data item to be any one of the following types: Long documents (typically > 500 words), short documents (100-500 words), sentence and words. In addition to storing properties against pre-stored data, the EDD can also automatically computes many of these attributes if some new data items are uploaded by author. For this, we have created a large lexicon of Hindi and Bangla words from the corpora mentioned below in Table 2. The attributes that are marked by a \* are to be provided manually by the author themselves. The data item along with the other associated attributes are stored in the metadata storage unit in XML format. Apart from the article properties and lexicon, the EDD is also linked to a Bangla SynNet (Sinha et al., 2012a) and a dependency information pool for which is a collection of text file where each text file corresponds to a sentence (figure 1).

### 3. Experiment Set and Analysis of Results

In the following subsections, we will briefly discuss about these experiments and the different analysis that can be done and recorded using the proposed framework. For each experimental, results are indexed and statistically analyzed

Corpus	Domain	Total Words
Bangla		
CIIL	Tourism	798649
Rabindra Rachanaboli	Literature	3434877
Anandabazar	News	23159022
Bankim Rachanaboli	Literature	1041784
Bangla Blog	Blogs	1916825
Hindi (taken from Emile corpus) <sup>1</sup>		
IndiaInfo	News	600000
Webduniya	News	5600000
Ranchi Express	News	3200000
CIIL	Miscellaneous	3000000

Table 2: Corpus statistics

according to different user category. Users are categorized by their background attributes as stated in table 1. The intergroup and intragroup variances are recorded for each experiment. Experiment statistics are updated dynamically as new results occur.

#### 3.1. Discourse Comprehension

I. **Comprehension Exercise:** Here, users are presented with articles in their preferred language. Each user has

1	Article type: long document, short document, sentence, phrase, or words	29*	Avg. lexical chain span
2	Font style	30*	Avg. no of lexical chains active at each word
3	Font size	31*	Avg. no. of lexical chains active at each NP
4	Language of the article	32	Total jukta-akshar (JUK) or complex character count
5*	Source: Type of the article or the source where it is taken from .	33*	Number of hard words (PSW)
6	Average syllable per word (ASW)	34	Hard words per 30 sentences (PSW30)
7	Average sentence length (ASL)	35	Flesch Reading Ease Score
8	Average word length (AWL)	36	Flesch Kincaid Grade level
9	Total syllable count.	37	Gunning Fog Index level
10	% of words in the document with 3+syllable	38	Automatic Readability Index (ARI)
11*	Avg. no. NPs per sentences		Coleman-Liau Index (CLI)
12*	Avg. no. common + proper nouns per sentence	39	SMOG Grade level (SMOG)
13*	Avg. VPs per sentence	40	Developed formulas for Indian Languages
14*	Avg. Adj per sentence		<b>Word Level Features</b>
15*	Avg. prep. phrases per sentence	41	Word length
16*	Total no. of NPs per sentence	42	Word length in terms of Akshars
17*	Total no. common + proper nouns per sentence	43	Total no. of complex characters
18*	Total VPs per sentence	44*	Total Number of syllables
19*	Total Adj per sentence	45	Corpus frequency and probability of words
20*	Total prep. Phrases per sentence	46	Morphological family size
21*	No. of named entities	47	Parts of speech
22*	No. of unique entities	48	Function or content word
23*	Avg. No. of named entities	49	n-gram from corpus
24*	Avg. No. of unique entities	50	No. of vowels and consonants
25*	No. of lexical chains in the document		<b>Sentence Level Features</b>
26	No of lines	51*	Sentence type: Simple, compound, complex
27*	Avg. lexical chain length	52*	Sentence category: assertive, negative, interrogative
28*	No. of lexical chains, spans, half document length	53*	Number of polysyllabic words
		54*	Number of dependencies
		55*	Average dependency distance
		56*	Dependency information

Table 3: The article metadata

to perform the following three tasks (See fig 4. for an illustration):

- Mark all the words that are difficult to read or comprehend.
- Rate the article on a scale of 1-10. 1 being easier and 10 being hard to comprehend.
- Answer some MCQ questions based on that article. The MCQ questions are evaluated based on the percentage of correct answers.

