Preverbal syntactic complexity leads to local coherence effects in Hindi

Sakshi Bhatia

Department of Linguistics, Central University of Rajasthan, Ajmer, India

Samar Husain

Department of Humanities and Social Sciences, Indian Institute of Technology Delhi,
India

Author Note

Correspondence concerning this article should be addressed to Sakshi Bhatia,
Email: sakshi.bhatia@curaj.ac.in or to Samar Husain, E-mail: samar@hss.iitd.ac.in

Abstract

The effective use of preverbal linguistic cues to make successful clause-final verbal prediction as well as robust prediction maintenance has been argued to be a cross-linguistic generalization for SOV languages such as German and Japanese. In this paper, we show that native speakers of Hindi (an SOV language) falter in maintaining clause-final verbal predictions in the presence of a center-embedded relative clause with a non-canonical word order. The fallibility of the parser is illustrated by the formation of a grammatically illicit locally coherent parse as well as by poor comprehension accuracy. Our investigations suggest that while plausibility is essential, presence of overt agreement features might not be necessary for forming a locally coherent parse in Hindi. The work highlights how top-down processing and bottom-up information interact during sentence comprehension in SOV languages – comprehension suffers with increased complexity of the preverbal linguistic environment.

Keywords: Local Coherence Effect; Syntactic prediction; Good-Enough Processing; Sentence Comprehension; SOV languages

Preverbal syntactic complexity leads to local coherence effects in Hindi

Introduction

Top-down processing forms a critical component of sentence comprehension (e.g., Altmann & Kamide, 1999, 2007; Fischler & Bloom, 1979; Kliegl, Grabner, Rolfs, & Engbert, 2004; Kutas & Hillyard, 1984, 1980; Levy & Keller, 2013; Marslen-Wilson, 1973; Schwanenflugel & Lacount, 1988; Schwanenflugel & Shoben, 1985; Staub & Clifton, 2006; Zola, 1984). For example, Altmann and Kamide (1999) have shown that English native speakers effectively use information about the verb's selectional restriction to preactivate the appropriate syntactic object prior to the object onset. In a visual world paradigm, participants selectively moved their gaze to the image of a 'cake' on hearing 'eat' in the utterance such as 1(a). In contrast, while attending to sentences such as 1(b), they started looking at the image of a 'cake' only at the onset of its utterance.

- (1) a. The boy will eat the cake
- b. The boy will move the cake

This line of research (Altmann & Kamide, 1999, 2007, among others) highlights how prediction can lead to the formation of grammatically licensed structures. See, Kuperberg and Jaeger (2016); Kutas, DeLong, and Smith (2011); Levy (2013); Staub (2015) for recent reviews on the role of prediction during sentence comprehension. Interestingly, the generated predictions need not always be licensed by the grammar. For example, in a recent study, Kamide and Kukona (2018) demonstrated that native speakers of English generate predictions that are not consistent with the global structure of the sentence. Rather, the generated parse is consistent only with the local input. Participants were more likely to look at the image of a 'bike' after hearing the 'the man from London' in 2a compared to 'the man very much' in 2b. In addition, participants also looked at the globally consistent object 'carousel' in condition 2a suggesting that both the local as well as the global parse were active.

- (2) a. The girl who likes the man from London will ride the carousel

- b. The girl who likes the man very much will ride the carousel

The above study highlights that in certain syntactic configurations, the parsing process can lead to grammatically illicit prediction. One such configuration is the locally coherent structure (the man from London will ride in 2) which has previously been shown to trigger parsing errors (Konieczny, Müller-Feldmeth, Hachmann, Schwarzkopf, & Wolfer, 2009; Tabor, Galantucci, & Richardson, 2004). Such locally consistent (but globally inconsistent) parses seem to arise in embedded structures such as relative clauses which are known to incur memory load (Gibson, 1998; Grodner & Gibson, 2005; Levy & Keller, 2013). In fact, another much discussed parsing error occurs in configurations involving multiple embeddings. In the context of the so called structural forgetting effect (Frazier, 1985), Gibson and Thomas (1999) showed that English native speakers rated sentences such as 3a as good as sentences such as 3b. Note that 3b is ungrammatical, as a verb ('was cleaning') in the embedded structure is missing; this missing verb is denoted by a ' $\emptyset$ ' symbol. Gibson and Thomas (1999) explained this effect by assuming that the parser is unable to maintain the prediction of multiple verbs in 3a due to increased memory load and therefore tends to forget the verb which incurs the maximum cost.

- (3) a. The apartment that the maid who the service had sent over was cleaning every week was well decorated.
- b. * The apartment that the maid who the service had sent over $\emptyset$ was well decorated.

Effects such as local coherence and structural forgetting highlight that the parser is susceptible to forming grammatically illicit parses. While in the case of local coherence, the global parse was presumably active along with the local parse; in the case of structural forgetting the global parse was completely abandoned in favor of a simpler structure. Both these effects happen with clausal embeddings, suggesting that the parser is particularly vulnerable in such syntactic configurations.

A parsing mechanism that can lead to grammatically illicit parses is compatible with a

variety of theoretical approaches to sentence processing. For example, a noisy-channel model of language processing (e.g., Gibson et al., 2013; Kuperberg & Jaeger, 2016; Piantadosi, Tily, & Gibson, 2011) is able to predict the formation of grammatically illicit parses. The noisy channel model assumes that the comprehension system is predictive because it needs to accommodate for the noisy transmission of information from the speaker (cf. Kurumada & Jaeger, 2015). Therefore, unexpected input during comprehension can lead to effects such as local coherence (Kamide & Kukona, 2018; Konieczny et al., 2009; Tabor et al., 2004) and misinterpretation due to non-canonical order (Ferreira, 2003). Similarly, the good-enough processing system (Ferreira, Bailey, & Ferraro, 2002) also assumes that speakers make mistakes and therefore comprehenders rely on various heuristics to accommodate their error-ridden input to arrive at meaning. Under a good-enough system, the parser can rely on very little information from the input, leading to a shallow or incorrect interpretation of the received input. It has been proposed that a processing of such a kind could be a result of limits on time and resources that the parsing system is typically subject to (Ferreira & Patson, 2007). Finally, local coherence has been explained by a self-organizing parsing mechanism (Tabor et al., 2004) where global syntactic structures compete with structures formed based on local input. Together, we refer to such proposals as the *fallibility hypothesis*. The *fallibility hypothesis* states that the comprehension system can generate grammatically invalid parses due to factors such as a processing load, etc.

Interestingly, the structural forgetting effect in the context of multiple clausal embeddings discussed above has been shown to disappear in SOV languages like German and Dutch. Vasishth, Suckow, Lewis, and Kern (2010) used the German equivalent of sentences 3a-b to demonstrate that German native speakers rate 3b worse than 3a (though see Häussler & Bader, 2015). Dutch has also been shown to pattern with the German results (Frank, Trompenaars, & Vasishth, 2016). Vasishth et al. (2010) explain this lack of forgetting in German by arguing that German native speakers get exposed to clause final SOV structures much more than speakers of English. Hence, the parser in German becomes adept at efficient handling of such verb-final structures. A

similar rationale has been invoked to explain the anti-locality effect in SOV languages. Anti-locality effects are characterised by a facilitation at the clause final verb with increased distance between the verb and its prior dependents. For example, Vasishth and Lewis (2006) used items such as 4 a-b to show that reading times at the relative clause verb *dekhaa* ‘saw’ were faster in the long conditions compared to the reading times in the short conditions.

(4) a. Short, Subject Relative

vo-lar.kaa jisne us-kaagaz-ko **dekhaa** bahut-jigyaasu thaa
 that-boy who that-paper-acc saw very-inquisitive was
 ‘That boy who saw that (piece of) paper was very inquisitive.’

b. Long, Subject Relative

vo-lar.kaa jisne us-kaagaz-ko mez ke-piiche gire-hue **dekhaa**
 that-boy who that-paper-acc table behind fallen saw
 bahut-jigyaasu thaa
 very-inquisitive was
 ‘That boy who saw that (piece of) paper fallen behind a/the table was very inquisitive.’

Such facilitatory effects are generally understood to arise due to the better prediction of the clause final verb in the long condition compared to the short condition. While Vasishth and Lewis (2006) explain anti-locality through memory reactivation, their model assumes a left-corner parsing mechanism to predict the relevant clause final verbal phrase. Note that the facilitation is found within a relative clause which is deemed to be a difficult to process syntactic environment (e.g. Gibson, 1998; Just & Carpenter, 1992; Miller & Chomsky, 1963; Yngve, 1960). Similar effects have been demonstrated in other head-final languages as well such as German (Konieczny, 2000) and Japanese (Nakatani & Gibson, 2010), pointing to the role of parser adaptability. Information theoretic metrics such as surprisal (Hale, 2006; Levy, 2008) have been used to account for the prediction-based effects such as anti-locality as high probability structures are associated with low surprisal values.

To summarize, effects such as lack of structural forgetting and anti-locality effects in

SOV languages has been explained by the *adaptability hypothesis*. According to this hypothesis, the parser becomes better at processing certain linguistic patterns in a language because of the frequent exposure to those patterns (Christiansen & Chater, 1999; Christiansen & MacDonald, 2009; Engelmann & Vasishth, 2009; Hsiao & MacDonald, 2013; Levy, 2008, 2013; MacDonald & Christiansen, 2002; Vasishth et al., 2010). In the current context, the parser for SOV languages becomes efficient in robust verbal prediction based on preverbal cues. Consequently, a key prediction of the adaptability hypothesis is that comprehenders of SOV languages such as Hindi, Japanese, German, etc. will effectively use the preverbal information (e.g., nominal semantics, case-markers, etc.) to correctly predict a valid clause-final verb. In addition, as demonstrated by the lack of structural forgetting in these languages, the hypothesis also predicts that such predictions should be maintained robustly even in syntactic configurations involving clausal embeddings.

Given the lack of structural forgetting and facilitation in processing of clause final verb within relative clauses, it is quite possible that local coherence effects that arise within RCs in English would fail to arise in an SOV language like Hindi. This is because in Hindi, the prediction of the RC structure as well as the matrix verb should be maintained robustly given previous evidence for efficient processing of clause final structures in SOV languages. In such a scenario the globally consistent structure should always be chosen over any locally consistent (grammatically illicit) structure.

In this paper, we investigate two questions, (1) Can a local coherence effect be observed in an SOV language like Hindi? and (2) If yes, then what are the possible factors due to which such an effect may arise? We report 6 experiments to investigate these questions. Following the frequent use of center-embedded relative clause (RC) structures to illustrate robust prediction in Hindi (Husain, Vasishth, & Srinivasan, 2014; Vasishth & Lewis, 2006), we also utilize center-embedded RCs in our experiments. In particular, we investigate if the prediction of a clause-final verb (and the relevant structure) is forgotten in the head final language Hindi in the presence of a center-embedded relative clause (RC) with non-canonical word order. This word-order manipulation is central to

our investigation into local coherence effects in Hindi. A failure to observe such an effect would be consistent with the *adaptability hypothesis*. On the other hand, if a local coherence effect is observed, that would be consistent with the *fallibility hypothesis*. In addition, we also investigate the role of two factors – verbal agreement and semantic plausibility – in inducing a locally coherent parse in Hindi.

