Proceedings of the Society for Computation in Linguistics

Volume 5 Article 11

2-1-2022

Linguistic Complexity and Planning Effects on Word Duration in Hindi Read Aloud Speech

Sidharth Ranjan
Indian Institute of Technology Delhi, sidharth.ranjan03@gmail.com

Rajakrishnan Rajkumar
IISER Bhopal, rajak@iiserb.ac.in

Sumeet Agarwal
Indian Institute of Technology Delhi, sumeet@iitd.ac.in

Follow this and additional works at: https://scholarworks.umass.edu/scil

Part of the Applied Linguistics Commons, Cognition and Perception Commons, Cognitive Psychology Commons, Cognitive Science Commons, Computational Linguistics Commons, Phonetics and Phonology Commons, Psycholinguistics and Neurolinguistics Commons, and the Typological Linguistics and Linguistic Diversity Commons

Recommended Citation

Ranjan, Sidharth; Rajkumar, Rajakrishnan; and Agarwal, Sumeet (2022) "Linguistic Complexity and Planning Effects on Word Duration in Hindi Read Aloud Speech," *Proceedings of the Society for Computation in Linguistics*: Vol. 5 , Article 11.
DOI: https://doi.org/10.7275/833t-w726
Available at: https://scholarworks.umass.edu/scil/vol5/iss1/11

This Paper is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Proceedings of the Society for Computation in Linguistics by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Linguistic Complexity and Planning Effects on Word Duration in Hindi Read Aloud Speech

Sidharth Ranjan
IIT Delhi
sidharth.ranjan03@gmail.com

Rajakrishnan Rajkumar
IISER Bhopal
rajak@iiserb.ac.in

Sumeet Agarwal
IIT Delhi
sumeet@iitd.ac.in

Abstract

Our study investigates the impact of linguistic complexity and planning on word durations in Hindi read aloud speech. Reading aloud involves both comprehension and production processes, and we use measures defined by two influential theories of sentence comprehension, *Surprisal Theory* and *Dependency Locality Theory*, to model the time taken to enunciate individual words. We model planning processes using an information-theoretic measure we call FORWARD SURPRISAL, inspired by surprisal theory which has been prominent in recent psycholinguistic work. Forward surprisal aims to capture articulatory planning when readers incorporate parafoveal viewing during reading aloud. Using a Linear Mixed Model containing memory and surprisal costs as predictors of word duration in read aloud speech (parts-of-speech and speakers being intercept terms), we investigate the following hypotheses: 1. High values of linguistic complexity measures (lexical+PCFG surprisal and DLT memory costs) lead to high word durations. 2. High values of forward lexical surprisal tend to induce high word durations. 3. High-frequency words are read aloud faster than low-frequency words. We validate the above hypotheses using data from the TDIL corpus of read aloud speech. Further, using a Generalized Linear Model to predict content and function word labels we show that lexical surprisal measures do not help distinguish between these 2 classes. Thus reading aloud might not involve distinct access strategies for content and function words, unlike spontaneous speech.

1 Introduction

Prior work on language production (Ganushchak and Chen, 2016; Navarrete et al., 2016) presents a long-standing debate on the cognitive processes involved in *spontaneous speech* and *reading aloud*. Although both the modalities deal with language production, their unifying accounts have been underexplored in the literature (Sulpizio and Kinoshita, 2016). Spontaneous speech involves the packaging of non-linear conceptual information into linear (sequential) ordering of words in a sentence. In this process, speakers optimize for words, syntactic alternations, and memory load (Slevc, 2011). On the contrary, the cognitive mechanism in reading aloud involves a two-step process, namely *word recognition* and *articulation*. Therefore, various representational levels of words, such as orthographic, phonological, phonemic, and visual information interact with on another to generate the pronunciation of a word.

Motivated by a long of line of previous work in both traditions, our current study investigates the relationship of word duration with linguistic complexity and planning effects in Hindi read aloud speech. To this end, we quantified linguistic complexity using contextual predictability measures defined by *Surprisal Theory* (Hale, 2001; Levy, 2008) and memory costs stipulated by Dependency Locality Theory (DLT, Gibson, 2000). Although surprisal and DLT measures were originally proposed for language comprehension, recent work points towards their efficacy in modelling language production. Mathematically, surprisal is the same as information density. Jaeger (2010) showed that the realization of the optional *that*-complementizer in English spontaneous speech is influenced by uniform information density considerations. Moreover, predictable words tend to be spoken fast (Bell et al., 2003) with reduced emphasis on fine-grained acoustic details (Pluymaekers et al., 2005). In order to investigate planning effects, we used the model-

Proceedings of the Society for Computation in Linguistics (SCiL) 2022, pages 119-132.
Held on-line February 7-9, 2022

ing framework proposed by Bell et al. (2009) for spontaneous speech and adapted their following bigram probability measure to capture *production planning* when reading aloud. We investigated 3 hypotheses using Linear Mixed Models (LMMs, Pinheiro and Bates, 2000) containing all the above measures and low-level predictors generally used in previous work (word frequency and length) to predict word durations (parts-of-speech and speakers being intercept terms). Our hypotheses and their motivation are provided below :

1. *High values of linguistic complexity measures (lexical+PCFG surprisal and DLT integration+storage costs) lead to high word durations*: Researchers have shown that such complexity measures account for production difficulties as well, such as disfluencies (Scontras et al., 2014; Dammalapati et al., 2021) and word duration (Demberg et al., 2012) in spontaneous speech.

2. *High values of forward lexical surprisal tend to induce high word durations*: We deployed a measure named *forward surprisal*, inspired from *Surprisal Theory*) and originally proposed by Ranjan et al. (2020). Cognitively, this measure (negative log probability of a word given upcoming words) models parafoveal preview in the reading part of reading aloud, and thus such look-ahead helps in articulatory planning during subsequent production processes.

3. *High-frequency words are read aloud faster than low-frequency words*: The *Dual Route Cascaded* model (Coltheart et al., 2001, DRC) of word recognition and reading aloud predicted and demonstrated this for isolated single words by means of lexical decision and reading aloud tasks.