For each article, average ratings with standard deviation are recorded with user details. Hard words are analyzed according the word attributes mentioned and are stored accordingly. For each hard word, clue from

the SynNet (refer to (Sinha et al., 2012a) for organization and description of SynNet) is provided (if the word exists), if the user selects an option, the options are noted as easy and preferred synonyms against a hard word.

- II. **Relative test:** Here, participants are provided with two articles (Baseline and Target) of a given language. The task is to identify which between the two articles is easier to comprehend then the other. The baseline article is primarily from standard textbooks.

The difference of attributes of the target article with the baseline article is stored along with the user response and user background categories.

- III. **Cloze test:** In a cloze test (Klare et al., 1972), the par-

ticipants were presented with text documents where every 5th word is missing. The task is to fill-up those missing words. Based on the performance of each participant, a score is generated according to the following formula:

$$Score(\%) = \frac{Number\ of\ correct\ answers}{Total\ number\ of\ blanks} * 100 \quad (1)$$

Wrong words are examined based on three scenarios: if they are exact synonyms (from Bangla SynNet), they are treated as equal; if they are related (distance measured by SynNet), they are stored separately according to their degree of relatedness; if above two conditions do not hold or the word is not present in SynNet, then it is stored separately. For all of the experiments, user can view and compare performances over time.

### 3.2. Visual Word Recognition

- **Lexical decision test:** Here, a participant is presented, either visually or auditory, a string of words, non-words or pseudo words. Their task is to indicate, through some key-press, whether the presented word (or the stimulus) is a valid word or not (Meyer and Schvaneveldt, 1971). We customize these experiments by changing the inter-stimulus time (typically ranges between 48 and 300ms). The reaction time by each subject is recorded for analysis.

Response time (both correct and wrong decision) against each individual and average response time for each word is calculated. Along with the absolute values, normalizing each user against their respective mean response time standardizes experimental results. The response time statistics are correlated with word attributes.

- **Naming task:** In this test, the user is shown a series of words one by one and he is asked to record the pronunciations of the words. The objective is to observe the proficiency and fluency of the user for the given word.

Responses are stored .wav format for downloading and further processing. As our framework does not contain any automatic speech recognition system, automatic detection of wrong responses is not available. The response times are analyzed as has been done in case of lexical decision task.

### 3.3. Sentence Comprehension Tests

#### 3.3.1. Self Paced Reading Task (SPRT)

The SPRT experiment uses a moving window paradigm, where each word of a sentence is revealed at a time, while the next and the previous word is kept hidden. The window shifts as user generates a keystroke event. SPRT has been found to be effective in studying how different parts of a sentence affect reader's on-line processing of the sentence (Ferreira and Henderson, 1990). At the end of each sentence, a multiple-choice question is asked based on the

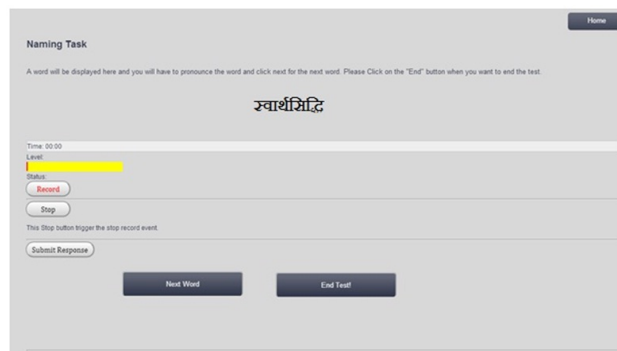


Figure 4: GUI for the lexical decision task

sentence. Among the three options: one is true answer, another is false and the third one says 'none of the above.' The individual and average response times for each word and the sentence as a whole is stored.

For every sentence, the frequency of choices of the answers and the percentage correct answers are recorded. The results can be correlated to the attributes of the different component words and of the sentence as well as its dependency information. Apart from studying the pattern of user responses against each sentence, the trend of an individual with respect to different types of sentences can also be examined.