The paper is arranged as follows: first, we discuss the key experiment that demonstrates a local coherence effect in Hindi (Experiment 1). Next, we discuss three experiments (Experiments 2-4) that investigate syntactic and semantic factors that could lead to a local coherence effect arising. We probe the dependencies formed in head-final structures with center-embedded RCs and word order variation through targeted comprehension questions in Experiment 5. These results are further investigated using a sentence completion study in Experiment 6. Finally, we consolidate the findings of all the experiments in the General Discussion and conclude the paper.

Experiment 1

The goal of this experiment was to investigate if native speakers of Hindi form a locally coherent parse during online comprehension. In order to probe this we employed the design in Husain et al. (2014), shown in 5. In condition 5a, the NP *vah laRkaa* ‘That boy_{Masc}’ is modified by a subject relative clause (RC) which has an SOV word order. In condition 5b, this relative clause has a non-canonical order such that the object of the relative clause verb *kitaab* ‘book_{Fem}’ appears after the relative clause verb *paRhii thii* ‘read_{Fem} had_{Fem}’; i.e., the RC has an SVO word order. Husain et al. (2014) found a significantly longer reading time at the RC verb in condition 5b compared to the RC verb in condition 5a. The RT difference at the RC verb was explained through an expectation-based account (Levy, 2008) by arguing that SOV structure of the RC in condition 5a is strongly expected compared to the non-canonical SVO structure found in 5b. Not encountering the expected OV order in 5b lead to a processing slowdown. Critically, their experiment demonstrates that the RC structure is successfully predicted during the comprehension process.

(5) a. Canonical, Subject Relative

Vah laRkaa jisne kal bahut dilchaspai se kitaab
 That boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem}
paRhii thii, meraa dost hai
read_{Fem} had_{Fem} my_{Gen.Masc} friend is
 ‘That boy, who read the book with great interest, is my friend’

b. Non-canonical, Subject Relative

Vah laRkaa jisne kal bahut dilchaspai se **paRhii thii**
 That boy_{Masc} who_{Erg} yesterday lots interest with **read_{Fem} had_{Fem}**
 kitaab, meraa dost hai
 book_{Fem} my_{Gen.Masc} friend is
 ‘That boy, who read the book with great interest, is my friend’

In the current experiment, we adapted their design to evaluate Hindi speakers’ propensity to build a locally coherent parse involving relative clause internal material and the post-RC material. Specifically, we used a post-RC verb which cannot be grammatically integrated with other matrix clause material but which is compatible with relative clause internal material in the non-canonical order as part of a locally coherent parse.

Material and methods

Participants. 52 native speakers of Hindi participated in this experiment. All the participants in the experiments reported in this paper were undergraduate or graduate students from the Indian Institute of Technology, Delhi (IIT, Delhi) between the ages of 18 to 40 years, with the exception of Experiment 3, where we also included students from Jawaharlal Nehru University, Delhi. Participants provided informed consent and were paid between INR 150 to 300 for their participation. The amount was pro-rated as a function of time spent in the lab: INR 150 was paid for 30 minutes of experiment participation; payments for experiments that lasted more than 30 minutes increased accordingly.

Items. We used a 2×2 design crossing RC-internal WORD ORDER (Canonical/Non-canonical) and post-RC CLAUSE TYPE (Copula/Transitive). The

Canonical condition has SOV order in the RC (6a,c), while the Non-canonical conditions has SVO order (6b,d). Importantly, all four experimental conditions are ungrammatical – the post-RC material cannot be integrated with the RC head noun NP_{Masc} (*vah laRkaa* ‘That boy_{Masc}’ in 6 a-d). In the Copula conditions (6a,b), the post-RC verb incorrectly bears a feminine agreement, matching the RC internal object NP Fem (*kitaab* ‘book_{Fem}’) rather than the features of NP_{Masc}. In the Transitive conditions (6c,d), in addition to the incorrect agreement, NP_{Masc} and the post-RC verb are thematically incompatible. We use underlining in the sample item below to draw attention to the locally coherent parse.

24 experimental items in 4 conditions were prepared. 56 filler sentences with varied levels of acceptability were also included. In the current experiment (and in subsequent experiments), some of the filler sentence were ungrammatical. These ungrammatical filler sentences were used to filter out participants who were not paying attention during the data analysis phase. A latin-square design was used to present different items.¹

(6) a. Canonical, Copula

vaha laRkaa jisne kal bahut dilchaspai se kitaab
 That boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem}
 paRhii thii **moTii thii** ...
 read_{Fem} had_{Fem} **fat_{Fem} was_{Fem}** ...

b. Non-canonical, Copula

vaha laRkaa jisne kal bahut dilchaspai se paRhii thii
 That boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem}
kitaab **moTii thii** ...
book_{Fem} **fat_{Fem} was_{Fem}** ...

c. Canonical, Transitive

vaha laRkaa jisne kal bahut dilchaspai se kitaab
 That boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem}
 paRhii thii mujhe **bechnii paRii** ...
 read_{Fem} had_{Fem} I_{Dat} **sell_{Fem} had-to_{Fem}** ...

d. Non-canonical, Transitive

¹ All the data and analysis files for this and subsequent experiments can be obtained from :

vaha laRkaa jisne kal bahut dilchaspai se paRhii thii
 That boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem}
 kitaab mujhe **bechnii paRii** ...
 book_{Fem} I_{Dat} sell_{Fem} had-to_{Fem} ...

Procedure. We used the self-paced reading (SPR) paradigm (Just, Carpenter, & Woolley, 1982) with centered presentation for the reading task. Stimulus items were presented using the online Ibex Farm platform. A Latin square design ensured that each participant sees each item in only one condition. The target items and fillers were randomized for each participant.

The experiment began by explaining the task to the participants verbally as well as in written form. After this, several practice sentences were presented in order to familiarize participants with the task. At the beginning of each trial, the participant saw a fixation cross at the center of the screen. When the participant pressed the space-bar key, the first word was displayed at the center of the screen. With each successive press of the space bar, the previous word disappeared and the next word or phrase was displayed. This successive replacement continued until the participant had read the whole sentence. Reading times or RTs (in milliseconds) were taken as the measure of relative momentary processing difficulty. In addition, participants were asked to rate the sentence on a scale of 1–7 (with 1 signifying completely unacceptable and 7 signifying completely acceptable) at the end of each trial.

Predictions

The *adaptability hypothesis* does not predict any difference in RTs between the conditions at the critical post-RC verb. This is because these accounts assume that the parser is actively predicting and successfully maintaining this clause-final verb (Levy, 2008; Vasishth et al., 2010). Since all the four conditions become ungrammatical at the critical post-RC region, the adaptability hypothesis, will predict equal difficulty in all the four conditions. Indeed, based on a metric like surprisal (Hale, 2006; Levy, 2008) that quantifies the processing difficulty at the word as its conditional probability in the sentential context, the probability of encountering the critical verb-forms in the experimental items will be close to zero. This is on account of a critical assumption that

surprisal makes: the probability mass is distributed over only grammatically licensed structures.

On the other hand, the *fallibility hypothesis* will predict that increased processing load due to the unexpected SVO order of the relative clause could make the parser susceptible to making errors. Under such a condition, a locally coherent structure (Tabor et al., 2004) could be formed such that the RC internal NP_{Fem} gets structurally integrated with the post-RC material in the Non-canonical condition. We would therefore expect to see a main effect of Word Order such that RTs at the (post-RC) matrix verb in the Non-canonical condition should be less than the Canonical condition. No such locally coherent structure is possible in the Canonical order. The predictions discussed above were pre-registered on AsPredicted.com.

Also, since the SPR paradigm is known to have spillover effects (Kaiser, 2014), the predicted RT difference can also appear at the post-critical regions. Consequently, similar to the critical region, the post-critical regions were kept identical across conditions.

In addition to the RTs, the *fallibility hypothesis* also makes predictions regarding grammaticality ratings: it may be the case that comprehenders are susceptible to an illusion of grammaticality for the non-canonical conditions due to the availability of a locally coherent parse.

Results

For the data analyses in Experiment 1 and all subsequent experiments the statistical analysis for the RT/acceptability data was done using linear mixed-effects model through the `lme4` package (Bates, Machler, Bolker, & Walker, 2015) in the *R* statistical computing environment (R Core Team, 2013). Maximal models were fit when possible (Barr, Levy, Scheepers, & Tily, 2013); in case of convergence failure, a less complex model was fit by successively removing the random slopes of the by-subject and by-item random effects component. Raw RTs were log transformed and acceptability ratings were scaled before fitting the lmer models.

In addition, participants who rated ungrammatical fillers more than 3.7 were removed from analysis. The threshold of 3.7 was chosen as it was the value for the 3rd quartile of all ungrammatical ratings in this experiment. This step led to the removal of data from 14 participants.

The average ratings for the experimental conditions were found to be significantly higher than the rating for the ungrammatical filler sentences ($t=-12.32$). The mean rating for the experimental items was 3.7 (Min=1, 1st Quartile=2, Median=4, 3rd Quartile=5, Max=7) while the mean rating for the ungrammatical fillers was 2.4 (Min=1, 1st Quartile=1, Median=2, 3rd Quartile=3, Max=7). The analysis of the acceptability ratings found a significant effect of Word Order ($t=-4.22$) such that the ratings for the non-canonical conditions were less than the ratings for the canonical condition. These figures can be seen in Tables 1, 2.

Table 1

Experiment 1: Average acceptability ratings for the four conditions.

Condition	Rating
a. Canonical, Copula	4.1
b. Non-Canonical, Copula	3.4
c. Canonical, Transitive	3.8
d. Non-Canonical, Transitive	3.4

RTs at the critical post-RC matrix verb region showed a significant main effect of Clause Type ($t=-2.0$) such that RTs in the Transitive conditions were higher than the RTs at the Copula conditions. In addition, a marginal interaction effect ($t=1.68$) was observed – in a nested contrast the Non-canonical Transitive condition was marginally faster than the Canonical Transitive condition ($t=-1.9$), while no difference was found in the Copula conditions ($t=0.01$). RTs at the post-critical region showed a significant effect of Word Order ($t=-2.77$) such that the Non-canonical canonical conditions was read faster than the Canonical conditions. Table 3 shows the results for both critical and the post-critical regions. The RTs at the critical and the post-critical regions are

Table 2

Experiment 1: lmer results for the acceptability ratings. Anova sum contrasts were used for the two main effects and the interaction.