All the above hypotheses were validated in our experiments conducted on the publicly available TDIL corpus of read-aloud Hindi speech. Forward surprisal is a significant positive predictor of word durations even in the presence of other factors, pointing towards planning effects in reading aloud. High values of trigram lexical surprisal and PCFG syntactic surprisal along with DLT storage costs induced high word durations. For English spontaneous speech, Bell et al. (2009) revealed asymmetric behavior of lexical predictability measures on function vs. content word duration. They attributed this finding to differences in how content and function words are accessed in the mind (*i.e.*., lexical access during spontaneous speech) apart from their properties pertaining to grammatical function. For *reading aloud* Hindi speech data, we found that lexical predictability of both content and function words have identical effects in predicting reading aloud times. An increase in both backward and forward surprisal measures of lexical surprisal led to identical effects on word durations (*i.e.*, increased durations) of both content and function words in read aloud speech. Going beyond Bell et al. (2009), for the separate task of predicting *content* and *function* class labels for each word using a Generalized Linear Model, we showed that trigram lexical surprisal measures are not significant predictors of word class. In contrast, PCFG surprisal induced a significant boost in prediction accuracy for this task. Thus we found differential effects of lexical and surprisal measures in reading aloud.

Our main contribution is that we extend the prior work motivating our hypotheses (as cited above) by validating them in the presence of a comprehensive host of factors in a language other than English. To the best of our knowledge, this is the first work that explores reading aloud production times in Hindi. Both Ranjan et al. (2020) and Demberg et al. (2012) did not incorporate DLT-based predictors, while the former work did not include syntactic surprisal in their regression models. Scontras et al. (2014) did not factor in surprisal-based factors in their spontaneous production experiments on relative clauses. Finally, the DRC model motivating the third hypothesis above deals with the recognition and production of isolated words. In this work, we extend its prediction to entire sentences. Based on the identical effects of both forward and backward lexical surprisal measures, we offer preliminary evidence that lexical access of items to the extent of the full semantic representation of a word may not be necessary during reading aloud processes. This finding is compatible with the DRC model assumption of word processing via the non-semantic lexical route.

The paper is structured as follows. Section 2 provides background on theories and models pertaining to this work. Section 3 presents the details about the dataset and methods used in this work. Section 4 illustrates our main experiments and their results. Section 5 summarizes our main findings and discusses their implications along with pointers to future work.

2 Background

The following subsections provide essential background on the Hindi language and its orthography, the Dual Route Cascaded (DRC) model, Dependency Locality Theory, and Surprisal Theory.

2.1 Hindi Language and Script

Hindi is a head-final language with relatively free word order (with Subject-Object-Verb being the canonical order) compared to English, and has a rich case-marking system realized as postpositions (Agnihotri, 2007). Hindi adopts the Devanagari alphasyllabary-based writing system. The Devanagari script is composed of 47 characters containing 33 consonants (क, ख, ग, etc.) and 14 vowels (अ, आ, इ, etc.). In terms of letter-sound correspondence, the orthography of the script mostly corresponds with grapheme pronunciation except for cases when vowel diacritics, conjunct consonants or ligatures are present (Vaid and Gupta, 2002). Further details of the script are provided in Appendix C.

2.2 Dual Route Cascaded (DRC) Model

The DRC model is a computational model of the *visual word recognition* and *reading aloud*. The model posits two separate cognitive routes i.e., *lexical* and *sub-lexical* that are involved in reading aloud, and within each route, the information processing occurs in a cascaded fashion (Coltheart et al., 2001). It is a computational implementation of the dual-route theory of reading and further stipulates three routes for word processing, *viz.* Grapheme-Phoneme Correspondence (GPC) route, Lexical Semantic route and Lexical Nonsemantic route. Figure 5 in Appendix B provides a visual illustration of the DRC model. Empirical evidence for the efficacy of the DRC model emerges from its ability to simulate human latencies in the tasks of reading aloud and lexical decision tasks. DRC adapts the rationale for *frequency effects* from earlier work on word processing. Morton (1969) demonstrated that high frequency words required lower evidence from visual input (*i.e.*, letters in reading) on account of their lower activation. Subsequently, word naming occurs on account of a lexical search procedure (Forster and Chambers, 1973) where activation levels affect search latencies.

2.3 Dependency Locality Theory

Dependency Locality Theory is a theory of sentence comprehension proposed by Gibson (2000) which posits two processing costs at each word, *viz*, INTEGRATION and STORAGE COSTS (defined and exemplified in Section 3). DLT predictions about the increased comprehension difficulty of object relative clauses over subject relative clauses have been validated using per-word reading time data in a variety of languages. Scontras et al. (2014) showed that object relative clauses are harder to produce than subject relative clauses and relative clause production times are connected to DLT-based memory costs. For Hindi, the eye-tracking based reading times in comprehension have been known to be influenced by DLT-inspired costs (Husain et al., 2015; Agrawal et al., 2017).

2.4 Surprisal Theory

Surprisal Theory (Hale, 2001; Levy, 2008) posits that comprehenders construct probabilistic knowledge based on previously encountered structures. Mathematically, *surprisal* of the $(k+1)^{th}$ word, w_{k+1}, is defined as negative logarithm of conditional probability of word, w_{k+1} given the preceding context which can be either sequence of words or a syntactic tree:

$$S_{k+1} = -\log P(w_{k+1}|w_{1...k}) = \log \frac{P(w_1...w_k)}{P(w_1...w_{k+1})} \quad (1)$$

Both the versions of surprisal i.e., *lexical and syntactic* configurations have been shown to account for eye-movements reading (Demberg and Keller, 2008; Agrawal et al., 2017; Staub, 2015) as well as self-paced reading time data (Smith and Levy, 2013). Pioneering work by Demberg et al. (2012) showed that both n-gram and PCFG-based syntactic surprisal measures were significant positive predictors of word duration in spontaneous speech. More recently, Dammalapati et al. (2021) demonstrated that surprisal and DLT-based metrics predict speech disfluency using English spontaneous speech corpus.

3 Data and Methods

Our dataset consists of 1531 sentences (from scientific and technical genre) from the TDIL corpus of Hindi read aloud speech[1]. One male and one female speaker were asked to record their speech by reading aloud 341 sentences (4,444 words) and 1,190 sentences (11,163 words), respectively. Table 4 in Appendix C illustrates pertinent word-level properties (overall and grammatical category-wise). Word durations were extracted from the recorded speech using the PRAAT software package. We estimated various word-level cognitive measures as described below:

1. **Word length:** Total number of consonants and vowels present in the word (इसलिए – *isliye; therefore* has word length of 4; 2 consonants (स, ल) and 2 vowels (इ, ए).

2. **Word frequency**: Count of each target word as obtained from the EMILLE Hindi corpus (Baker et al., 2002).

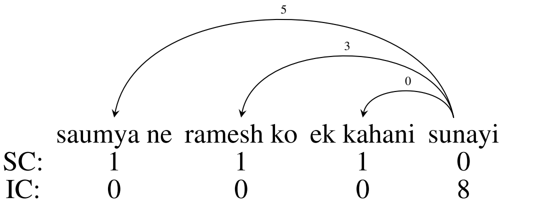
3. **Unigram surprisal:** Negative log probability of individual target word.