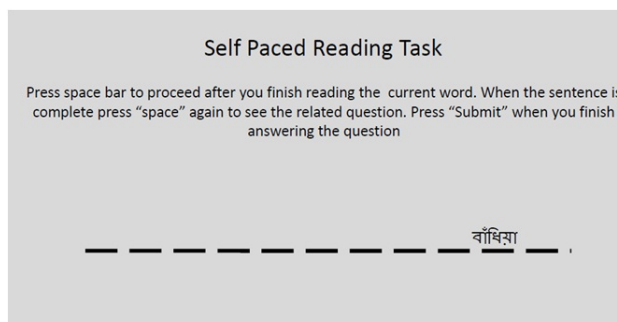


Figure 5: GUI of the self paced reading task.

#### 3.3.2. Effect of Surface Forms

As languages like Bangla and Hindi have relatively free word order, sentences can have multiple surface forms, which are grammatically correct. As word ordering strongly affects sentence processing. We have incorporated two types of experiment to record user response w.r.t to different surface forms.

- Sentence is correct or false: different surface forms of the same sentence are presented to different users and they are asked to respond whether the given sentence is valid or not, as quickly as possible. The variations in response time w.r.t. syntactic and dependency structure can be used to predict the effect of the word ordering on sentence processing.
- Ranking of surface forms: Unlike, the previous test, here a user is presented with three different surface forms of a sentence and she has to rank them as 1, 2 or

Please read the following passage and mark the difficult words that you found or did not understand... Also rate the difficulty of the passage according to you on the scale of 1 to 10. Your readability score will be calculated based on your feedback of difficult words and also your difficulty scale input will be taken into account.

বৈদ্যনাথ গ্রামের মধ্যে বিজ্ঞ ছিলেন সেইজন্য তিনি ভবিষ্যতের শিকে দুটি রাখিয়া বর্তমানের সমস্ত কাজ করিতেন। যখন বিবাহ করিলেন তখন তিনি বর্তমান নববধুর অপেক্ষা ডাবী নবকুমারের মুখ স্পষ্টতরূপে দেখিতে পাইয়াছিলেন। শুভদুষ্টির সময় এতটা দুরদৃষ্টি প্রায় দেখা যায় না। তিনি পাকা লোক ছিলেন সেইজন্য প্রেমের চেয়ে পিওটাকেই অধিক বৃদ্ধিভেদ এবং পুত্রার্থে ক্রিয়তে ভার্য এই মর্মেই তিনি বিনোদিনীকে বিবাহ করিয়াছিলেন। কিন্তু এ সংসারে বিজ্ঞ লোকও ঠকে। বৌবনপ্রাপ্ত হইয়াও যখন বিনোদিনী তাহার সর্বপ্রধান কর্তব্যটি পালন করিল না তখন পুত্রাস নরকের দ্বার খোলা দেখিয়া বেদনাথ বড়ো চিন্তিত হইলেন। মৃত্যুর পরে তাহার বিপুল ঐশ্বর্যই বা কে ভোগ করিবে এই ভাবনায় মৃত্যুর পূর্বে তিনি সেই ঐশ্বর্য ভোগ করিতে বিমুখ হইলেন। পুত্রই বলিয়াছি, বর্তমানের অপেক্ষা ভবিষ্যৎটাকেই তিনি সত্য বলিয়া জানিতেন কিন্তু যুবতী বিনোদিনীর নিকট হঠাৎ এতটা প্রাজ্ঞতা প্রত্যাশা করা যায় না। সে বেচারার দুর্বল্যে বর্তমান, তাহার নববিকশিত বৌবন, বিনা প্রেমে বিকলে অতিবাহিত হইয়া যায় এইটাই তাহার পক্ষে সবচেয়ে শোচনীয় ছিল। পারলৌকিক পিওনের মুখটা সে ইহলৌকিক চিত্তমুখ্যদায়ে একেবারেই জুলিয়া বসিয়াছিল, মনুর পবিত্র বিধান এবং বৈদ্যনাথের অধ্যাত্মিক ব্যাখ্যায় তাহার বুদ্ধিস্তিত হৃদয়ের তিপমাত্র তৃপ্তি হইল না। যে যাহাই বলুক, এই বয়সটাতে ভালোবাসা দেওয়া এবং ভালোবাসা পাওয়াই রমণীর সকল সুখ এবং সকল কর্তব্যের চেয়ে স্বভাবতই বেশি মনে হয়। কিন্তু বিনোদ্যার ভাগ্যে নবপ্রেমের বর্ষাবারিসিকনের বদলে স্বামীর, পিসশাশুড়ীর এবং অন্যান্য গুরু ও গুরুতর লোকের সমৃদ্ধ আকাশ হইতে উর্জন-গর্জনের শিলাবৃষ্টি ব্যবস্থা হইল। সকলেই তাহাকে বন্দ্য্য বলিয়া অপরাধী করিত। একটা ফুলের চারাকে আলোক এবং বাতাস হইতে লক্ষ্য করে রাখিলে তাহার বৈরাগ্য অবস্থা হয়, বিনোদ্যার বক্তিত বৌবনেরও সেইরূপ অবস্থা ঘটয়াছিল। সদাসর্বদা এই-সকল চাপাচুপি ও বকাবকির মধ্যে থাকিতে না পারিয়া যখন সে কুমুমের বাড়ি তাস খেলিতে যাইত সেই সমবর্তা তাহার বড়ো ভালো লাগিত। সেখানে পুংনরকের কীৰ্ত্তন ঘাষা সর্বদা বর্তমান না থাকতে হাসি-ঠাট্টা গভের কোনো বাধা ছিল না। কুমুম যেদিন তাস খেলিবার সাথি না পাইত সেদিন তাহার ভরূপ দেবর নগেন্দ্রকে ধরিয়া আনিত। নগেন্দ্র ও বিনোদ্যার আপত্তি হাসিয়া উড়াইয়া দিত। এ সংসারে এক হইতে আর হয় এবং খেলা ক্রমে স