	Coefficient	SE	t-value
Intercept	0.014	0.09	0.14
Word Order	-0.13	0.03	-4.22
Clause Type	0.05	0.05	0.99
Word Order:Clause Type	-0.03	0.03	-1.00

shown in Figure 1. Figure 2 shows the RTs for all the regions across all the experimental conditions. We also note that not removing the participants based on the exclusion criterion mentioned above did not change the salient findings of the presented analyses.

Table 3

Experiment 1: lmer results for the RTs at the critical and post-critical regions. Anova sum contrasts were used for the two main effects and the interaction.

	Critical Region			Post-critical Region		
	Coefficient	SE	t-value	Coefficient	SE	t-value
Intercept	6.72	0.05	127.08	6.48	0.04	157.61
Word Order	-0.01	0.02	-0.72	-0.04	0.01	-2.77
Clause Type	-0.05	0.02	-2.0	-0.02	0.01	-1.43
Word Order:Clause Type	0.02	0.01	1.68	-0.01	0.01	-1.02

Discussion

The RT results show that Hindi native speakers are susceptible to local coherence in the non-canonical conditions where the RC internal NP_{Fem} object appears after the RC verb. Additionally, the acceptability ratings shows that non-canonical order conditions are dispreferred compared to the canonical order conditions.

The RT results are surprising as they show that, in the face of a locally coherent parse,

Figure 1. Experiment 1: Reading time at the critical and the post-critical region in the 4 conditions. C.Cop=Canonical, Copula; NC.Cop=Non-canonical, Copula; C.Trans=Canonical, Transitive; NC.Trans=Non-canonical, Transitive.

native speakers of the head-final language Hindi are unable to sustain the prediction of the matrix verb that was to be integrated with the head noun, NP_{Masc} . This goes against the *adaptability hypothesis* that predicts robust prediction and successful maintenance of clause final structures even in the face of clausal embeddings. These results cannot be explained by the surprisal metric that would predict no difference between various conditions.

What can explain this effect? There are two possibilities that we consider. The first relates the use of clause final verb as a sentence boundary marker: the finite verb in the RC could be a strong cue for a clause boundary and the RC-final NP_{Fem} may be treated as being beyond this boundary allowing NP_{Fem} to be integrated with the upcoming string. However, this still does not explain as to why the prediction of the matrix verb due to the head NP is simply discarded. In fact, as we show in Experiment

Figure 2. Experiment 1: Reading time for all the regions in the target sentence for the 4 conditions.

5, the RC verb is not simply treated as the clause boundary and the RC-object is predicted at the RC verb.

The second possibility relates to the processing load at the RC verb and its NP_{Fem} object in the non-canonical conditions. RTs at these regions show a clear effect of word order ($t=3.43$) such that the canonical conditions were faster than the non-canonical conditions (see Figure 2). This increase in RTs in the non-canonical conditions can be explained as a revision cost due to dashed expectation of the SOV order in the RC structure (Husain et al., 2014). On this account, the head NP and its associated prediction could become less activated as more time needs to be spent to process the

RC due to the revision process at the RC verb and its object.

In order for the local coherence effect to arise, not only does the prediction of the matrix verb have to be discarded, but in addition, the structural relation of the RC object with its verb also needs to be discarded. In Experiment 5, we report a completion study that shows that the NP_{Fem} object is indeed predicted at the RC verb (also see Husain et al., 2014). This fact makes the local coherence effect even more striking. Such a pattern is however compatible with the *fallibility hypothesis* that predicts parsing errors due to increased processing load. For example, the self-organizing parsers (SOPARSE) (Tabor et al., 2004) could explain the RT patterns. As stated earlier, SOPARSE assumes a formation of local structures that compete with non-local structure. In the current experiment, in the non-canonical structure, an infrequent verb before object RC structure is formed, while the post RC material leads to a more frequent object before verb structure. For instance, in a Hindi dependency treebank (Bhatt et al., 2009), there were 8620 instances of Object before Verb order in the treebank, compared to a mere 33 instances of Verb before Object cases. This asymmetry in frequency between the word orders could lead the local structure to win over the non-local structure.

Interestingly, the acceptability rating results show a different pattern than the RT data – Canonical conditions were rated higher than the Non-canonical conditions. This seems to go against the SOPARSE account which would predict higher ratings for the locally coherent (Non-canonical) structures. One way to reconcile the RT and the rating pattern would be to assume the parser forms a locally coherent parse temporarily before realizing that it had made a mistake. However, a post-hoc analysis of the RTs in the post-critical regions showed no evidence for an effect of word order ($t=0.28$). Indeed, in Experiment 5 we report comprehension accuracy results on similar items that also do not support this possibility. The rating study seems to be capturing just an overall preference for canonical vs non-canonical order in Hindi.

To summarize, the results show a local coherence effect during online comprehension in Hindi, thus highlighting fallibility in the prediction process in a head-final language. The effect was demonstrated using an embedded RC structure; a similar structure has

been used by Vasishth and Lewis (2006) and Husain et al. (2014) to show robust prediction in Hindi. We suggested that one possible explanation for observing local coherence effect in head-final language could be reduced activation of the head noun and its associated predicted structure due to increased processing in the non-canonical conditions. But could there be certain linguistic features in the post-RC region that cause the locally coherent structure to arise? In the three experiment discussed in the following sections we explore two such features, (a) verbal agreement and (b) semantic plausibility.

Experiment 2

The goal of the current experiment was to investigate if certain linguistic factors are necessary for the local coherence observed in Experiment 1 to arise. In particular, this experiment evaluated the contribution of verb agreement which is a key device to encode syntactic relations between words in Hindi. Hindi has a mixed agreement system - verb agreement morphology can index the features of the subject or the object or no argument depending on the sentence structure. The verb agrees with the subject if the subject lacks overt case-marking; failing that the verb agrees with the object if the object lacks overt case-marking; failing that the verb exhibits default agreement morphology (Kachru, 2006; Pandharipande & Kachru, 1977). In order to probe the hypothesis that agreement constitutes a *necessary* condition for local coherence to arise we manipulated agreement morphology on the post-RC verb for transitive sentences. We compared sentences where this verb incorrectly bears overt agreement morphology indexing the features of the RC internal NP to sentences where the verb bears default agreement morphology (not indexing any NP's features) to evaluate if the presence or absence of agreement morphology impacts local coherence.

Material and methods

Participants. 63 native speakers of Hindi participated in this experiment at IIT Delhi.

Items. We used a 2×2 design crossing RC-internal WORD ORDER (Canonical/Non-canonical) and AGREEMENT-TYPE (Overt/Default). Canonical order is SOV, while non-canonical order is SVO. All four experimental conditions are ungrammatical as in Experiment 1 – the post-RC verb is thematically incompatible with the RC head noun, NP_{Masc} (*vah laRkaa* ‘That boy_{Masc}’). In the Overt Agreement conditions (7a,b) the post-RC verb incorrectly bears Feminine agreement matching the RC internal object NP_{Fem} (*kitaab* ‘book_{Fem}’) rather than the NP_{Masc}. The Overt Agreement conditions are identical to the transitive sentences in Experiment 1 (6c,d). In the Default Agreement conditions (7a,b), the post-RC verb bears default agreement morphology which does not index any NP’s features. Note that in order to allow for the possibility of a locally coherent parse in the Default agreement conditions, the RC internal object NP bears an overt case-marker *ko* which blocks agreement in the language and is consistent with default agreement on the post-RC verb.

24 experimental items in 4 conditions were prepared. 56 filler sentences with varied levels of acceptability were also included. Some of the filler sentence were ungrammatical. These ungrammatical fillers were used to filter out participants who were not paying attention during the data analysis phase. A latin-square design was used to present different items.

(7) a. Canonical, Overt

vah larka/ jisne/ kal/ bohota dilchaspī se/ kitaab/ parhi
 that boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem} read_{Fem}
 thī/ mujhe/ bechnī parhī
 had_{Fem} I_{Dat} sell_{Fem} had-to_{Fem}

b. Non-Canonical, Overt

vah larka/ jisne/ kal/ bohota dilchaspī se/ parhi thī/
 that boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem}
 kitaab/ mujhe/ **bechnī parhī**
 book_{Fem} I_{Dat} sell_{Fem} had-to_{Fem}

c. Canonical, Default

vah larka/ jisne/ kal/ bohota dilchaspī se/ kitaab ko/ parhaa
 that boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem} ACC read_{Def}
 tha/ mujhe/ bechnaa parhaa
 had_{Def} I_{Dat} sell_{Def} had-to_{Def}

d. Non-Canonical, Default

vah larka/ jisne/ kal/ bohot dilchaspī se/ parhaa thaa/
 that boy_{Masc} who_{Erg} yesterday lots interest with read_{Def} had_{Def}
 kitaab ko/ mujhe/ **bechnaa parhaa**
 book_{Fem} ACC I_{Dat} sell_{Def} had-to_{Def}

Procedure. We followed the same procedure discussed in Experiment 1.

Predictions

The predictions of the *adaptability hypothesis* will be similar to those mentioned in Experiment 1. With respect to the *fallibility hypothesis*, the SOPARSE account predicts that the presence of overt agreement would make the local coherence effect stronger. This is because the account assumes each node in an activated treelet to be a vector of semantic and syntactic features. Increased match between nodes of different treelets leads to increased strength of link between the treelets (see, Villata, Tabor, & Franck, 2018, for more details). Therefore, if agreement is a necessary condition for the building of a locally coherent parse, then we expect the local coherence effect on RTs to manifest differently across the Overt Agreement and Default Agreement conditions. The reading time difference between the Canonical and Non-canonical condition ought to be larger in the Overt Agreement conditions since the morphology on the post-RC verb overtly indexes the features of the RC-internal object NP here but not in the Default Agreement conditions.

Results

We follow the same statistical analysis procedures as discussed in Experiment 1. In addition, participants who rated ungrammatical fillers more than 3.5 were removed from analysis. The threshold of 3.5 was chosen as it was the value for the 3rd quartile of all ungrammatical ratings. This step led to the removal of data from 17 participants in this experiment.

The average rating for the experimental conditions were found to be significantly higher than the rating for the ungrammatical filler sentences ($t=-12.51$). The mean rating for the experimental items was 3.6 (Min=1, 1st Quartile=2, Median=4, 3rd Quartile=5,

Max=7) while the mean rating for the ungrammatical fillers was 2.5 (Min=1, 1st Quartile=1, Median=2, 3rd Quartile=4, Max=7). The analysis of the acceptability ratings found a marginal effect of Word Order ($t=-1.74$) such that the ratings for the non-canonical conditions were less than the ratings for the canonical condition. In addition, a significant effect was found for Agreement ($t=-2.26$) such that Overt agreement conditions were rated higher than the Default agreement conditions. These figures can be seen in Tables 4, 5.

Table 4

Experiment 2: Average acceptability ratings for the four conditions.