4. **Backward surprisal:** Negative log of probability of target word given two preceding words in the context (Equation 1).

5. **Forward surprisal:** Negative log of probability of target word given two following words in the context. So the surprisal of the k^{th} word is estimated as: $S_k = -\log P(w_k \mid w_{k+1}, w_{k+2})$

6. **PCFG surprisal:** Negative log probability of target word given contextual syntactic tree (Equation 1).

7. **Integration cost (IC)**: Backward looking cost denoting the sum of distances between the word to be integrated into the structure processed so far and its previous heads/dependents. Distance is the number of intervening words between each head and dependent.



	saumya ne	ramesh ko	ek kahani	sunayi
SC:	1	1	1	0
IC:	0	0	0	8

Figure 1: Integration and storage cost calculations for the sentence '*Saumya narrated a story to Ramesh*', with head-dependent distance indicated above each dependency link; example sentence adapted from Husain et al. (2015)

8. **Storage cost (SC)**: Forward-looking cost corresponding to the number of incomplete dependencies in the upcoming structure.

Unigram and Trigram Surprisal measures for each word in a sentence was computed using unigram and trigram language models respectively trained on the EMILLE corpus of written text with mixed genre (Baker et al., 2002) using the SRILM toolkit (Stolcke, 2002) with Good-Turing discounting smoothing algorithm. **PCFG surprisal** for each word was estimated by training an incremental probabilistic left-corner parser (van Schijndel et al., 2013) on 13,000 phrase structure trees (converted from HUTB dependency trees) using ModelBlocks toolkit[2] (Refer Appendix D for more details on training data and settings). We calculated DLT IC and SC costs automatically following the definitions adopted by Husain et al. (2015). See Figure 1 for an illustration. They computed DLT costs by hand for a small corpus, while our DLT SC and IC costs were computed from dependency trees obtained by parsing TDIL sentences using the ISC dependency parser[3] (Bhat, 2017) trained on HUTB gold standard dependency trees (parser performance documented by Bhat: UAS of 93.52% and a LAS of 87.77%).

4 Experiments and Results

In the following subsections we describe the specific experiments and results of this study.

4.1 Correlation Results

Prior to performing the regression experiments described in the next few subsections, we computed

[1] https://tdil-dc.in

[2] https://github.com/modelblocks

[3] https://bitbucket.org/account/user/iscnlp/projects/ISCNLP

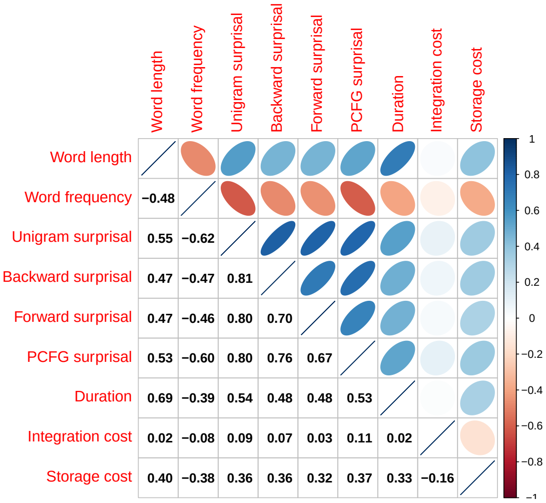


Figure 2: Pearson's correlation coefficients amongst the different predictors and word duration

the Pearson's coefficient of correlation between the different predictors. We also computed the correlation between each predictor and the dependent variable, word duration. Figure 2 displays the correlation results. The high positive correlation between word duration and all surprisal scores suggests that the words which are easy to produce by virtue of high predictability in context tend to have lower reading time and vice versa. DLT-storage costs display low correlation with other predictors, while integration cost shows negligible correlation with any other predictor, indicating their independent impact. SC and IC costs show low negative correlation with one another as they are forward and backward-looking costs respectively and thus might work differently. We also observe that word length is highly correlated with word duration as is observed in previous production (Bell et al., 2009) and comprehension studies (Husain et al., 2015; Agrawal et al., 2017).

4.2 Regression Experiments

We trained Linear Mixed Models (LMMs) to predict per-word duration (transformed to a logarithmic scale following previous work). The logarithmic scaling of the independent variables, *viz.* surprisal measures, took care of highly varied frequencies during model training. All the independent variables were normalised to z-scores, *i.e.*, the predictor's value (centered around its mean) was divided by its standard deviation. We have used the Glm package in R to perform our regression experiments using a very basic model, given below in R GLM format (independent variable $\sim$ dependent variables + 1| random intercept terms):

$$Duration \sim word\ length + word\ frequency + unigram\ surprisal + backward\ surprisal + forward\ surprisal + PCFG\ surprisal + IC + SC + 1|Speaker + 1|POS$$

The POS intercepts were based on tags obtained by converting HUTB POS tags to 11 universal POS tags corresponding to content words (verb, noun, adjective, and adverb) as well function words (postposition, pronoun, determiner, particle, conjunction, question, and quantifier).

Predictors	**Estimate**	**Std. Error**	**t-value**
Intercept	5.525	0.098	56.364
Word length	0.217	0.003	62.430
Unigram surprisal	0.027	0.006	4.284
Word frequency	-0.034	0.004	-7.643
SC	0.010	0.004	2.309
IC	-0.016	0.003	-5.830
Backward 3g-surprisal	0.015	0.005	3.128
Forward 3g-surprisal	0.032	0.004	7.181
PCFG surprisal	0.051	0.005	10.412

Table 1: Fixed effects of an LMM predicting reading aloud time (15607 data points; all predictors are significant for the |t|=2 threshold)

Our regression results documented in Table 1 reveal that all the measures are significant in predicting the read-aloud word duration and their regression coefficients are in the expected direction, thus validating our original hypotheses stated in Section 1. Frequency and unigram surprisal capture the frequency and predictability effects of individual words, *i.e.*, frequent words require less time and effort to activate phonemes for articulation (as predicted by the DRC model). The positive coefficients of all surprisal and DLT SC measures show that with an increase in each predictor's value, the word duration in read-aloud speech increases. However, DLT IC has an unexpected negative coefficient, an anomaly which has been also reported in the comprehension literature (Demberg and Keller, 2008; Husain et al., 2015). Demberg and Keller (2008) analyzed this anomaly rigorously and showed that in the presence of other predictors, integration cost works in the expected direction (*i.e.*, high integration costs induce high reading times) only for higher range IC values. Future inquiries need to examine whether this result carries over to the production setting and the implications of such a finding for integrated models of both processes (a theme we take up at the end of Section 5). In the following subsections, we now discuss the impact of selected measures on reading aloud word duration.