### Marked Words

ভবিষ্যতের, স্পষ্টতরূপে, ঐশ্বর্য

Your Rating : 7

Submit Words and Rating.

Figure 2: GUI of the comprehension exercise.

3 according to the comprehension difficulty experienced by her. The ranking information can be studied against sentence attributes and dependency description.

Mother tongue	Hindi	Bangla
1st Language	50	80
2nd Language	30	NIL
Graduate	35	58
Post Graduate	15	22

Table 5: User details

### Ranking of surface forms

Rank the following surfaces forms according to the comprehension load experienced by you.

- 1 → Easy
- 2 → Moderate
- 3 → Difficult

১. চার বছর ধরে তাড়া করে চলেছে কথাটা আমায় উম্মাদের মত
২. চার বছর ধরে কথাটা আমায় উম্মাদের মত তাড়া করে চলেছে
৩. চার বছর ধরে উম্মাদের মত তাড়া করে চলেছে কথাটা আমায়

Submit

Figure 6: GUI for sentence with different surface forms

## 4. Creating Data for Bangla and Hindi Experiments

We have created a pilot dataset to run Bangla and Hindi comprehension experiments. Table 4 summarizes the collected dataset. Table 5 contains user details; the age of the subjects varies between 24 to 30 years. Each comprehension experiment has been performed by around 15 participants. Corresponding to each article, we have recorded around  $110 \times 15 = 3250$  annotations, for sentence,  $500 \times 15 = 7500$  and for word,  $1500 \times 15 = 22500$  annotations.

## 5. Conclusion

This paper presents an on-line integrated computational framework to facilitate language comprehension research on Indian languages. The framework provides seven different experiments that are useful for the study of language comprehension. Along with carrying out user experiments it also provide options to analyse and process the experimental data. Presently the framework supports only Bangla and Hindi language but other languages can easily be integrated.

## 6. Acknowledgements

We are thankful to the Society for Natural Language Technology Research for providing us with the necessary Bangla corpus to carry our research. We are also thankful to all those participants who helped us to prepare the Readability annotated Bangla and Hindi texts.

## 7. References

Ans, B., Carbonnel, S., and Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review; Psychological Review*, 105(4):678.