Condition	Rating
a. Canonical, Overt	4.0
b. Non-Canonical, Overt	3.7
c. Canonical, Default	3.5
d. Non-Canonical, Default	3.4

Table 5

Experiment 2: lmer results for the acceptability ratings. Anova sum contrasts were used for the two main effects and the interaction.

	Coefficient	SE	t-value
Intercept	0.004	0.09	0.04
Word Order	-0.04	0.02	-1.74
Agreement	-0.09	0.04	-2.26
Word Order:Agreement	0.02	0.02	0.78

RTs at the critical and post-critical region showed a trend for faster RTs in the Non-canonical conditions compared to the Canonical conditions; however, this difference was not statistically significant. The effects of Agreement and the interaction term were found to be non-significant. Table 6 shows the results for both critical and the post-critical region. RTs at the critical and the post-critical regions are shown in

Figure 3. Figure 4 shows the RTs for all the regions across all the experimental conditions.

We also note that not removing the participants based on the exclusion criterion mentioned above did change the results in the following ways – for the acceptability ratings there was a significant effect of Word Order ($t=2.5$), in addition to the main effect of Agreement ($t=-3.6$). For the RT results at the critical region, the effect of Word Order was marginally significant ($t=-1.8$) with the Non-canonical conditions being read faster compared to the Canonical conditions. For this unfiltered data, the average rating for the experimental conditions were still found to be significantly higher than the rating for the ungrammatical filler sentences ($t=-10.35$).

Table 6

Experiment 2: lmer results for the RTs at the critical and post-critical regions. Anova sum contrasts were used for the two main effects and the interaction.

	Critical Region			Post-critical region		
	Coefficient	SE	t-value	Coefficient	SE	t-value
Intercept	6.71	0.05	123.90	6.44	0.03	164.28
Word Order	-0.01	0.01	-1.01	-0.01	0.01	-1.35
Agreement	-0.009	0.01	-0.70	0.007	0.01	0.61
Word Order:Agreement	-0.007	0.01	-0.46	-0.007	0.01	-0.55

Discussion

The RT results and the acceptability ratings results in the current experiment show similar numerical trends to Experiment 1 with regard to the Word Order manipulation. However, the effect of Word Order was non-significant for both dependent variables. A main effect of Agreement was also observed in the acceptability data.

The lack of a differential impact of overt agreement morphology versus default agreement morphology on the reading times suggest a number of possibilities. One, it could be the case that this morphological difference is not particularly contrastive to

Figure 3. Experiment 2: Reading time at the critical and the post-critical region in the 4 conditions. C.Overt=Canonical, Overt Agreement; NC.Overt=Non-canonical, Overt Agreement; C.Default=Canonical, Default Agreement; NC.Default=Non-canonical, Default Agreement.

Hindi speakers who may be treating the Overt Agreement conditions at par with the Default Agreement conditions in terms of agreement evaluation since the locally coherent parse in both the conditions is consistent the agreement grammar of the language. Alternatively, it could be the case that the structure may not be processed deeply enough to allow agreement constraints to be evaluated immediately.

Given the effect of word order in Experiment 1, and a similar trend observed in the current experiment, but no observable effect of agreement on RTs, it is possible that overt agreement does not form a necessary condition for local coherence to arise.

A further possibility is that while overt agreement on the post-RC verb is not a necessary condition for local coherence, overt agreement elsewhere in the sentence, for example on the RC-internal verb, impacts local coherence effects. In the next section,

Figure 4. Experiment 2: Reading time for all the regions in the target sentence for the 4 conditions. C.Overt=Canonical, Overt Agreement; NC.Overt=Non-canonical, Overt Agreement; C.Default=Canonical, Default Agreement; NC.Default=Non-canonical, Default Agreement.

we present a sentence completion study which investigates this possibility.

Experiment 3

This section presents a sentence completion study to investigate the errors after the RC verb in the two non-canonical conditions. While agreement morphology on the post-RC verb, as manipulated in Experiment 2, does not appear to impact local coherence, one further test of the contribution of agreement would be at the RC-verb. We evaluated participant performance in a sentence completion task involving Non-canonical sentence preambles truncated at the RC-verb with Agreement type manipulated at this verb. The linking hypothesis is that the more susceptible participants are to locally coherent parses after encountering the RC verb, the more likely they are to make completion errors. If participants make similar amounts of errors after seeing Overt agreement and Default agreement morphology on the RC-verb in Non-Canonical sentence preambles then it can be concluded that they will be equally susceptible to the local coherence effect irrespective of the agreement morphology at the RC-verb. On the other hand, if agreement morphology is contrastive for Hindi speakers here, then a differential error

rate should be observed, which would suggest that the susceptibility to building a locally coherent parse rests on agreement morphology on the RC-verb.

Material and methods

Participants. 21 Hindi native speaker participated in the experiment. The participants were undergraduate or graduate students at Jawaharlal Nehru University Delhi and IIT Delhi.

Items. We used a 1×2 design manipulating AGREEMENT-TYPE (Overt/Default) at the RC-internal verb in sentences with non-canonical word order in the relative clause. Participant were shown 24 items following example 8. The items were presented using a latin square design. In addition, 122 filler items were also included in the experiment.

(8) a. Non-Canonical, Overt

vah larka/ jisne/ kal/ bohoh dilchaspī se/ parhī thī ...
that boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem} ...

b. Non-Canonical, Default

vah larka/ jisne/ kal/ bohoh dilchaspī se/ parhaa tha ...
that boy_{Masc} who_{Erg} yesterday lots interest with read_{Def} had_{Def} ...

Procedure. The sentence completion task was employed as the experimental paradigm (Taylor, 1953). Each sentence was presented using the centered self-paced reading (SPR) paradigm. Participants were provided an incomplete sentence and their task was to complete it such that it was meaningful. Each sentence appeared on the screen in the self-paced reading format. Initially, a ‘+’ sign appeared on the center of the computer screen. When the participant pressed the space-bar key this ‘+’ sign got replaced with the first word of the sentence. Successive button presses displayed the remaining words of the sentence at the center of the screen. A ... symbol signalled the end of the sentence fragment and prompted the participant to complete the sentence. This was done in a text box that appeared by pressing ‘space-bar’ after the ... symbol. After typing in the text the participants pressed the ‘enter’ key to move to the next trial. The experiment was conducted using Douglas Rohde’s Linger software (version 2.94). Items were automatically randomized by Linger.

Response coding

The completion data was coded for two variables (a) whether the direct object of the RC verb was posited (1 if yes, 0 otherwise), and (b) whether the completion was grammatical (1 for yes, 0 otherwise).

Predictions

As stated previously, the aim of the current completion study was to investigate if the susceptibility to form locally coherent structures is affected by the agreement morphology at the RC-verb.

Note that given the SV order in the relative clause, speakers ought to predict an object argument for the RC-verb, and that this step would precede the building of any locally coherent parses. Object prediction may have been directly affected by the agreement morphology on the RC-verb – Overt agreement requires the positing of a case-unmarked NP with the relevant person/number/gender features. On the other hand Default agreement may not be as strongly predictive because in the language default agreement can arise in a number of contexts – when argument NPs are case-marked or when there are no argument NPs present (e.g. in infinitive clauses). Therefore, completions with a valid RC-object are predicted to be higher in the Overt condition compared to the Default condition.

In order to form the locally coherent parse the prediction of the head NP-matrix verb structure has to be discarded by the parser. By hypothesis, increased completion errors in this experiment constitute a signal of the discarding of this prediction and can be used as a measure of the strength of local coherence in the two conditions.

If the local coherence effect is impacted by agreement morphology at the RC-verb we should expect a differential rate of ungrammatical completions in the Overt condition compared to the Default condition. It may be the case that in the Overt agreement condition, the RC internal structure is processed more deeply compared to the Default agreement condition, leading to reduced susceptibility to a locally coherent parse in sentence completion. On the other hand, if the susceptibility to form a locally coherent

parse is not influenced by agreement morphology, we would see no difference in the number of valid grammatical completions in the two conditions.

Results

The statistical analysis for the completion data was done using the generalized linear mixed-effects model with logit link function. This has been done using the `lme4` package (Bates et al., 2015) in R. Maximal models were fit when possible (Barr et al., 2013); in case of convergence failure, a less complex model was fit by successively removing the random slopes of the by-subject and by-item random effects component. The percentage completion for the two response variables coded by us is shown in Table 7. Results show no difference in the completions either for the RC object or for grammatical completions.

Table 7

Experiment 3: Percentage completions for valid RC object and valid grammatical continuations in the two conditions.

Condition	RC object	Grammatical completions
Non-Canonical, Overt	62	57
Non-Canonical, Default	57	59

Table 8

Experiment 3: glmer results for the completion data. Treatment contrast was used with Non-Canonical, Default as the baseline condition.

	RC object			Grammatical completion		
	Coefficient	SE	z-value	Coefficient	SE	z-value
Intercept	0.45	0.71	0.64	0.51	0.43	1.19
Non-Canonical, Overt	0.52	0.45	1.17	-0.10	0.28	-0.38

Discussion

These results show that completions in sentences with overt agreement features on the RC verb do not differ from cases where there were default agreement features on the RC verb. Hindi speakers are just as likely to (a) posit an RC-object in the Overt agreement condition as the Default agreement condition; and (b) grammatically complete the sentences in both the conditions. While there may be some generalized susceptibility to locally coherent parses in these sentences – as signalled by the relatively low proportion of grammatical completions across the board – overt agreement at the RC-verb does not modulate this effect.

Experiment 4

The aim of the current experiment was to evaluate if plausibility underlies the local coherence effect observed in Experiment 1. Words in a structure are in a semantic relationship with one another and plausibility is a key device to encode the well-formedness of these semantic relations. In this experiment, we specifically ask if local coherence arises only when the RC-internal object is a *plausible* argument of the post-RC verb.

Material and methods

Participants. 63 native speakers of Hindi participated in this experiment at IIT Delhi. Please note that although, the number of participants in this experiment and Experiment 2 is same, these experiments were conducted separately.

Items. We used a 2×2 design crossing RC-internal WORD ORDER (Canonical/Non-canonical) and PLAUSIBILITY (Plausible/Implausible). All experimental conditions are ungrammatical as in Experiment 1. As before, Canonical order is SOV, and Non-canonical SVO. In the Plausible conditions, the RC-internal NP-object is a semantically plausible argument of the post-RC verb, while in the Implausible conditions, NP-object is not a plausible argument of the post-RC verb: for example, in the sample item given in 9, while the sale of a book is a plausible event, the

sale of an eye is less so.