4.2.1 Forward Surprisal

The positive regression coefficient of forward surprisal (Table 1) suggests that the difficulty associated with the upcoming words has a role in determining the reading time of the current word. The effect of forward surprisal on duration is illustrated using the following examples (region of interest: *vidyalaye*; *school*):

1. pahle pitaji bacchon=ko **vidyalaye=se** lene jaate the
   before father child=ACC school=ABL take go be-PST.3SG
   Earlier father used to take children from school

2. bacche **vidyalaye=se** aate hi khelne chale gaye
   children school=ABL come EMPH play go-PST.PL
   The children went to play as soon as they came from school

In the first example above, the word *vidyalaye* (550ms duration; 4.55 forward surprisal) has a higher surprisal and longer duration compared to the same word in the second sentence (510ms; 3.90 bits). This is because *vidyalaye se aate* is a much more frequent sequence than *vidyalaye se lene* in the trigram training corpus. Thus planning effects are modelled by this measure, a theme we explore in the next subsection.

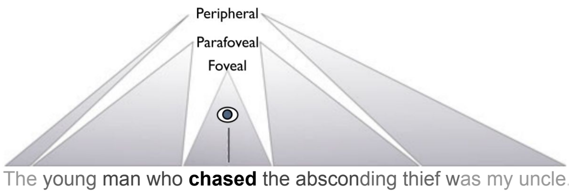


Figure 3: Parafoveal preview in reading; adapted from Schotter et al. (2012)

Interactions	Estimate	Std. Error	t-value
MODEL 1			
Word length x Backward 3g-surp	**-0.024**	0.004	-5.491
Word length x Forward 3g-surp	**-0.031**	0.004	-8.061
Word length x PCFG surprisal	0.001	0.005	0.314
MODEL 2			
Function word x Backward 3g-surp	**0.028**	0.009	2.936
Function word x Forward 3g-surp	**0.041**	0.008	4.855
Function word x PCFG surprisal	**-0.039**	0.009	-3.953

Table 2: Two different LMMs displaying only the interaction terms of surprisal with word length (top) and function word (bottom) respectively predicting reading aloud time; see full model results in Appendix E (15607 data points; all significant predictors denoted by |t|>2)

4.2.2 Parafoveal Preview and Word Length Effects

It is well understood that the length of words influences the reader's eye movements as long words induce more fixations of greater duration than short words (Just and Carpenter, 1980; Rayner et al., 1996). In this context, Bicknell and Levy (2012) argue that uncertainty about the length of words affects the word reading duration. They posit that the uncertainty increases proportionally with an increase in word length, leading to more fixation and longer word duration. We hypothesize that if the forward surprisal effect is driven by parafoveal previewing (as illustrated in Figure 3), there should be smaller predictability effects with longer target words. This is because longer target words will lead to less linguistic material visible in the parafoveal region, thus not allowing for informative computation of the target word's forward surprisal. We investigated the effect of word length on word duration using another linear mixed model containing word length and surprisal interaction terms. Table 2 (top block) documents the interaction results, which show that the effect of forward trigram surprisal on reading-aloud times decreases by 0.02 with every unit increase in the word length, thus confirming our hypothesis. A similar result is obtained in case of backward trigram surprisal as well. See Table 5 in Appendix E for full regression model results. The relative strengths of forward and backward surprisal measures in both production and comprehension needs to be systematically investigated in future inquiries.

4.2.3 Word Class and Duration

This section and the next one are motivated by the findings of Bell et al. (2009). For spontaneous speech, they showed that both function and content word duration were significantly predicted by the following word (*forward probability*). However, unlike content words, function word duration was determined significantly by the previous word only (*backward probability*). Content words are associated more with semantics, whereas function words are linked to the syntactic aspects of the sentence (see Table 4 of Appendix C for more details about their properties). In order to investigate the relationship between predictability measures and word class in read-aloud speech, we deployed a Linear Mixed Model with speaker and POS random effect terms for duration prediction. Fixed effects included all the predictors along with interaction terms between word class and trigram lexical+PCFG syntactic surprisal measures. Each word in our dataset was annotated with a word class label (*viz.*, *content* or *function* word) derived from its universal POS tag. Table 2 (see bottom block of table) depicts the significant interaction effects between both lexical surprisal measures and word class. High values of both forward and backward trigram surprisal induced high function word duration in read aloud speech after controlling for several other factors. This result is in contrast to the asymmetric behavior observed by Bell et al. (2009) for function words in conversational English speech. See Table 6 in Appendix E for full regression model results.

Predictor(s)	**10-fold CV prediction accuracy (%)**
Word length	68.91
+Word frequency	76.10
+Unigram surp	77.65
+Backward 3g-surp	77.02
+Forward trigram surp	77.14
+PCFG surprisal	79.61
+SC	80.21
+IC	83.94

Table 3: Prediction accuracy for content and function word classification (on the entire dataset of 15607 data points) via Generalized LMs where features are added incrementally (all differences between successive pairs of models significant at $p < 0.001$ via McNemar's test)

Counter-intuitively, the interaction term between word class and PCFG surprisal has a negative coefficent, signifying that high values of PCFG surprisal result in low word durations for function words. Examining this anomaly, we looked at function word distributions in our dataset (TDIL corpus) and the corpus used to train the PCFG parser (HUTB corpus). Table 4 in Appendix C lists grammatical category-wise distribution of HUTB and TDIL words. Particles (3.73%) and question words (1.38%) words have higher mean surprisal and lower mean duration compared to the corresponding mean values for the function word class in TDIL corpus. The high surprisal of words belonging to these grammatical categories can be attributed to the fact that the PCFG parser training data from the HUTB corpus (particles: 1.59%, questions: 0.11%) has very few words belonging to these categories, thus impacting PCFG surprisal estimates. The following examples illustrate question words like *kis* (183ms duration and 12.16bits PCFG surprisal) and particles like *toh* (675ms and 9.5bits):

(1) a.

yeh	aag	**kis**	hanuman	dwara	lagayi
this	fire	WHICH	hanuman	by	set

gayi	hogi?
would?	

Which Hanuman would have set this fire?

b.

ab	tak	**toh**	pitaji	so	gaye	honge
by	now	PARTICLE	father	sleep	must	

By now, father must have been asleep.

The information profiles and per-word read-aloud word duration of the above examples from our dataset are presented in Figure 4 of Appendix A. Cognitively, it is also conceivable that *WH*-markers and particles might be easy to articulate being very common function words. However, they might potentially introduce complex mental operations like movement (or linking to other words in non-movement based accounts) in the upcoming structure, which are reflected in the duration of the next word (akin to spillover in reading studies). This conjecture is supported by the fact that words following question words and particles have higher duration on an average compared to the mean duration of these target function words themselves (question words: 225ms & next word 274ms; particles: 155ms & next word 292ms mean duration).

