Non-projectivity and processing constraints: Insights from Hindi

Samar Husain
Indian Institute of Technology, Delhi
Department of Humanities and Social Sciences
India
samar@hss.iitd.ac.in

Shravan Vasishth
Universität Potsdam
Department of Linguistics
Germany
vasishth@uni-potsdam.de

Abstract

Non-projectivity is an important theoretical and computational concept that has been investigated extensively in the dependency grammar/parsing paradigms. However, from a human sentence processing perspective, non-projectivity has received very little attention. In this paper, we look at existing work and propose new factors related to processing non-projective configuration. We argue that (a) counter to the claims in the psycholinguistic literature (Levy et al, 2012), different aspects of prediction maintenance can lead to higher processing cost for a non-projective dependency, (b) parsing strategies can interact with the expectation for a non-projective dependency, and (c) memory (re)activation can explain processing cost in certain non-projective configurations.

1 Introduction

Within the dependency grammar framework, non-projectivity has received considerable attention from both the theoretical as well as the computational perspectives. Non-projective structures are assumed to be both more complex to analyze as well as more difficult to parse. Figure 1 shows a Hindi sentence involving a non-projective dependency between *abhay kaa* 'Abhay's' and *caSamaa* 'spectacles'.

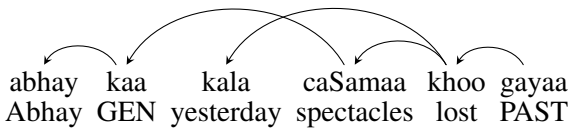

Figure 1: A Hindi sentence involving a non-projective dependency. English translation: *'Abhay's spectacles got lost yesterday.'*

Formally, an arc $i \rightarrow j$ is projective if and only if there is no word k between i and j that i does not dominate[1] (Nivre and Nilsson, 2005).

While some parsing paradigms can handle such dependencies, others either cannot or have special mechanisms to process them (e.g., Kuhlmann and Nivre (2010); Rambow and Joshi (1994)). Many theoretical approaches have special mechanisms to account for these constructions within their framework (e.g., Chomsky (1981); Pollard and Sag (1994)).

It is unclear if the complexity arising from non-projectivity has any processing cost in human language comprehension. That is, does the human sentence processing system find such sentences difficult to process, compared to projective dependencies? Previous work has addressed this question. In a classic study, Bach et al. (1986) showed that Dutch speakers find cross-serial dependencies in Dutch more acceptable compared to German speakers who read matched set of embedded constructions in German. Other work has looked at filler-gap dependencies, but these have generally focused on the question of wh movement (e.g., Traxler and Pickering (1996)). More recently, Levy et al. (2012) have directly taken up the issue of non-projectivity and sentence processing. They raised the following questions:

1. *Under what circumstances are non-projective dependency structures easier or harder to comprehend than corresponding projective-dependency structures?*

2. *How can these differences in comprehension difficulty be understood with respect to existing theories of online comprehension?*

Levy et al. (2012) try to answer the above questions using right-extraposed relative clauses in English. They show that the right-extraposed version

[1] Linearly, i could either precede j or follow it.

Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015), pages 141–150,
Uppsala, Sweden, August 24–26 2015.

is more costly than the embedded relative clause (RC), hence demonstrating that non-projective structures are indeed costlier than the projective counterpart. Additionally, they argue that the expectation-based theory of surprisal (Levy, 2008) explains the experimental results better than other competing theories like the cue-based memory model of Lewis and Vasishth (2005) and the derivational theory of complexity (Miller, 1962).

In this paper, we take up Levy's questions by investigating non-projectivity in Hindi participle clauses. We confirm that non-projectivity is indeed costly. However, we show that surprisal is unable to account for the increased processing cost, and that the cue-based memory model of Lewis and Vasishth (2005) can partly account for the results. To anticipate the conclusion, we argue that while expectation (formalized as conditional probability of the head in a dependency given previous syntactic dependencies) is relevant for explaining processing of non-projective dependencies, other factors (that can be orthogonal to predictive processing) can be equally critical. In particular, the following factors are implicated in the processing of non-projective dependencies: (a) The nature of the intervening material between a head and its dependent; (b) The nature of the head-dependent relation; (c) The length/complexity of the intervening material; (d) Memory activation; and (e) Parsing strategies.