- (9) a. Canonical, Plausible (= 6c / 7a)

vah larka/ jisne/ kal/ bohoh dilchaspai se/ kitaab/ parhi
 that boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem} read_{Fem}
 thii/ mujhe/ bechnii parhi
 had_{Fem} I_{Dat} sell_{Fem} had-to_{Fem}

- b. Non-Canonical, Plausible (= 6d / 7b)

vah larka/ jisne/ kal/ bohoh dilchaspai se/ parhi thii/
 that boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem}
 kitaab/ mujhe/ **bechnii parhi**
 book_{Fem} I_{Dat} sell_{Fem} had-to_{Fem}

- c. Canonical, Implausible

vah larka/ jisne/ kal/ bohoh dilchaspai se/ aankh/ dekhii
 that boy_{Masc} who_{Erg} yesterday lots interest with eye_{Fem} seen_{Fem}
 thii/ mujhe/ bechnii parhi
 had_{Fem} I_{Dat} sell_{Fem} had-to_{Fem}

- d. Non-Canonical, Implausible

vah larka/ jisne/ kal/ bohoh dilchaspai se/ dekhii thii/
 that boy_{Masc} who_{Erg} yesterday lots interest with seen_{Fem} had_{Fem}
 aankh/ mujhe/ **bechnii parhi**
 eye_{Fem} I_{Dat} sell_{Fem} had-to_{Fem}

24 experimental items in 4 conditions were prepared. 56 filler sentences with varied levels of acceptability were also included.

Procedure. We followed the same procedure discussed in Experiment 1.

Predictions

The predictions for the *adaptability hypothesis* for the items in current experiment are similar to those mentioned in Experiment 1.

With respect to the *fallibility hypothesis*, both the SOPARSE account as well as the good-enough processing account would predict that for local coherence to arise, semantic compatibility should be a necessary condition. In the SOPARSE account, semantic match is a precondition for forming links between nodes of various treelets since the treelets are assumed to be vectors of semantic and syntactic features (see, Villata et al., 2018, for more details). Similarly, the good-enough account assumes use

of various heuristics during the comprehension process. According to this account, heuristics based on semantics and world knowledge can override the grammatical constraints during online structure building (Ferreira, 2003). So both these accounts predict semantic plausibility to be a necessary prerequisite for local coherence to arise, as plausibility is a key device to encode semantic relations between words. If this is correct then, we would expect to observe a faster reading time in the Plausible Non-canonical condition compared to the Plausible Canonical condition. No difference in RT should be observed in the two Implausible conditions.

Results

We follow the same statistical analysis procedures as discussed in Experiment 1. In addition, participants who rated ungrammatical fillers more than 4.1 were removed from analysis. The threshold of 4.1 was chosen as it was the value for the 3rd quartile of all ungrammatical ratings. This step led to the removal of 16 participants.

The average rating for the experimental conditions were found to be marginally higher than the rating for the ungrammatical filler sentences ($t=-1.8$). The mean rating for the experimental items was 3.1 (Min=1, 1st Quartile=1, Median=2, 3rd Quartile=5, Max=7) while the mean rating for the ungrammatical fillers was 2.9 (Min=1, 1st Quartile=1, Median=2, 3rd Quartile=5, Max=7). The analysis of the acceptability ratings found a significant effect of Word Order ($t=-2.69$) such that the ratings for the non-canonical conditions were less than the ratings for the canonical condition. A significant effect of Plausibility ($t=8.27$) was also observed such that plausible conditions were rated higher than the implausible conditions. These figures can be seen in Tables 9, 10. In addition, a nested contrast showed that the ratings in the Plausible conditions were marginally different ($z=-1.9$) while the rating in the Implausible conditions were not ($z=-1.6$).

RTs at the critical region showed a significant main effect of Word Order ($t=-2.83$) such that RTs in the Non-canonical conditions were faster than the Canonical conditions. See Table 11 for the RT results for the critical region. A nested contrast showed that

Table 9

Experiment 4: Average acceptability ratings for the four conditions.

Condition	Rating
a. Canonical, Plausible	4.06
b. Non-Canonical, Plausible	3.7
c. Canonical, Implausible	2.41
d. Non-Canonical, Implausible	2.1

Table 10

Experiment 4: lmer results for the acceptability ratings. Anova sum contrasts were used for the two main effects and the interaction.

	Coefficient	SE	t-value
Intercept	-0.0008	0.07	-0.01
Word Order	-0.06	0.02	-2.69
Plausibility	0.38	0.04	8.27
Word Order:Plausibility	-0.01	0.02	-0.38

this effect was driven by the Plausible conditions ($t=-3.49$) – the Plausible non-canonical condition was faster than the Plausible canonical condition while no difference was found in the Implausible conditions ($t=-0.98$). See Table 12 for the results of the nested contrast. No significant effect was observed in the post-critical region. RTs at the critical and the post-critical regions are shown in Figure 5. Figure 6 shows the RTs for all the regions across all the experimental conditions. We also note that not removing the participants based on the exclusion criterion mentioned above did not change the salient findings of the presented analyses.

Discussion

The results from Experiment 4 show that the local coherence effect appeared in the non-canonical condition only when the post-RC object and the clause final verb and its

Table 11

Experiment 4: lmer results for the RTs at the critical region. Anova sum contrasts were used for the two main effects and the interaction.

	Coefficient	SE	t-value
Intercept	6.67	0.08	79.94
Word Order	-0.05	0.01	-2.83
Plausibility	-0.01	0.02	-0.45
Word Order:Plausibility	-0.02	0.02	-1.07

Table 12

Experiment 4: lmer results for the RTs at the critical region. Anova sum contrasts were used for the nested comparisons.

	Coefficient	SE	t-value
Intercept	6.67	0.08	79.94
Plausibility	-0.01	0.02	-0.45
Implausible.Order	-0.03	0.03	-0.98
Plausible.Order	-0.07	0.02	-3.49

subject formed a plausible parse. It also showed that while the implausible conditions were rated lower than the plausible conditions, the two implausible conditions were rated similarly.

The results replicate the key result of the speed-up in the non-canonical word order found in Experiments 1. It shows that semantic plausibility is a key precondition for the local coherence effect to arise. With regard to the processes assumed by SOPARSE and good-enough processing accounts, the results show that syntactic compatibility alone cannot allow a local parse to override competing parses and that semantic coherence is critical.

So far we have observed that a local coherence effect arises in a head-final language like Hindi. We noted that agreement is not a necessary precondition for this effect to arise,

Figure 5. Experiment 4: Reading time at the critical and the post-critical region in the 4 conditions. C.Plaus=Canonical, Plausible; NC.Plaus=Non-canonical, Plausible; C.Implaus=Canonical, Implausible; NC.Implaus=Non-canonical, Implausible.

while semantic plausibility is. At the same time, the acceptability ratings clearly show that while the experimental items are rated higher than the ungrammatical fillers, they are also consistently rated worse than the grammatical fillers. This suggests that the purported illicit parse (that shows up as facilitation in reading time) in the non-canonical conditions could be short-lived. Given the SPR+acceptability rating paradigm employed so far it is difficult to say if the locally coherent parse persists after the participant has comprehended the sentence. In order to assess this we also conducted an SPR experiment where comprehension questions probe dependencies formed between verbs and nouns in the sentences during the comprehension process.

Figure 6. Experiment 4: Reading time for all the regions in the target sentence for the 4 conditions. C.Plaus=Canonical, Plausible; NC.Plaus=Non-canonical, Plausible; C.Implaus=Canonical, Implausible; NC.Implaus=Non-canonical, Implausible.

Experiment 5

The aim of the current experiment was to probe whether Hindi speakers are able to successfully form the various structural dependencies in the sentences presented in previous experiments by utilizing targeted comprehension questions. We presented Hindi speakers with copular verb sentences in a self-paced reading paradigm which was followed by comprehension questions targeting the integration of NPs in the sentences. Our primary focus was the integration of the RC-internal object NP – we tested a licit integration (RC object NP – RC verb) and an illicit integration (RC object NP – matrix verb) to evaluate whether faulty NP integrations underlie the observed local coherence effects – but we also tested the integration of the head NP. Furthermore, in addition to ungrammatical copular sentences, we also tested grammatical copular sentences to establish a baseline for local coherence effects.

Material and methods

Participants. 46 native speakers of Hindi participated in this experiment. While the planned number of participants was 70, data collection had to be stopped partway due

to COVID related disruptions.

Items. We used a 2×2 within-subjects repeated measures design crossing RC-internal Word order (Canonical/Non-canonical) and Grammaticality (Grammatical/Ungrammatical). Canonical order is SOV (9a,c), while Non-canonical order is SVO (9b,d). Grammaticality is manipulated via the RC head NP and the relationship between this NP and the agreement morphology on the post RC verb – distinct NPs with different gender specifications were used in the grammatical and ungrammatical conditions. In (10), the ungrammatical conditions have a masculine NP as the head NP (*vah laRkaa* ‘That boy_{Masc}’) while the grammatical conditions have a feminine NP as the head NP (*vah raajkumaarii* ‘That princess_{Fem}’). The post-RC verb bears the same agreement morphology across all conditions of an item, in 10 feminine agreement. In the ungrammatical conditions (10a,b), this feminine agreement morphology on the post-RC verb does not correspond to the features of the grammatical target of agreement, the RC head NP (*vah laRkaa* ‘That boy_{Masc}’) but rather to the features of the RC internal object (*kitaab* ‘book_{Fem}’), as in Experiments 1. In the grammatical conditions (10c,d), the feminine agreement morphology on the post-RC verb correctly matches the RC head NP (*vah raajkumaarii* ‘That princess_{Fem}’).

(10) a. Canonical, Ungrammatical

vah laRkaa *jisne* *kal* *bahut dilchaspai se kitaab*
 That boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem}
paRhii thii *shaayad* **moTii thii** ...
read_{Fem} had_{Fem} maybe **fat_{Fem} was_{Fem}** ...

b. Non-canonical, Ungrammatical

Vah laRkaa *jisne* *kal* *bahut dilchaspai se paRhii thii*
 That boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem}
kitaab shaayad **moTii thii** ...
book_{Fem} maybe **fat_{Fem} was_{Fem}** ...

c. Canonical, Grammatical

vah raajkumaarii *jisne* *kal* *bahut dilchaspai se kitaab*
 That princess_{Fem} who_{Erg} yesterday lots interest with book_{Fem}
paRhii thii *shaayad* **moTii thii** ...
read_{Fem} had_{Fem} maybe **fat_{Fem} was_{Fem}** ...

d. Non-canonical, Grammatical

vah raajkumaarii jisne kal bahut dilchaspai se paRhii thii
 That princess_{Fem} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem}
 kitaab shaayad **moTii thii** ...
book_{Fem} maybe **fat**_{Fem} **was**_{Fem} ...

24 experimental items in 4 conditions were prepared. Every experimental trial was followed by a Yes-No comprehension question. 88 filler sentences were also included. The filler sentences included both monoclausal and biclausal sentences in canonical and non-canonical word orders. The complete list of experimental items can be obtained from the authors. A latin-square design was used to present different items. Overall, comprehension questions were included for 70% of trials.