4.2.4 Word Class Prediction and PCFG Surprisal

Extending the work by Bell et al. (2009) (who do not factor in syntactic predictability estimates) described in the previous section, we explored the impact of all our measures for predicting word class using Generalized Linear Models (GLMs). For this binary classification task, function words were coded as class 1, while content words were coded as 0. Subsequently, we added each predictor incrementally to a GLM and measured the prediction accuracy of the model via 10-fold cross-validation (CV). The corpus was divided into 10 sections and 10 models trained on 9 sections each were used to generate predictions for the remaining section, thus obtaining predictions over the entire dataset. Table 3 provides CV prediction accuracies of all our incremental models. Low-level predictors, frequency and unigram surprisal, confer significant gains over a basic word length baseline. However, adding backward and forward surprisal actually worsens model performance and hence these measures do not help distinguish between content and function words. This result thus validates our findings pertaining to word class and lexical surprisal measures reported in Table 2 (bottom block). In contrast, PCFG surprisal confers a 2% increase in predicting the word class. PCFG surprisal is a more powerful measure compared to word-based surprisal models as it factors in POS tag information and syntactic context and hence outperforms word-based trigram models. DLT-costs also induce significant gains over and above models containing low-level and all other surprisal predictors. In particular, integration cost induces close to a 2.5% increase over a model containing all the other predictors.

5 Discussion

Overall, our results validate our initial hypotheses motivating the study. Linguistic complexity measures (lexical+PCFG surprisal and DLT's integration + storage costs) are significant positive predictors of word duration in reading aloud speech, mirroring trends reported in the literature on spontaneous speech production (Demberg et al., 2012; Dammalapati et al., 2021). Our measure of planning, FORWARD SURPRISAL, is also a positive predictor of reading aloud times. It potentially models parafoveal preview in the reading aspect of reading aloud. Such look-ahead during reading likely helps articulatory planning during the reading aloud process. This finding advances further support to the "*involvement-in-planning*" account, as proposed by Pluymaekers et al. (2005). The cited work shows that articulatory processes are continuous and incremental in nature; upcoming words affect the planning of the target word. Finally, our data and analyses validate the *frequency effects* (high frequency words are read aloud faster than low frequency words) predicted by the DRC model of word recognition and reading aloud.

Going further, we show that an increase in both our measures of lexical surprisal (*viz.,* backward and forward surprisal) led to identical effects on word duration of content and function words in read aloud speech, *i.e.*, increased duration for both classes of words. For the binary classification task of predicting content and function words, PCFG surprisal induces a notable boost in accuracy over a baseline containing low-level predictors and lexical surprisal measures. However, forward and backward surprisal do not help discriminate between content and function words. This is in direct contrast to the results reported by Bell et al. (2009) for spontaneous speech (Switchboard corpus). They show evidence for differential lexical access mechanisms for content and function words as attested to by the long line of work in the production literature (Garrett, 1975, 1980; Lapointe and Dell, 1989). Thus via this work, we have compared the cognitive processes in reading aloud with spontaneous speech production, an underexplored direction highlighted by Sulpizio and Kinoshita (2016) whom we cited at the outset.

Our results indicate that both content and function words might rely on the non-semantic lexical route or grapheme-phoneme correspondence (GPC) rules as hypothesized by the DRC model of reading aloud. Speakers might not be doing semantic processing during this task. The close symbol-sound correspondence in Hindi orthography (Vaid and Gupta, 2002) might be a factor contributing to this effect, a conjecture that needs to be validated using further experiments. The measure of word complexity proposed by Husain et al. (2015) and character-based surprisal models of reading difficulty proposed in recent work (Hahn et al., 2019; Oh et al., 2021) might be viable approaches towards this end. Situations where the connection between orthographic length and pronunciation length is complex (say "535" in written text articulated as *panch sau paintis*, *i.e.*, “five hundred and thirty five”) are best investigated using more controlled experimental designs.[4]

In a recent survey, Staub (2015) summarized that lexical predictability induces the graded activation of multiple upcoming words during reading comprehension (as opposed to the prediction of a single word). Moreover, lexical predictability effects occur either at the very early stages of lexical access or pre-lexical stages (processing visual features of letters in the script), rather than at post-lexical stages involving meaning identification. Based on insights from prior work, high syntactic predictability (low PCFG surprisal values in our setup) can be linked to high accessibility and hence the ease of word retrieval from memory, which in turn facilitates production ease (Bock and Warren, 1985; Arnold, 2010). Future inquiries need to tease apart the contributions of lexical and syntactic predictability in reading aloud, quantifying the impact of language-specific properties of the Hindi language on reading aloud durations. In particular, the verb-final nature of Hindi and prior findings about the interplay between expectation and locality effects (Husain et al., 2014; Ranjan et al., 2021) need to be explored. Other salient aspects like predictability and case marking (Ranjan et al., 2019), and the impact of the argument-adjunct distinction (Pandey et al., 2022), could also be investigated to contribute to a comprehensive theory of reading aloud, which accounts for data from multiple language families.

We also plan to develop reading aloud speech corpora with a larger number of participants. Moreover, the current task of reading the printed text aloud can be modified to include comprehension questions (à la reading studies) to ensure that participants engage with the material. We also plan to collect eye-tracking times to study comprehension during the reading phase prior to reading aloud. Thus this paradigm can catalyze research in integrated models of production and comprehension (MacDonald, 2013; Pickering and Garrod, 2013). Levy and Gibson (2013) point out that the surprisal measure is an *incremental* and *localized* measure of comprehension difficulty, which can be used to formalize such integrated models. Since this measure can be used to model production difficulty as well, it facilitates cross-linguistic hypothesis testing on both comprehension and production as well as interactions between these processes.

Acknowledgements

We would like to thank Marten van Schijndel, Samvit Dammalapati, Rupesh Pandey, and members of the Cornell Computational Psycholinguistics Discussions research group (C.Psyd) for their invaluable comments and feedback on this work. We also thank IISER Bhopal undergraduate students, Chandan Upadhyay, Vipul Bhadani, Priya Kumari, Naman Joshi, and Kshitiz Dixit for their help in speech transcription. The authors acknowledge extramural funding from the Cognitive Science Research Initiative, Department of Science and Technology, Government of India (grant no. DST/CSRI/2018/263). Finally, we are also indebted to the anonymous reviewers for SCiL 2022 and AMLaP 2020 for their detailed and insightful feedback.