Hindi[2] is a useful language for investigating non-projectivity because its relatively free-word order allows non-projective dependencies to occur quite frequently (see Mannem et al. (2009) for a more detailed discussion).

The paper is organized as follows, we first discuss relevant processing theories and their predictions regarding non-projectivity in Section 2. Following this, in Section 3 we discuss experiments that investigate processing of non-projective structures in Hindi. In Section 4 we discuss these findings and discuss potential factors that could influence processing non-projective configurations. Section 5 concludes.

[2]Hindi is one of the official languages of India. It is the fourth most widely spoken language in the world [source: http://www.ethnologue.com/statistics/size]. It is a free-word order language and is head final. It has relatively rich morphology with verb-subject, noun-adjective agreement. See Kachru (2006) for more details on the grammatical properties of Hindi.

2 Two theories of sentence comprehension

Here, we introduce two well-established theories of sentence comprehension, surprisal and the cue-based memory model, and discuss their predictions regarding the processing of non-projective dependencies.

2.1 Surprisal

Expectation-based theories appeal to the predictive nature of the human sentence comprehension system. On this view, processing becomes difficult if the upcoming sentential material is less predictable. Surprisal (Levy, 2008) is one such account. Surprisal presupposes that sentence-comprehenders know a grammar describing the structure of the word-sequences they hear. This grammar not only says which words can combine with which other words but also assigns a probability to all well-formed combinations. Such a probabilistic grammar assigns exactly one structure to unambiguous sentences. But even before the final word, one can use the grammar to answer the question: what structures are compatible with the words that have been read (or heard) so far? This set of structures may contract more or less radically as a comprehender makes their way through a sentence. Intuitively, surprisal increases when a parser is required to build some low-probability structure. Surprisal formalises the processing difficulty of a non-projective dependency (for that matter any dependency) as the conditional probability of encountering the head of the dependency given previous context. The processing cost at word *n* can be formally represented as (1).

$$\text{surprisal}(n) = \log \frac{1}{\Pr(n|context)} \qquad (1)$$

It is easy to see that surprisal can predict higher processing cost of a non-projective dependency because such dependencies are generally quite infrequent compared to their projective counterpart.

2.2 The cue-based memory model

The cue-based memory model is a working memory-based theory of human sentence processing proposed by Lewis and Vasishth (2005). Here sentence processing is modeled as skilled memory retrieval, where independently motivated principles of memory and cognitive skill play an im-

portant role in formulating the overall model. It uses the notion of decay as one determinant of memory retrieval difficulty. Elements that exists in memory without being retrieved for a long time will decay more, compared to elements that have been retrieved recently or elements that are recent. In addition to decay, the theory also incorporates the notion of interference. Memory retrievals are feature based, and feature overlap during retrieval, in addition to decay, will cause difficulty. The activation of a word i is computed using (2).

$$A_i = B_i + \left(\sum_j W_j S_{ji} \right) + \epsilon_i \qquad (2)$$

Activation is based on two separate quantities. One is the word's baseline activation B_i, which calculates activation decay due solely to the passage of time. The second variable that is used in determining a word's activation is the amount of similarity-based interference that occurs with other words that have been parsed (see Lewis and Vasishth, 2005 for a more extensive discussion).

The cue-based memory model also predicts higher processing cost for certain non-projective configurations such as the one shown in figure 2. Vasishth and Lewis (2006) have proposed that the reactivation of upcoming VPs by adjuncts, and/or reactivation of arguments by intervening adjuncts might lead to facilitation at the reactivated VP. This is because such modifications lead to an activation boost of the upcoming verb. Now assume a non-projective structure for figure 2 where *adjunct1* does not modify the non-finite verb, rather it modifies the matrix verb that follows the non-finite verb. This will make NP-gen ← non-finite verb a non-projective dependency. The cue-based model will predict higher processing cost at the non-finite verb in the non-projective case as fewer pre-modifers will reactivate the critical non-finite verb compared to when all intervening phrases modify the verb in the projective configuration.