Comprehension Questions

The comprehension questions for the critical experimental items sought to evaluate the success of various NP integrations in the sentences. Question type was manipulated as a between items factor to give three types of question-item pairs probing distinct NP integrations.

The first two subsets of questions are concerned with the integration of the RC-internal NP and are our primary focus. Of these two subsets, one subset tested the integration of the RC internal object NP with the RC internal material. This is a grammatically licit dependency since both these elements are part of the the relative clause. To illustrate, if the sample item in (10) was part of the subset probing RC internal NP-RC internal material integration, then the corresponding question would be as in (11). 8 out of 24 items in the experiment had this kind of comprehension question.

(11) RC internal NP-RC internal integration

kyaa kitaab kal paRhii gayii thii?
 Q book yesterday read go was?
 ‘Was the book read yesterday?’

The second subset tested the integration of the RC internal NP with the post-RC material. Since the RC internal object NP and the post RC material are from distinct clauses and the RC internal object NP is not an argument of the post RC verb, this is a

grammatically illicit dependency. If the sample item in (10) was part of the subset probing RC internal NP-post RC integration, then the question would be as in (12). 8 out of 24 items in the experiment had this kind of comprehension question.

(12) RC internal NP-post RC integration

kyaa kitaab moTii thii?

Q book fat was

‘Was the book thick?’

In addition, we also had a third subset of questions which probed the integration of the head NP with distinct parts of the structure. Distinct questions had to be used in the grammatical and ungrammatical conditions due to the non-identical head NP across these conditions. In the grammatical conditions, the questions probed the integration of the head NP with the post-RC material. Since both the head NP and the post RC material are part of the matrix clause structure, this integration is licit. However, in the ungrammatical conditions we probed the integration of the head NP with the RC internal verb since the integration with the post RC verb is illicit here. If the sample item in (10) was part of the subset probing head NP-RC integration/post RC integration, then the questions would be as in (13).

(13) head NP integrations

a. kyaa laRke ne kuchh paRhaa thaa?

Q boy erg something read was

‘Did the boy read something?’

b. kyaa raajkumaarii motii thii?

Q princess_{FeM} fat was

‘Was the princess fat?’

Note that while we illustrate the various question types for a single item here (10), in the experiment any one item corresponded to only one question type subset for all participants.

Procedure

We followed the same procedure discussed in Experiment 1. However, instead of ratings, the SPR reading was followed by yes-no comprehension questions. The experiment was conducted using Douglas Rohde's Linger software (version 2.94).

Predictions

We expected to replicate the reading time pattern observed in previous experiments, wherein Word Order has an impact on RTs. If the RC internal NP gets integrated with the post-RC material, then the RTs for the Non-canonical condition should be lower than the Canonical condition as in the previous experiments.

Turning to question comprehension accuracy, the primary predictions are regarding the integration of the RC-internal NP. If Hindi speakers are successfully integrating the RC internal object NP with the RC internal material, then they should have high accuracy on questions probing this dependency. Furthermore, since the RC internal object NP and the post RC material are from distinct clauses, Hindi speakers ought not to integrate these elements. If Hindi speakers do not attempt such an integration, then they should exhibit high accuracy on questions targeting this dependency. However, if Hindi speakers erroneously attempt to integrate the RC internal NP with the post RC material then they are expected to exhibit low accuracy on questions targeting this dependency.

In addition, a more general point can be made regarding the integration of the head NP – if Hindi speakers are able to make a successful prediction about upcoming matrix material at the head NP and maintain that prediction, then the integration of the head NP with the matrix material ought to proceed smoothly. However, if they fail to maintain that prediction they should subsequently have difficulty in answering comprehension questions about this integration which would manifest as low accuracy in the grammatical conditions of the head NP-integration subset.

Results

We follow the same statistical analysis procedures as discussed in Experiment 1. Data from participants with less than 70% accuracy across all the trials (including fillers) were excluded from analysis. This led to the removal of one participant's data.

RTs at the critical region showed a significant effect for Grammaticality ($t=-2.07$) such that RTs in the Grammatical condition were faster than the Ungrammatical condition.

Table 13 shows the RT results for critical region. No significant effect was observed in the post-critical region. There was a numerical trend in the post-critical2 region consistent with the effect of Word order observed in previous experiments –

Non-canonical conditions were marginally faster than the Canonical conditions ($t=-1.945$). Table 14 shows the RT results for the post-critical2 region. RTs at the critical and the post-critical2 regions are shown in Figure 7. Figure 8 shows the RTs for all the regions across all the experimental conditions.

Table 13

Experiment 5: lmer results for the RTs at the critical region. Anova sum contrasts were used for the two main effects and the interaction.

	Coefficient	SE	t-value
Intercept	6.425	0.053	120.778
Word Order	0.018	0.012	1.478
Grammaticality	-0.026	0.012	-2.070
Word Order:Grammaticality	0.006	0.012	0.496

Comprehension accuracy results are given in Table 15. The results suggest that Hindi speakers are fairly accurate at answering questions probing the integration of the RC-internal NP with the RC internal material. At the same time, they struggle with questions probing the integration of the RC internal NP with the post RC material. The accuracy data for the two question types probing the RC internal NP was analyzed using generalized linear mixed effects model. There is a significant main effect of Question Type ($p < 0.001$). This shows that Hindi speakers are quite susceptible to the

Figure 7. Experiment 5: Reading time at the critical and the post-critical2 region in the 4 conditions. C.Ungram=Canonical, Ungrammatical; NC.Ungram=Non-canonical, Grammatical; C.Gram=Canonical, Ungrammatical; NC.Gram=Non-canonical, Grammatical

Table 14

Experiment 5: lmer results for the RTs at the post critical 2 region. Anova sum contrasts were used for the two main effects and the interaction.

	Coefficient	SE	t-value
Intercept	6.308	0.04	156.385
Word Order	-0.02	0.01	-1.945
Grammaticality	0.003	0.01	0.309
Word Order:Grammaticality	-0.004	0.01	-0.409

incorrect integration of the RC-NP with post-RC material.

In addition, we evaluated accuracy on comprehension questions targeting the

Figure 8. Experiment 5: Reading time for all the regions in the target sentence for the 4 conditions. C.Ungram=Canonical, Ungrammatical; NC.Ungram=Non-canonical, Grammatical; C.Gram=Canonical, Ungrammatical; NC.Gram=Non-canonical, Grammatical.

Table 15

Experiment 5: Comprehension accuracy for questions

Question Type	Accuracy %
RC internal NP – RC Verb	75
RC internal NP – post RC material	26

Table 16

Experiment 5: lmer results for comprehension accuracy for questions probing integration of RC-internal NP. An anova sum contrast was used.

	Coefficient	SE	z-value	p-value
Intercept	0.001	.32	0.003	0.997
Question Type	1.436	.316	4.551	< 0.001

integration of the head NP with post-RC matrix material in grammatical conditions separately. Hindi speakers appear to struggle with this integration – speakers’ accuracy is close to chance levels here (53%). Finally, we note that not removing the participants

based on the exclusion criterion mentioned above did not change the salient findings of the presented statistical analyses.

Discussion

The comprehension question accuracy results from this experiment allow us to establish that Hindi speakers are forming faulty integrations during structure building in sentences with center embedded relative clauses. Accuracy on questions targeting the RC internal NP suggests that licit integrations and illicit integrations coexist – Hindi speakers integrate the RC object NP with RC internal material and they also integrate RC object NP with post RC material. Alongside this susceptibility to incorrect integrations, we also observed that Hindi speakers struggle with the integration of the head NP with the post-RC material. The near chance performance on questions targeting this integration suggests that Hindi speakers failed to adequately maintain the predictions associated with this dependency.

The reading times results in this experiment are consistent with the key result of speed up in non-canonical word order observed in previous experiments, although the effect is only marginally significant in the present study. We also observe a main effect of grammaticality in the reading times at the critical region – Hindi speakers are faster at reading grammatical sentences than ungrammatical sentences. The existence of this grammaticality effect in RTs alongside a susceptibility to building locally coherent parses (as suggested by the comprehension question accuracies) may appear puzzling at first glance. However, this can be treated as an ambiguity advantage effect explicable in terms of underspecification (Swets, Desmet, Clifton, & Ferreira, 2008). In the Grammatical conditions, where the (sub-)strings associated with the global parse (head NP integration with post RC verb) and the local parse (RC-NP integration with post RC verb) are well-formed, speakers do not need to commit to a particular parse at the post-RC verb in the absence of disambiguating information. In contrast, in ungrammatical sentences, the global parse may be ruled out immediately upon encountering the post-RC verb which signals the ungrammaticality, and only the local

parse remains viable.

In sum, this experiment shows that Hindi speakers are struggling with the various NP integrations in sentences with center-embedded relative clauses. This may be due to failure to maintain various details of predictive structure building during incremental processing of head final structures with clausal embedding. In order to further assess the nature of predictive processes, we conducted a sentence completion experiment comparing canonical and non-canonical word orders.

Experiment 6

The goal of the current experiment was to investigate if Hindi speakers are successfully predicting upcoming structure in sentences with embedded relative clauses and if prediction is modulated by relative clause internal word order in such cases. In order to investigate this we conducted a sentence completion study.

Material and methods

Participants. 36 Hindi native speakers participated in the experiment but data from one participant had to be excluded due to no responses on nearly all trials.

Items. We used a 1×2 design manipulating RC-internal WORD ORDER (Canonical/Non-canonical) with the sentence preambles truncated at the RC-internal verb in both cases. In the Canonical conditions the RC internal order was SOV, but in the Non-canonical conditions, no object was present and the order was SV.

Participants were shown 24 items following example (14). Items presented using a latin square design. In addition 52 filler items were also shown.

(14) a. Canonical

vah larka/ jisne/ kal/ bohoh dilchaspī se/ kitaab/ parhī
that boy_{Masc} who_{Erg} yesterday lots interest with book_{Fem} read_{Fem}
thī ...
had_{Fem} ...

b. Non-Canonical

vah larka/ jisne/ kal/ bohoh dilchaspī se/ parhī thī ...
that boy_{Masc} who_{Erg} yesterday lots interest with read_{Fem} had_{Fem} ...

Procedure. The sentence completion task was employed as the experimental paradigm in this experiment in a manner identical to that outlined in Experiment 3. The experiment was conducted using Douglas Rohde’s Linger software (version 2.94). Items were automatically randomized by Linger.

Response coding

The completion data was coded for two variables (a) whether the completion was grammatical (1 for yes, 0 otherwise), and (b) whether the direct object of the RC verb was posited (1 for yes, 0 otherwise).

Results

The statistical analysis for the completion data was done using the generalized linear mixed-effects model with logit link function. This has been done using the `lme4` package (Bates et al., 2015) in R. Maximal models were fit when possible (Barr et al., 2013); in case of convergence failure, a less complex model was fit by successively removing the random slopes of the by-subject and by-item random effects component. The percentage completion for the two variables is shown in Table 17. A glmer analysis showed that grammatical completions in the Canonical condition were significantly different from the Non-canonical condition ($z=4.06$). The glmer analysis is provided in Table 18. In addition, note that the RC object was posited frequently in the Non-canonical condition while it was not predicted in the Canonical condition.