References

Rama Kant Agnihotri. 2007. *Hindi: An Essential Grammar*. Essential Grammars. Routledge.

Arpit Agrawal, Sumeet Agarwal, and Samar Husain. 2017. Role of expectation and working memory constraints in Hindi comprehension: An eyetracking corpus analysis. *Journal of Eye Movement Research*, 10(2).

Jennifer E. Arnold. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4):187–203.

Paul Baker, Andrew Hardie, Tony McEnery, Hamish Cunningham, and Robert Gaizauskas. 2002. Emille: a 67-million word corpus of indic languages: data collection, mark-up and harmonization. In *Proceedings of LREC 2002*, pages 819–827. Lancaster University.

Alan Bell, Jason M Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.

Alan Bell, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in english

[4] We are indebted to an anonymous reviewer for this suggestion and the example.

conversation. *The Journal of the Acoustical Society of America*, 113(2):1001–1024.

Riyaz Ahmad Bhat. 2017. *Exploiting linguistic knowledge to address representation and sparsity issues in dependency parsing of indian languages*. Ph.D. thesis, IIIT Hyderabad India.

Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, and Fei Xia. 2009. A multi-representational and multi-layered treebank for Hindi/urdu. In *Proceedings of the Third Linguistic Annotation Workshop*, ACL-IJCNLP '09, pages 186–189, Stroudsburg, PA, USA. Association for Computational Linguistics.

Klinton Bicknell and Roger Levy. 2012. Why long words take longer to read: the role of uncertainty about word length. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics*, pages 21–30. Association for Computational Linguistics.

J. Kathryn Bock and Richard K Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21:47–67.

Max Coltheart, Kathleen Rastle, Conrad Perry, Robyn Langdon, and Johannes Ziegler. 2001. Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1):204.

Samvit Dammalapati, Rajakrishnan Rajkumar, and Sumeet Agarwal. 2021. Effects of duration, locality, and surprisal in speech disfluency prediction in english spontaneous speech. In *Proceedings of the Society for Computation in Linguistics*, volume 4, page 10.

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Vera Demberg, Asad B. Sayeed, Philip J. Gorinski, and Nikolaos Engonopoulos. 2012. Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 356–367, Stroudsburg, PA, USA. Association for Computational Linguistics.

Kenneth I. Forster and Susan M. Chambers. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6):627–635.

Lesya Y Ganushchak and Yiya Chen. 2016. Incrementality in planning of speech during speaking and reading aloud: Evidence from eye-tracking. *Frontiers in psychology*, 7:33.

Merrill Garrett. 1980. Levels of processing in sentence production. In *Language production Vol. 1: Speech and talk*, pages 177–220. Academic Press.

Merrill F Garrett. 1975. The analysis of sentence production. In *Psychology of learning and motivation*, volume 9, pages 133–177. Elsevier.

Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, pages 95–126.

Michael Hahn, Frank Keller, Yonatan Bisk, and Yonatan Belinkov. 2019. Character-based surprisal as a model of reading difficulty in the presence of errors. In *Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation, Montreal, Canada, July 24-27, 2019*, pages 401–407. cognitivesciencesociety.org.

John Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, NAACL '01, pages 1–8, Pittsburgh, Pennsylvania. Association for Computational Linguistics.

Samar Husain, Shravan Vasishth, and Narayanan Srinivasan. 2014. Strong expectations cancel locality effects: Evidence from Hindi. *PLOS ONE*, 9(7):1–14.

Samar Husain, Shravan Vasishth, and Narayanan Srinivasan. 2015. Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eye-tracking corpus. *Journal of Eye Movement Research*, 8(2).

T. Florian Jaeger. 2010. Redundancy and reduction: Speakers manage information density. *Cognitive Psychology*, 61(1):23–62.

Marcel A Just and Patricia A Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329.

Steven G Lapointe and Gary S Dell. 1989. A synthesis of some recent work in sentence production. In *Linguistic structure in language processing*, pages 107–156. Springer.

Roger Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126 – 1177.

Roger Levy and Edward Gibson. 2013. Surprisal, the pdc, and the primary locus of processing difficulty in relative clauses. *Frontiers in Psychology*, 4(229).

Maryellen C. MacDonald. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology*, 4(226):1–16. Published with commentaries in Frontiers.

John Morton. 1969. Interaction of information in word recognition. *Psychological review*, 76(2):165.

Eduardo Navarrete, Bradford Z Mahon, Anna Lorenzoni, and Francesca Peressotti. 2016. What can written-words tell us about lexical retrieval in speech production? *Frontiers in psychology*, 6:1982.

Byung-Doh Oh, Christian Clark, and William Schuler. 2021. Surprisal estimators for human reading times need character models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3746–3757, Online. Association for Computational Linguistics.

Rupesh Pandey, Sidharth Ranjan, and Rajakrishnan Rajkumar. 2022. Locality effects in the processing of argument structure and information status using reading aloud paradigm. In *Proceedings of the 8th Annual conference of the Association for Cognitive Science (ACCS)*, India. Amrita University.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, ACL-44, pages 433–440, Stroudsburg, PA, USA. Association for Computational Linguistics.

Martin J. Pickering and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36:329–347.

José C Pinheiro and Douglas M Bates. 2000. Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, pages 3–56.

Mark Pluymaekers, Mirjam Ernestus, and R Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken dutch. *The Journal of the Acoustical Society of America*, 118(4):2561–2569.

Sidharth Ranjan, Sumeet Agarwal, and Rajakrishnan Rajkumar. 2019. Surprisal and interference effects of case markers in Hindi word order. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 30–42, Minneapolis, Minnesota. Association for Computational Linguistics.

Sidharth Ranjan, Rajakrishnan Rajkumar, and Sumeet Agarwal. 2020. Forward surprisal models production planning in reading aloud. In *Proceedings of the 26th Architectures and Mechanisms for Language Processing Conference (AMLaP)*, Potsdam, Germany. University of Potsdam.

Sidharth Ranjan, Rajakrishnan Rajkumar, and Sumeet Agarwal. 2021. Locality and Expectation Effects in Hindi Preverbal Constituent Ordering. *Cognition*, in press.

Keith Rayner, Sara C Sereno, and Gary E Raney. 1996. Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5):1188.

Elizabeth R Schotter, Bernhard Angele, and Keith Rayner. 2012. Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, 74(1):5–35.

Gregory Scontras, William Badecker, Lisa Shank, Eunice Lim, and Evelina Fedorenko. 2014. Syntactic complexity effects in sentence production. *Cognitive Science*, 39(3):559–583.

L. Robert Slevc. 2011. Saying what's on your mind: working memory effects on sentence production. *Journal of experimental psychology. Learning, memory, and cognition*, 37(6):1503–1514.