So, both surprisal (via expectation) and cue-based memory model (via memory activation) predict higher processing cost for certain non-projective configurations. The first experiment described in the next section tests this prediction using self-paced reading. The second experiment is a sentence completion study and tests the hypothesis that subjects tend to avoid producing non-projective dependencies when they can. Together,

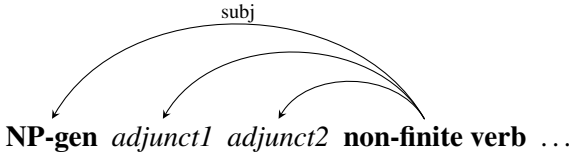

Figure 2: The base activation of a memory chunk gets a boost everytime it gets retrieved after it has been created. Above we show a schematic configuration where the non-finite verb is created/predicted at NP-gen, and it gets reactivated by its modifiers, *adjunct1* and *adjunct2*. NP-gen: Noun phrase with a genitive postposition.

these two studies suggest that reactivation can attenuate the cost of non-projective dependencies, and non-projective structures are hard (otherwise subjects would not try to avoid building them).

3 Experiments

We discuss two experiments in this section. In the first experiment, we test whether expectation and memory activation affect non-projective dependency configuration.

3.1 Experiment 1: Role of Memory Activation

The experiment has a $2 \times 2 \times 2$ factorial design, with factors Distance, Attachment, and Context. The critical region, where the dependency of interest is completed, is the non-finite verb *hasnaa* 'laughing' (see examples 1). In the context condition, the subject of the non-finite verb *raam kaa* and the non-finite verb *hasnaa* are expected, while in the no-context conditions they are not. As shown in Figure 3 and the examples 1, the attachment factor has two levels, an intervening phrase either attaches with the main verb (AttachMV) (Figure 3a), or it attaches to the non-finite verb (AttachNFV) (Figure 3b). The intervening phrase, *mere Xayaal se* 'according to me', does not modify the non-finite verb (rather it modifies the main verb); by contrast, *meri vajah se* 'because of me', modifies the non-finite verb. The Distance factor has two levels; in the short condition there is an adverbial modifying the upcoming non-finite verb (example 1a) compared to three adverbials in the long condition (example 1b). The Distance manipulation modulates the activation of the critical non-finite verb; as explained in section 2.2, in the cue-based model, more preverbal modification can

lead to higher memory activation.

Note that in examples 1, some conditions are not shown due to space constraints, but they can be derived from the other conditions. In the context conditions participant first see a screen with *kyaa raam kaa haMsnaa Thiik thaa?* 'Was it ok for Ram to laugh' (literally: Was Ram's laughing ok?). Following this, they see the critical sentence (shown below) on the next screen. In the no-context condition, they see *kyaa huaa?* 'What happened?' prior to seeing the critical sentence (shown below). The dots after each sentence represent the continuation *bilkul Thiik thaa, aisaa karne meM koii buraaii nahi hai* 'was absolutely ok, there is no harm in doing that'. All experimental items can be obtained from http://web.iitd.ernet.in/∼samar/data/experimental-items-depling2015.txt

(1) a. **Short, AttachMV, Context**

haan, / **[raama kaa** / mere Xayaal se
yes, Ram GEN according to me
/ zor zor se / **haMsnaa]** / . . .
loudly laughing . . .

'Yes, according to me it was absolutely ok for Ram to laugh loudly, there is no harm in doing that.'

b. **Long, AttachMV, Context**

haan, / **[raama kaa** / mere Xayaal se
yes, Ram GEN according to me
/ do din pehle / sabke saamne /
two days ago in front of everyone
zor zor se / **haMsnaa]** / . . .
loudly laughing . . .

'Yes, according to me it was absolutely ok for Ram to laugh loudly two days ago infront of every one, there is no harm in doing that.'

c. **Short, AttachNFV, Context**

haan, / **[raama kaa** / merii vajah se /
yes, Ram GEN because to me
zor zor se / **haMsnaa]** / . . .
loudly laughing . . .

'Yes, it was absolutely ok for Ram to laugh loudly because of me, there is no harm in doing that.'

d. **Long, AttachNFV, Context**
see above

e. **Short, AttachMV, No context**

[raama kaa / mere Xayaal se /
Ram GEN according to me
zor zor se / **haMsnaa]** / . . .
loudly laughing . . .