Table 17

Experiment 6: Percentage completions

Condition	Grammatical	RC object completions
Canonical	71	0
Non-Canonical	42	73

Table 18

*Experiment 6: glmer results for the completion data. Treatment contrast was used with the Non-canonical condition as the baseline. * signifies significant at $\alpha < 0.05$.*

	Grammatical completion		
	Coefficient	SE	z-value
Intercept	-0.87	0.56	-1.55
Canonical	2.13	0.52	4.06 *

Discussion

The results of the completion study show that the head NP and its prediction is frequently forgotten during prediction in the structures used in the reported experiment. This is signalled by the proportion of ungrammatical completions in the study. The results also show that the relative clause internal prediction is fine - the direct object of the RC verb in the Non-canonical condition is predicted frequently. These findings support the comprehension accuracy patterns observed in the previous experiment. The results suggests that the parser is vulnerable to making errors in the Post-RC region, specially in the Non-canonical condition. It is therefore likely that in such situations the parser could commit to a locally coherent parse.

In Experiment 1, we entertained two hypothesis regarding the manifestation of a locally coherent parse in the post-RC region. One of the hypothesis was that the RC verb could be treated as clause boundary and therefore in the Non-canonical order, any material following the RC verb (i.e. the RC object here) could be treated as belonging to a new clause. The results of the completion study suggests that this verb-as-boundary hypothesis is incorrect. The other hypothesis that we entertained was that there is a revision cost at the RC verb - RC object region in the Non-canonical condition. In light of the completion data and the RT results where Hindi speakers frequently posit an object NP after the RC verb, it is more likely that the NP head and its associated predictions are forgotten in the Non-canonical condition owing to increased processing cost due to this revision. This suggests that simultaneously

maintaining the prediction about the RC-internal object and the matrix verb in the non-canonical sentences proves to be difficult for Hindi speakers.

General Discussion

The key results of the series of experiments on the processing and production of sentences with word order variations in embedded relative clauses in Hindi were as follows. In Experiment 1, Hindi speakers were faster at reading the post-RC material when the RC-internal word order was non-canonical (VO) compared to when it was canonical (OV). This suggests that they are susceptible to a local coherence effect where the RC-object in the non-canonical order and post-RC verb are part of a grammatically illicit but locally coherent parse. In end-of-sentence acceptability ratings, sentences with canonical word order were rated better than sentences with non-canonical word order consistent with the generalized difficulty observed for non-canonical order in the absence of licensing discourse contexts in languages allowing flexible word order (Kaiser & Trueswell, 2004).

In Experiments 2 and 3, we tested the role of syntactic dependencies, here verb agreement, in giving rise to a locally coherent parse. In Experiment 2, where we manipulated agreement on the post-RC verb in addition to word order, we observed a marginal effect of word order on RTs and on ratings consistent with Experiment 1. In addition, an effect of agreement-type was observed on the ratings – sentences with overt agreement were rated better than sentences with default agreement – but not on the RTs. Experiment 3 was a sentence completion study where we manipulated agreement on the RC-verb in non-canonical sentence preambles truncated at the RC-verb. Hindi speakers were observed to be just as likely to grammatically complete the sentences irrespective of the agreement-type on the RC-verb. Together these two experiments suggest that the building of grammatically illicit locally coherent parses doesn't *necessarily* rest on agreement dependencies.

In Experiment 4, we tested the role of semantic information, here plausibility, in giving rise to a locally coherent parse. We manipulated whether the RC-object could be a

plausible argument of the post-RC verb, in addition to manipulating word order. We observed a significant effect of word order on RTs consistent with Experiment 1. Crucially this effect was driven by the plausible conditions pointing to the role of semantic plausibility in the building of locally coherent parses. In addition, an effect of plausibility was also observed in ratings: sentences where the RC-object was a plausible argument of the post-RC verb were rated better. We also replicated the effect of word-order on the ratings observed in Experiment 1: here too sentences with canonical word order were rated better than sentences with non-canonical word order.

Experiment 5 probed the building of structural dependencies by utilizing targeted comprehension questions. We observed that while Hindi speakers were able to successfully integrate the RC-object with the RC-verb, they struggled with inhibiting the ungrammatical integration between the RC-object and the post-RC material. In addition, integrations involving the head-NP and post-RC material also seemed to be non-trivial for Hindi speakers. Finally, in Experiment 6 we used a sentence completion task to evaluate the nature of predictive processing. We observed that Hindi speakers struggled with grammatical sentence completions in sentences with non-canonical word order pointing to the forgetting of the (predictions of the) head-NP in such structures. Together Experiments 5 and 6 suggest that the parser is subject to significant processing difficulty during structure-building in sentences with clausal embeddings, particularly in non-canonical orders, even in head-final languages.

In summary, through a series of online and offline experiments, this work demonstrated the local coherence effect in Hindi. We highlighted the possible reasons for the parser becoming susceptible to forming a locally coherent parse. In the context of the RC construction used in various experiments, we used multiple methods to show that parsing errors take place in sentences with non-canonical SVO order where there is a higher processing cost due to revision in the RC-verb RC-object region. Under such conditions, the susceptibility to form locally coherent parse increases when there is a semantic compatibility between the RC object and the matrix verb. To our knowledge, this is the first demonstration of local coherence in an SOV language during online

sentence comprehension.

Is the locally coherent structure formed bottom-up or top-down?

The evidence for the formation of a locally coherent structure from the present set of experiments is compatible with the parser *fallibility hypothesis* wherein the comprehension system can generate invalid parses. There are a number of potential explanations to the observed local coherence effect such as a SOPARSE account of parsing, good-enough processing or the noisy-channel processing framework. In particular, it is unclear from the results so far if the faster RTs in the non-canonical condition reflect active prediction of a locally coherent parse on the part of the parser or is it the case that the parser forms a locally coherent parse bottom-up using the input encountered in the post-RC region? There are key differences between the two kinds of parsing processes in the current set up: while a top-down account assumes that a locally coherent parse is *predicted* at the post-RC object, a bottom-up account assumes that a locally coherent parse is formed because the prediction of the head NP and associated structure is less activated. Furthermore, while a bottom-up account of local coherence is compatible with SOPARSE and good-enough processing mechanisms, a top-down account would be more compatible with the noisy-channel processing framework. In order to explore this issue further, we coded the completion data discussed in Experiment 6 for two additional variables. These were, (a) whether a resumptive pronoun referencing the head NP was used in the completion (1 for yes, 0 otherwise), and (b) whether the completion was locally coherent with the RC object (1 for yes, 0 otherwise). Sentence 15(a) shows an instance where a resumptive pronoun (*vah* ‘that’) is used during the completion. The use of resumptive pronouns in this context can be understood to constitute a successful attempt at reactivating the head NP (cf. Keshtiari & Vasishth, 2012) especially when the continuation is otherwise well-formed vis á vis the integration of the head NP and post RC-verb. Sentence 15(b) shows an instance of a locally coherent parse.

- (15) a. Use of resumptive pronoun

vah raajkumaarii_i jisne kal bahut dilchasp_{ii} se kitaab
 That princess_{Fem} who_{Erg} yesterday lots interest with book_{Fem}
 paRhii thii **vah_i shaayad moTii thii** ...
 read_{Fem} had_{Fem} that maybe fat_{Fem} was_{Fem} ...

b. Locally coherent completion

vah aadmi jisne parson dhiire dhiire
 That man who_{Erg} day-before-yesterday slowly
 cunii thii **divaar_i aaj vah_i dheh gaai** ...
 constructed_{Fem} had_{Fem} wall today that fall_{Fem} had_{Fem} ...

Table 19 shows percentage completion for the two variables. These results show that compared to the Non-canonical condition, the number of resumptive pronouns was higher in the Canonical condition ($z=4.2$). The number of locally coherent completions did not differ between the two conditions ($z=0.55$). A higher number of resumptive pronouns in the Canonical condition suggests that compared to the Non-canonical condition, the head NP was more likely to be retrieved successfully in the Canonical condition. Combined with the difference in grammatical completions between the two conditions, the results suggest that the parser's tendency to form locally coherent structure is mostly driven by bottom-up information rather than by a top-down process.

Table 19

Percentage instances of resumptive pronouns and locally coherent structures in Experiment 6

Completion type	% Completion	
	Canonical	Non-canonical
Resumptive pronoun for head NP	47	25
Locally coherent	7	9

However, the issue of which of the two bottom-up accounts discussed above – SOPARSE and good-enough processing – is the best explanation for the current set of observations needs further investigation.

One possibility worth considering is that the two bottom-up accounts may be linked to differing parsing strategies which arise in different contexts. For instance, while

SOPARSE may correspond to a more general parsing strategy used to navigate bottom-up structure building that is not necessarily triggered by any particular processing difficulty, good-enough processing may be a parsing strategy that is adopted in the face of processing difficulty due to high memory load (Ferreira & Patson, 2007). Furthermore, given that it is plausibility rather than agreement that underlies the observed local coherence effects in our experiments, and SOPARSE is argued to utilize both semantic and syntactic features, SOPARSE in its current form will be inadequate, tilting the scales towards good-enough processing. Alternatively, our findings could be used to motivate a more curtailed SOPARSE model with weighted vectors such that the model privileges semantic information over syntactic information during bottom-up structure building.

Understanding the mechanisms that subserve the formation of a locally coherent structure during online comprehension is critical because it highlights the interaction between top-down and bottom-up processing. The locally coherent parse in previous work that probed this interaction (e.g. Gibson, 2006) was mainly driven by lexical ambiguity (also see, Bicknell & Levy, 2009). The locally coherent structures explored in our work, on the other hand, are driven by syntactic dependency resolution involving argument structure, agreement operation, etc. (cf. Kamide & Kukona, 2018). Indeed, a fuller understanding of this interaction is critical in uncovering the parsing mechanisms in SOV languages.