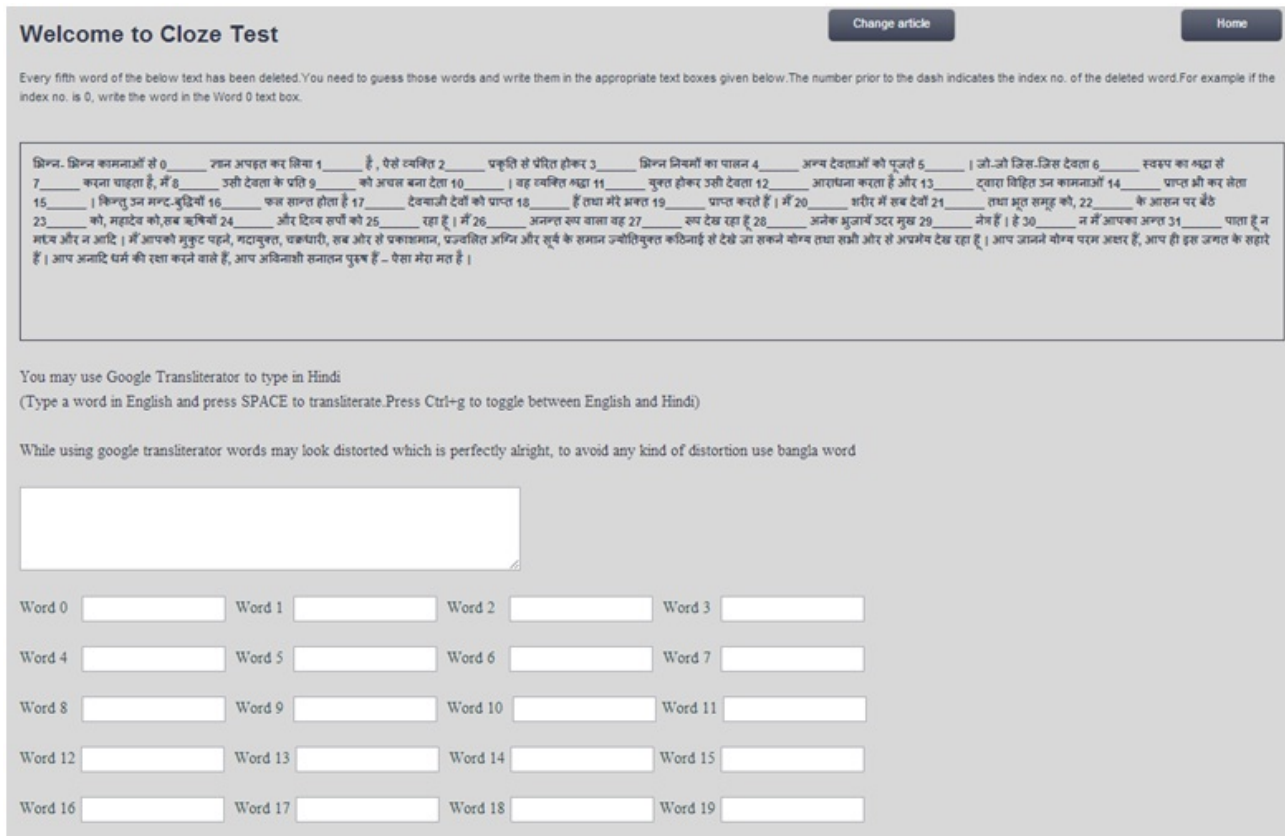


Figure 3: GUI of the Cloze test

Table 4: Details of collected empirical data

Experiments	Total No. of Documents (in each language)	Document Source		
		Textbooks	News	Literature
Reading Comprehension	50	25	9	16
Cloze test	50	25	13	12
Relative tests	10	30	-	-
Sentence Comprehension		Simple	Compound	Complex
SPRT	250	100	100	50
Surface form	250	70	130	50
Word Recognition		High Freq.	Average Freq.	Low Freq.
Lexical Decision Task	900	300	300	300
Naming Task	600	200	200	200

Bardovi-Harlig, K. (1992). A second look at t-unit analysis: Reconsidering the sentence. *TESOL quarterly*, 26(2):390–395.

Benjamin, R. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24:1–26.

Britton, B. and Gülgöz, S. (1991). Using kintsch’s computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83(3):329.

Collins-Thompson, K. and Callan, J. (2005). Predicting reading difficulty with statistical language models. *Journal of the American Society for Information Science and Technology*, 56(13):1448–1462.