'According to me it was absolutely ok for Ram to laugh loudly, there is no harm in doing that.'

f. **Long, AttachMV, No context**
see above

g. **Short, AttachNFV, No context**
see above

h. **Long, AttachNFV, No context**
see above

3.1.1 Procedure and Participants

We used the centered self-paced reading (SPR) method (Just et al., 1982); centering was used to prevent readers from using the sentence-length cue to adapt their processing strategy. Stimulus items were presented using Douglas Rohde's Linger software, version 2.94 (http://tedlab.mit.edu/∼dr/Linger/). A Latin square design ensured that each participant saw each item in only one condition. The target items and fillers were pseudo-randomized for each participant.

The experimenter (Husain) began by explaining the task to the participants. After this, six practice sentences were presented in order to familiarize participants with the task. At the beginning of each trial, the computer screen showed a single hyphen that covered the first word of the upcoming sentence; the hyphen appeared in the center of the computer screen. When the space bar was pressed, the word was unmasked. With each successive press of the space bar, the next word or phrase replaced the previous word in the center of the screen. This successive replacement continued until the participant had read the whole sentence. Reading times or RTs (in milliseconds) were taken as a measure of relative momentary processing difficulty. The f-key for was pressed for answering a question with a 'yes' response and the the j-key was pressed for answering with a 'no' response.

Eighty two native speakers of Hindi in Jawaharlal Nehru University, New Delhi, India, par-

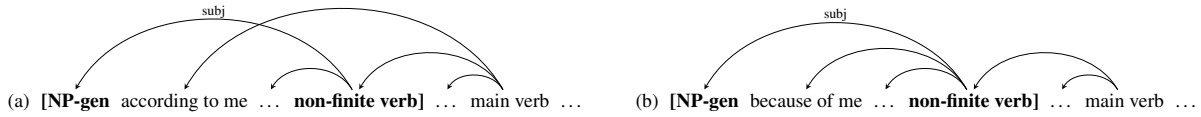

Figure 3: Projectivity manipulation in the self-paced reading (SPR) experiment discussed in section 3.1; see examples 1. (a) shows AttachMV, the main verb attachment condition, the non-projective dependency, while (b) shows AttachNFV, the embedded verb attachment condition, the projective dependency. NP-gen: Noun phrase with a genitive postposition.

ticipated for payment. Their mean age was 23.7 years, SD 3.3 years.

3.1.2 Statistical analyses

All analyses for fixation measures were carried out using the package `lme4`, version 1.1-7, (Bates et al., 2014) for fitting linear mixed models, which is available for R, version 3.1.2 (R Development Core Team, 2006). In the `lme4` models, we fit cross varying intercepts for subjects and items, no varying slopes for subject and item were estimated, as data of this size is usually insufficient to estimate these parameters with any accuracy. The data analysis was done on log-transformed reading times to achieve approximate normality of residuals. From the `lme4` analyses, we present the t-values (z-values for response data).

3.1.3 Pretest

Before conducting the SPR study, we carried out a sentence completion study to ensure that the experimental items used in the study had the appropriate properties. Participants were asked to complete the incomplete version of the items shown in (1); for example, for 1(a) they were supposed to complete the incomplete string *haan, raama kaa mere Xayaal se zor zor se …* Twenty four sets of items, each with eight versions were presented using the centered self-paced reading method in the standard Latin square design. Items were presented using Douglas Rohde's Linger software, version 2.94 (http://tedlab.mit.edu/∼dr/Linger/). The critical items were presented with 122 filler items unrelated to this study. Twenty-one Hindi native speaker in Jawaharlal Nehru University participated for payment. Their mean age was 22.7 years, SD 3.1 years.

The sentence completion confirmed that there were more exact predictions[3] in the context conditions (70.75%) compared to just 2.25% in the no-context condition; this confirms that the context condition allows us to manipulate the conditional probability of the upcoming critical non-finite verb. If considering the prediction of a non-finite verb category (i.e. any non-finite verb), then the percentage prediction in the context condition is 86.25%, and 56% in the no-context condition. This shows that in the no-context condition a non-finite verb is being predicted. Similarly, the exact prediction of the main verb was 81% and 31% respectively for the context and no-context conditions. If considering only the finite category information, i.e. any finite verb, this percentage prediction was 98% and 87% for context and no-context conditions respectively. Analysis of the binomial responses[4] using generalized linear mixed models with a logit link function also shows a significant main effect of context (z=5.76) on non-finite verb prediction accuracy.