Nathaniel J. Smith and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319.

Adrian Staub. 2015. The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8):311–327.

Andreas Stolcke. 2002. SRILM — An extensible language modeling toolkit. In *Proc. ICSLP-02*.

Simone Sulpizio and Sachiko Kinoshita. 2016. bridging reading aloud and speech production. *Frontiers in psychology*, 7:661.

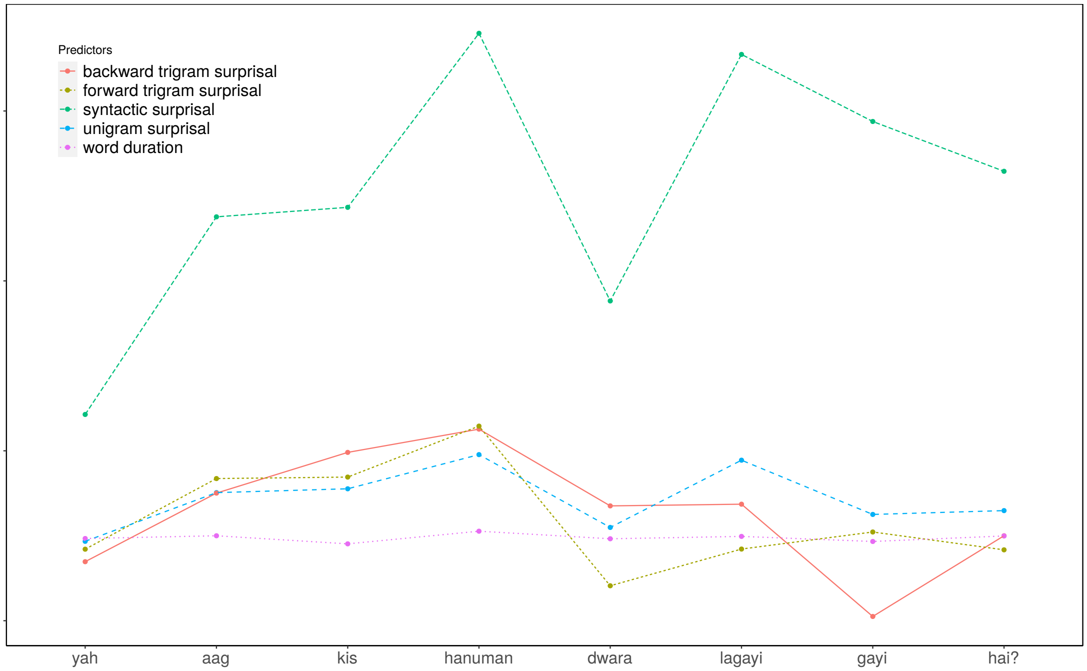
J Vaid and Anshum Gupta. 2002. Exploring word recognition in a semi-alphabetic script: The case of devanagari. *Brain and Language*, 81:679–90.

Marten van Schijndel, Andy Exley, and William Schuler. 2013. A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, 5(3):522–540.

Himanshu Yadav, Ashwini Vaidya, and Samar Husain. 2017. Keeping it simple: Generating phrase structure trees from a Hindi dependency treebank. In *TLT*.

A Information Profile

Figure 4 depicts the information profiles of Examples 1a and 1b respectively from the TDIL corpus discussed in Section 4.2.3 of the paper.



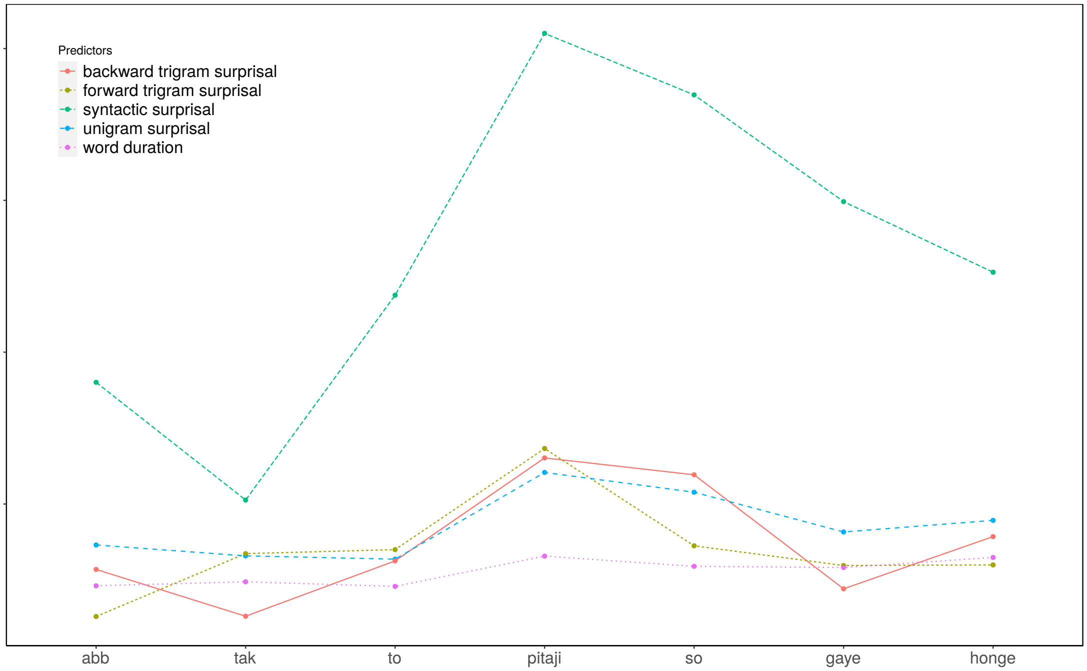


Figure 4: Word duration and information profiles of sentences containing a question marker (*kis*; top figure) and particle (*toh*; bottom figure)

B Dual Route Cascaded (DRC) Model

The DRC model is shown in Figure 5. Each route consists of several interacting layers containing a set of units (representing words in the orthographic lexicon or letters in the letters layer). Units of different layers interact via *inhibition* (an activated unit impedes activation levels of other units) or *excitation* (an activated unit facilitates activation of other units). Figure 3 shows a snapshot of parafoveal preview in reading.