Dale, E. and Chall, J. (1948). A formula for predicting readability. *Educational research bulletin*, pages 11–28.

Diependaele, K., Ziegler, J., and Grainger, J. (2010). Fast phonology and the bimodal interactive activation model. *European Journal of Cognitive Psychology*, 22(5):764–778.

- DuBay, W. (2004). The principles of readability. *Impact Information*, pages 1–76.
- Ferreira, F. and Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4):555.
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, 32(3):221.
- Gibson, E. and Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2):88–124.
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., and Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*.
- Gordon, P. C., Hendrick, R., and Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51(1):97–114.
- Graesser, A., McNamara, D., Louwerse, M., and Cai, Z. (2004). Coh-matrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2):193–202.
- Grainger, J. and Dufau, S. (2012). 8 the front end of visual word recognition. *Visual Word Recognition: Models and Methods, Orthography and Phonology*, page 159.
- Grainger, J. and Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in psychology*, 2.
- Hall, M. L., Mayberry, R. I., and Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, 129(1):1 – 17.
- Heilman, M., Collins-Thompson, K., and Eskenazi, M. (2008). An analysis of statistical models and features for reading difficulty prediction. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 71–79. Association for Computational Linguistics.
- Kaiser, E. and Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2):113 – 147.
- Kello, C. and Andrews, S. (2006). Considering the junction model of lexical processing. *From inkmarks to ideas: Current issues in lexical processing*, pages 50–75.
- Kincaid, J. et al. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.
- Kintsch, W. and Van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological review*, 85(5):363.
- Klare, G., SINAICO, H., and STOLUROW, L. (1972). The cloze procedure: A convenient readability test for training materials and translations. *Applied Psychology*, 21(2):77–105.
- Landauer, T., Foltz, P., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284.
- McLaughlin, G. (1969). Smog grading: A new readability formula. *Journal of reading*, 12(8):639–646.
- Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, 90(2):227.
- Oakland, T. and Lane, H. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing*, 4(3):239–252.
- Oya, M. (2011). Syntactic dependency distance as sentence complexity measure. In *Proceedings of The 16th Conference of Pan-Pacific Association of Applied Linguistics*.
- Perry, C., Ziegler, J., and Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (cdp++) model. *Cognitive Psychology*, 61(2):106–151.
- Petersen, S. E. and Ostendorf, M. (2009). A machine learning approach to reading level assessment. *Computer Speech & Language*, 23(1):89–106.
- Rabin, A., Zakaluk, B., and Samuels, S. (1988). Determining difficulty levels of text written in languages other than english. *Readability: Its past, present & future. Newark DE: International Reading Association*, pages 46–76.
- Rosch, E. (1978). Principles of categorization. *Fuzzy grammar: a reader*, pages 91–108.
- Schwarm, S. and Ostendorf, M. (2005). Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 523–530. Association for Computational Linguistics.
- Seidenberg, M. S. and Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. *From inkmarks to ideas: Current issues in lexical processing*, pages 25–49.
- Si, L. and Callan, J. (2003). A semisupervised learning method to merge search engine results. *ACM Transactions on Information Systems (TOIS)*, 21(4):457–491.
- Sinha, M., Jana, A., Dasgupta, T., and Basu, A. (2012a). A new semantic lexicon and similarity measure in bangla. In *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon*, pages 171–182, Mumbai, India, December. The COLING 2012 Organizing Committee.
- Sinha, M., Sharma, S., Dasgupta, T., and Basu, A. (2012b). New readability measures for Bangla and Hindi texts. In *Proceedings of COLING 2012: Posters*, pages 1141–1150, Mumbai, India, December. The COLING 2012 Organizing Committee.
- Sinha, M., Rudra, K., Dasgupta, T., and Basu, A. (2013). How word order affects sentence comprehension in bangla: A computational approach to simple sentence. In *Mining Intelligence and Knowledge Exploration*, pages 769–779. Springer.
- SWINNEY, D. A. (1998). The influence of canonical word order on structural processing. *Syntax and semantics*, 31:153–166.
- Vasishth, S. and Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, pages 767–794.