3.2 Results

As mentioned above, the critical region in the SPR study was the non-finite verb. We find a main effect of context (t=-12.11), such that the non-finite verb was read faster in the context condition compared to the no-context condition. This is expected given the results of the sentence completion study just discussed. We also get an interaction between the three factors, distance, attachment, and context (t=-2.04). A nested contrast shows that this interaction is driven by the no-context, AttachNFV condition, such that the reading time at the non-finite verb is faster in the long condition compared to the short condition. Figure 4 shows the reading times for all the eight conditions.

[3] A response is considered as an exact prediction if it matches in type and tense/aspect features with the expected verb.

[4] Non-finite category prediction was coded as 1, while wrong category prediction was coded as 0. Data from two subjects were removed during the analysis as they did not understand the task.

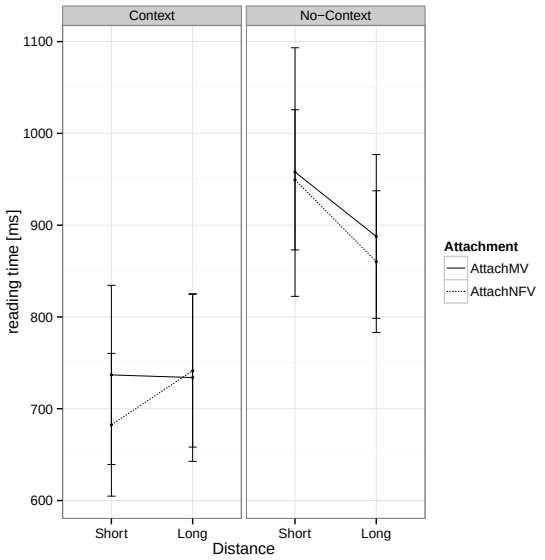

Figure 4: Reading times in ms (with 95% CIs) at the critical region (non-finite verb). The Distance × Attachment × Context interaction (t=-2.04) is driven by the No-Context condition. A nested contrast (details omitted due to lack of space) shows that RT in AttachNFV, Short, No-Context is longer than AttachNFV, Long, No-Context, this is evidence for reactivation effects as suggested by Vasishth and Lewis (2006). Note that the difference between the No-Context, AttachMV conditions is not significant.

3.2.1 Discussion

The three-way interaction is driven by a speedup in the attach non-finite verb (projective) condition when we compare the long vs short conditions in the no-context case. This is established by a nested contrast comparison. Additionally, in the attach main verb condition (the non-projective condition), when we compare long vs short conditions in the no-context case, we see no such speedup. This absence of a speedup could be due to the additional cost of non-projectivity. We suggest that the facilitation in reading time in the projective condition in long vs short cases (in the no-context condition) may be due to reactivation of the non-finite verb, and this is attenuated if the dependency is non-projective. This reactivation-based speedup is not seen in the context conditions (nested contrasts, not presented here, show that there is no significant interaction between distance and attachment in the context case). Thus, the underlying cause for the three-way interaction seems to be the reactivation-based speedup in the no-context condition. In other words, expectation in the context condition could be playing a role in eliminating any effect of reactivation between the two attachment types. These results can therefore be partly explained by Vasishth and Lewis (2006).[5]

The surprisal account cannot easily account for these results. As noted in section 3.1.3, a sentence completion study using the same items shows no significant difference in prediction type for the projective vs non-projection condition in the no-context condition. Surprisal will therefore only predict a main effect of the context condition and not predict any interactions. This does not seem to hold.

3.3 Experiment 2: The Role of Prediction Revision

Next, we investigate the role of prediction revision in processing non-projective configuration. We employ a sentence completion task with a modified design of example 1.