Implications for processing of SOV languages

As discussed in the Introduction section, it is generally understood that clause final verbal prediction is quite robust in SOV languages (e.g., Konieczny, 2000; Levy & Keller, 2013; Vasishth et al., 2010). On this account, the current results are quite striking. They show that when the processing load is high, the prediction mechanism in an SOV language can falter. In the experiments reported, an embedded RC as well as a cost of revision due to dashed expectation, was assumed to lead to high processing load. These factors are known to cause processing difficulty during comprehension (e.g.,

Frazier, 1985; Gibson, 1998). One implication of this would be that processing in SOV languages should be easier when the preverbal linguistic configuration is simple and when the core verbal arguments are closer to the the clause final verb, thus highlighting the influence of working-memory limitation (cf. Futrell, Gibson, & Levy, 2020; Gibson et al., 2013; Ros, Santesteban, Fukumura, & Laka, 2015; Ueno & Polinsky, 2009). Recent experimental/corpus studies in Hindi and other SOV languages do support this hypothesis. For example, Apurva and Husain (2020a, 2020b) find that parsing errors in Hindi increase with complex argument structure involving multiple embeddings and increased linear distance between the verb and its arguments. In a cross-linguistic corpus study involving multiple SOV languages (including Hindi), Yadav, Vaidya, Shukla, and Husain (2020) found that the linguistic context intervening a verb and its dependent has few embeddings. Similarly, using a corpus study, Sharma, Futrell, and Husain (2020) found that, compared to adjuncts, arguments (subject, objects and indirect objects) in Hindi, are, on average, linearly close to the verb. Together these results suggest that processing in a head-final language like Hindi involves robust prediction when the preverbal linguistic material is simple; with increased preverbal complexity, prediction can become fallible and the parser can falter.

Broader implications for prediction and future directions

The results discussed in this paper allow us to segue into more specific questions about predictability and the load-bearing capacity of the parser. In the current set of experiments, speakers needed to parse various dependencies within the relative clause (e.g. argument structure, agreement) and resolve the connection between the relative clause and matrix clause, all while also navigating the word order variations and retaining the details of the predicted matrix structure. While it is clear that Hindi speakers struggle with the prediction of matrix verb as generated at the head NP in this context, the details of what is remembered and what is forgotten needs further probing. For instance, we can ask which features of the upcoming verb are predicted, and which of these predicted features are retained at the end of a period of processing complexity.

Recent work suggests that native speakers of Hindi predict valency of the verb and its selection restrictions (Apurva & Husain, 2020b; Husain et al., 2014), event structure (Eva Wittenberg & Vaidya, 2020), tense and agreement features (Dillon, Nevins, Austin, & Phillips, 2012). However it is still unclear if some of these predictions are more robust than others and therefore more likely to be retained.

Verb agreement in Hindi may be one domain where forgetting of specific predictions could be investigated. Bhatia (2019) has demonstrated agreement attraction effects for matrix agreement in sentences with a center embedded relative clause. However, attraction effects were restricted to contexts where the distractor noun in the relative clause was itself an agreement controller within that relative clause. These attraction effects can be viewed as an instance of a failure to retain the specific prediction for the matrix clause verb – it may be the case that the agreement dependency between the distractor NP and the relative clause internal verb interferes with the specific prediction about the agreement features of the matrix verb generated based on the features of the head NP leading to the forgetting of this prediction. More broadly, it will be important to investigate the underlying representation of the predictions and how this representation relates to the linguistic entity that generated the predictions. This should help to identify the limits on the adaptability of the parser to head-final structures, thereby giving us a fuller picture of human cognitive capacities.

Conclusion

In this paper we examined the nature of predictive structure building during incremental processing in head-final languages in the context of high processing load created by varying word order in embedded relative clauses. We contrasted two hypotheses: the adaptability hypothesis, which predicted that speakers ought to face no difficulty in building head-final structures on account of the parser’s adaptation to language specific statistics, and the fallibility hypothesis which predicted that in contexts with high processing load the parser would become susceptible to errors even in head-final languages. Across various experiments on processing and production we

observed that Hindi speakers struggled with head-final structures. They formed grammatically invalid parses as indexed by a local coherence effect in reading ungrammatical sentences, with this effect resting on the plausibility of the local parse. Hindi speakers also struggled with maintaining predictions regarding the matrix structure as indexed by ungrammatical completions and faulty structural integrations. The results draw attention to the limits of parser adaptability in head-final languages and demonstrate that the parser can falter with respect to predicting and maintaining linguistic structure in the face of increased complexity. Further we suggest that in the context of the fallibility hypothesis, the pattern of results is best explained by bottom-up processes, with good-enough processing being the most likely model.

Acknowledgements

We would like to thank Apurva, Niralee Gupta, Vasundhara Srivastava, and Srabasti Dey for their help with data collection.

Funding

This work was partially supported by a Department of Science and Technology - Cognitive Science Research Initiative grant (SR/CSRI/29/2015) to Samar Husain.

References

- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–256.
- Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502–518.
- Apurva, & Husain, S. (2020a). Anti-locality effect in hindi: Evidence for shallow parsing. *In submission*.
- Apurva, & Husain, S. (2020b). Parsing errors in hindi: Investigating limits to verbal prediction in an sov language. *In submission*.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255–278.
- Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bhatia, S. (2019). *Computing agreement in a mixed system* (Doctoral dissertation, University of Massachusetts, Amherst, USA). (Doctoral Dissertation) doi: <https://doi.org/10.7275/csbe-8c55>
- Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., & Xia, F. (2009). A multi-representational and multi-layered treebank for Hindi/Urdu. In *Proceedings of the third linguistic annotation workshop* (pp. 186–189).
- Bicknell, K., & Levy, R. (2009, June). A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs. In *Proceedings of human language technologies: The 2009 annual conference of the north American chapter of the association for computational linguistics* (pp. 665–673). Boulder, Colorado: Association for Computational Linguistics. Retrieved from <https://www.aclweb.org/anthology/N09-1075>
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205.

- Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, 59(s1), 126–161.
- Dillon, B., Nevins, A., Austin, A. C., & Phillips, C. (2012). Syntactic and semantic predictors of tense in hindi: an erp investigation. *Language and Cognitive Processes*, 27(3), 313–344.
- Engelmann, F., & Vasishth, S. (2009). Processing grammatical and ungrammatical center embeddings in English and German: A computational model. In A. Howes, D. Peebles, & R. Cooper (Eds.), *Proceedings of 9th International Conference on Cognitive Modeling*. Manchester, UK.
- Eva Wittenberg, M. G., & Vaidya, A. (2020). How does case marking constrain event representations? In *The first south asian forum on the acquisition and processing of language (safal)*.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164 - 203. doi: [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)
- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11-15.
- Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, 1, 71-83.
- Fischler, I. S., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18(1), 1-20.
- Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science*, 40(3), 554–578.
- Frazier, L. (1985). Syntactic complexity. In L. K. D. Dowty & A. Zwicky (Eds.), *Natural language parsing* (pp. 129–189). Cambridge: Cambridge University Press.
- Futrell, R., Gibson, E., & Levy, R. (2020). Lossy-context surprisal: An

- information-theoretic model of memory effects in sentence processing. *Cognitive Science*.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Gibson, E. (2006). The interaction of top–down and bottom–up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54(3), 363 – 388. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0749596X0500149X>
doi: <https://doi.org/10.1016/j.jml.2005.12.005>
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*, 24(7), 1079–1088.
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3), 225–248.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2), 261–290.
- Hale, J. T. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4).
- Häussler, J., & Bader, M. (2015). An interference account of the missing-vp effect. *Frontiers in Psychology*, 6, 766.
- Hsiao, Y., & MacDonald, M. C. (2013). Experience and generalization in a connectionist model of mandarin chinese relative clause processing. *Frontiers in psychology*, 4, 767.
- Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong Expectations Cancel Locality Effects: Evidence From Hindi. *PloS one*, 9(7), e100986.
- Just, M. A., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in

- reading comprehension. *Journal of experimental psychology: General*, 111(2), 228.
- Kachru, Y. (2006). *Hindi*. John Benjamins Publishing Company.
- Kaiser, E. (2014). Experimental paradigms in psycholinguistics. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (p. 135–168). Cambridge University Press. doi: 10.1017/CBO9781139013734.009
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94, 113–147.
- Kamide, Y., & Kukona, A. (2018). The influence of globally ungrammatical local syntactic constraints on real-time sentence comprehension: Evidence from the visual world paradigm and reading. *Cognitive Science*, 42(8), 2976–2998. doi: 10.1111/cogs.12694
- Keshtiari, N., & Vasisht, S. (2012). Reactivation of antecedents by overt vs null pronouns: Evidence from persian. *Journal of Language Modelling*, 1. doi: <http://dx.doi.org/10.15398/jlm.v1i2.54>
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1-2), 262–284.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6), 627–645.
- Konieczny, L., Müller-Feldmeth, D., Hachmann, W., Schwarzkopf, S., & Wolfer, S. (2009, 01). Local syntactic coherence interpretation. evidence from a visual world study. In *Proceedings of the 31th annual conference of the cognitive science society* (pp. 1133–1138). Austin, TX: Cognitive Science Society.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in japanese. *Journal of Memory and Language*, 83, 152 - 178.

- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York, NY: Oxford University Press.
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161 – 163.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.
- Levy, R. (2013). Memory and surprisal in human sentence comprehension. In R. P. G. van Gompel (Ed.), *Sentence processing*. London: Psychology Press.
- Levy, R., & Keller, F. (2013). Expectation and locality effects in german verb-final structures. *Journal of memory and language*, *68*(2), 199–222.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on just and carpenter (1992) and waters and caplan (1996). *Psychological Review*, *109*(1), 35–54.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.
- Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. B. R.D. Luce & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419–492). New York: Wiley.
- Nakatani, K., & Gibson, E. (2010). An on-line study of japanese nesting complexity. *Cognitive Science*, *34*(1), 94-112.
- Pandharipande, R., & Kachru, Y. (1977). Relational grammar, ergativity, and Hindi-Urdu. *Lingua*, *41*(3-4), 217–238.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

- R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Ros, I., Santesteban, M., Fukumura, K., & Laka, I. (2015). Aiming at shorter dependencies: the role of agreement morphology. *Language, Cognition and Neuroscience*, 30(9), 1156–1174.
- Schwanenflugel, P. J., & Lacount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning Memory and Cognition*, 14(2), 344–354.
- Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24(2), 232–252.
- Sharma, K., Futrell, R., & Husain, S. (2020). What determines the order of verbal dependents in hindi? effects of efficiency in comprehension and production. In *Proceedings of cognitive modeling and computational linguistics (cmcl)*. (Accepted)
- Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9, 311–327.
- Staub, A., & Clifton, J., C. (2006). Syntactic prediction in language comprehension: Evidence from either ... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 425–436.
- Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, 36(1), 201–216.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355 – 370.
- Taylor, W. (1953). ‘cloze’ procedure: A new tool for measuring readability. *Journalism*

- Quarterly*, 30, 415–433.
- Ueno, M., & Polinsky, M. (2009). Does headedness affect processing? a new look at the vo–ov contrast. *Journal of Linguistics*, 45(3), 675–710. doi: 10.1017/S0022226709990065
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 767–794.
- Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, 25(4), 533–567.
- Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, 9, 2. doi: 10.3389/fpsyg.2018.00002
- Yadav, H., Vaidya, A., Shukla, V., & Husain, S. (2020). Word order typology interacts with linguistic complexity: a cross-linguistic corpus study. *Cognitive Science*, 44(4).
- Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American philosophical society*, 104(5), 444–466.
- Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics*, 36(3), 277–284.