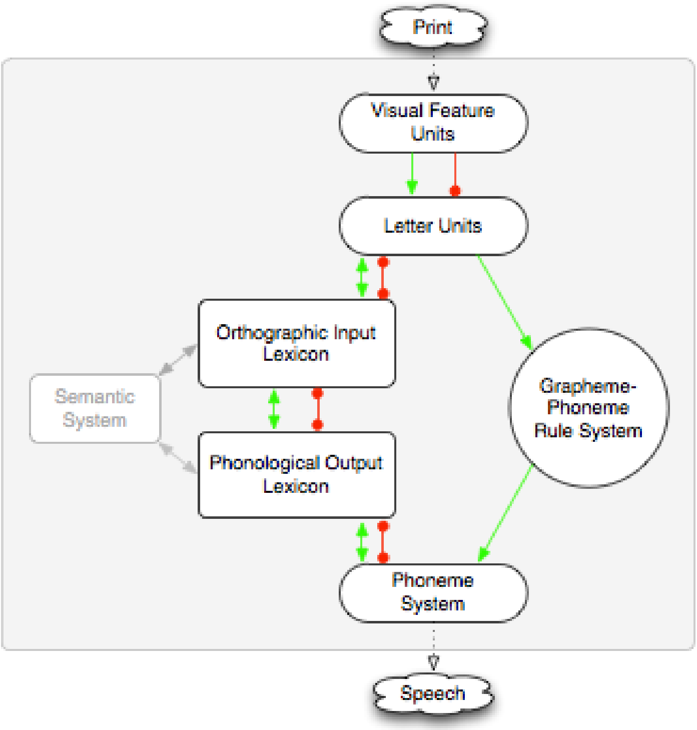

Print
Visual Feature Units
Letter Units
Orthographic Input Lexicon
Semantic System
Phonological Output Lexicon
Grapheme-Phoneme Rule System
Phoneme System
Speech

Figure 5: DRC model[a] of visual word recognition and reading aloud by Coltheart et al. (2001)

[a] Reproduced from: https://maxcoltheart.wordpress.com/drc/

C Details of Hindi Script and Grammatical Categories

Unlike the Latin alphabet, Hindi has no concept of letter case (upper/lower) except for sinistrodextral (left-to-write) writing system. Each unit of word is written in horizontal direction separated by space and follows standard punctuation markers alike English except for full stop (.) where a pipe (।) is used as an end of sentence marker. Vowel diacritics (glyph) combines with consonants to form another syllabic letter (आ + क = का). For example, the vowel – आ ($\bar{a}$) combines with consonant – क (k) to give a letter का ($k\bar{a}$) with added vowel sign in diacritic form. Conjunct consonants is understood to offer most difficulty during reading consist of two consonants grouped together but with a missing vowel sound between them. For example, the two consonants (च, छ्) when combined together (च + छ् = च्छ्), the letter च्छ् (as in the word– अच्छा) has a missing vowel (अ) diacritic i.e., ा between them.

Table 4 illustrates the distribution of various grammatical categories in TDIL and HUTB corpora of Hindi written text as well as properties of content and function words. The mean word length of a content word in the TDIL corpus was 2.66 (minimum: 1, maximum: 8), and the function word was 1.74 (minimum: 1, maximum: 5).

Category	**%Freq**	**%Freq**	**Length**	**PCFG**	**RT**
■	273013 words	15607 words	characters	**surprisal**	ms
Corpus	HUTB	TDIL	Mean values in TDIL		
CONTENT					
Verb	18.12	32.15	1.98	11.26	274.99
Noun	38.47	26.96	2.86	13.92	375.45
Adjective	5.91	3.71	3.01	14.53	399.65
Adverb	0.47	0.78	2.88	14.21	367.91
FUNCTION					
Postposition	21.42	11.14	1.22	5.58	178.22
Pronoun	4.34	11.07	2.20	10.62	258.12
Det	4.65	4.64	1.86	8.87	242.32
Particle	1.59	3.73	1.16	**8.42**	**155.02**
Conjunction	4.13	3.17	1.87	7.65	206.01
Question	0.11	1.38	0.11	**12.59**	**225.97**
Quantifier	0.81	1.27	0.81	11.23	294.96
Content words	62.97	63.70	2.66	13.96	354.29
Function words	37.03	36.40	1.74	8.60	226.54
All words	100.00	100.00	2.17	11.10	286.17

Table 4: Grammatical category-wise descriptive statistics in TDIL and HUTB corpora

D PCFG Parser Training Procedures

Following steps were involved in training the Modelblocks parser using the HUTB corpus:

1. The parser training requires phrase-structure trees as input. Due to the unavailability of such resources in Hindi, we created our own corpus by converting the existing dependency parsed trees (Dependency structure; DS) of HUTB corpus (Bhatt et al., 2009) into constituency parsed trees (Phrase structure; PS) using an approach described in Yadav et al. (2017).

2. However, we had to do some extra post-processing of the obtained phrase structure trees (removal null nodes, unary nodes, punctation and coordination fixes, inter-alia) to make it compatible with the format expected by the Berkeley parser. The corrected final phrase structures thus produced were used to train the Berkeley parser model.

3. Parser training involved estimating a sophisticated grammar using 4 iterations of the split-merge algorithm (Petrov et al., 2006) and a beamwidth of 5000 (shown to be effective for reading time studies).

E Interaction analysis of word class and word length with surprisal

Predictors	**Estimate**	**Std. Error**	**t-value**
Intercept	5.550	0.098	56.825
Word length	0.237	0.004	60.946
Unigram surprisal	0.039	0.006	6.118
Word frequency	-0.004	0.005	-0.777
IC	-0.018	0.003	-6.550
SC	0.005	0.004	1.106
Backward surprisal	0.028	0.005	5.556
Forward surprisal	0.044	0.005	9.653
PCFG surprisal	0.034	0.005	6.904
INTERACTIONS			
Word length x Backward 3g-surp	-0.024	0.004	-5.491
Word length x Forward 3g-surp	-0.031	0.004	-8.061
Word length x PCFG surprisal	0.001	0.005	0.314

Table 5: Fixed effects of LMM (with word length as interaction term) predicting reading aloud time (15607 data points; all significant predictors denoted by |t|>2)

Predictors	**Estimate**	**Std. Error**	**t-value**
Intercept	5.512	0.099	55.652
Word length	0.216	0.003	62.147
Unigram surprisal	0.036	0.007	5.451
Word frequency	-0.028	0.005	-6.038
SC	0.012	0.005	2.517
IC	-0.016	0.003	-5.171
Backward 3g-surp	0.007	0.006	1.095
Forward 3g-surp	0.018	0.005	3.364
PCFG surprisal	0.065	0.007	9.192
Word class	0.024	0.011	2.264
INTERACTIONS			
Function word x Backward 3g-surp	0.028	0.009	2.936
Function word x Forward 3g-surp	0.041	0.008	4.855
Function word x PCFG surprisal	-0.039	0.009	-3.953

Table 6: Fixed effects of LMM (with word class as interaction term) predicting reading aloud time (15607 data points; all significant predictors denoted by |t|>2)