Similar to experiment 1, we use embedded non-finite constructions. This experiment also has a $2 \times 2 \times 2$ design: Distance × Attachment × Context. Context either generates a strong expectation for an upcoming non-finite verb or does not. The Distance factor has two levels; the short condition has one adverbial modifying the upcoming non-finite verb, while the long condition has three adverbials. The Attachment factor has two levels, AttachMV and AttachNFV. Compared to experiment 1, this manipulation has a subtle difference. While the phrase 'according to me' in the AttachMV condition of Experiment 1 was clearly an adjunct, in Experiment 2, the phrase used has an Accusative case-marker. The Accusative case marker in Hindi generally appears with arguments. In the AttachNFV condition, the phrase has the genitive case-marker, which generally appears with adjuncts. This is shown in example 2(a); the phrase *abhay ko* 'Abhay ACC[6]' is an argument of the matrix verb *lagaa thaa* 'found'. By modifying the matrix verb, *abhay ko* makes the dependency between *raama kaa* ← *haMsnaa* non-projective. In example 2(b), on the other hand, the phrase *ab-*

[5] An important caveat here is that the results are rather weakly supportive of the account we present. A stronger result would have entirely parallel lines in the context conditions, and a stronger effect size for the interaction seen in the no-context condition. We intend to try to replicate this effect in a future study.

[6] ACC: Accusative case-marker

hay par 'Abhay LOC[7]' is an adjunct of the upcoming non-finite verb *haMsnaa* 'laughing'. Example 2 shows only the attachment manipulation, we don't list all the items due to space constraints. In the context conditions participant first see a screen with *kyaa kal raam kaa haMsnaa Thiik thaa?* 'Was it ok for Ram to laugh yesterday' (literally: Was Ram's laughing yesterday ok?), following this, on the next screen, they see fragment of the critical sentence upto *zor zor se* 'loudly' (shown below). In the no-context condition, they see *kyaa huaa?* 'What happened?' prior to seeing the critical sentence. All experimental items can be obtained from http://web.iitd.ernet.in/∼samar/data/experimental-items-depling2015.txt

(2) a. **Short, AttachMV, Context**
haan Thiik thaa, magar,
yes ok was, but,
mere Xayaal se **[raama kaa**
according to me Ram GEN
abhay ko do din pehle zor zor se
Abhay ACC two days ago loudly
haMsnaa] Thiik nahii lagaa thaa
laughing good not find was
'Yes it was ok, however, according to me Abhay did not find it was ok for Ram to laugh loudly two days ago.'

b. **Short, AttachNFV, Context**
haan Thiik thaa, magar, man hi man
yes ok was, but, in my heart
mujhko **[raama kaa** abhay par
I ACC Ram GEN Abhay LOC
do din pehle zor zor se **haMsnaa]**
two days ago loudly laughing
Thiik nahii lagaa thaa
good not find was
'Yes it was ok, however, in my heart I did not find it ok for Ram to laugh loudly on Abhay two days ago.'

The question here was: when the reader is given a context in which an embedded non-finite verb is highly predictable, if he encounters a phrase that requires a non-projective dependency, would the prediction for the specific non-finite verb be revised such that a projective dependency is built with a different non-finite verb?

[7]LOC: Locative case-marker

Condition	% exact predictions
AttachMV	10
AttachNFV	53

Table 1: Exact prediction (in percentage) of the non-finite verb (*haMsnaa* 'laughing') in the sentence completion study for the AttachMV and AttachNFV conditions in the context, short conditions.

3.3.1 Procedure

The same procedure as discussed in section 3.1.3 was followed. The same subjects participated in the experiment.

3.3.2 Results

The dependent measure is the proportion of exact predictions for the non-finite verb in the different conditions. There are more exact predictions of the non-finite verb in the context conditions (29%) compared to just 3% in the no-context condition. This is as expected; however, note that the proportion of exact predictions is relatively low in the context condition (cf. table 1). This is because of the AttachMV condition—the non-projective dependency causes a reduction in the proportion of exact predictions; in this condition, participants tend to use verbs that would form a projective structure (more details in the next section). We found a a significant main effect of Attachment (z=-5.05) and of context (z=5.41).[8]

3.3.3 Discussion

Together, the main effect of Attachment, Context and the percent of exact predictions shown in table 1 suggests that subjects override the prediction generated by the context in order to avoid forming a non-projective dependency. The sentence completion data show that in the AttachMV (non-projective dependency) conditions subjects used verbs that were compatible with the critical case-markers (genitive and accusative), rather than using the verb used in the context. In doing so, they form a projective structure, rather than forming a non-projective structure using the context verb. For example, subjects tend to use a transitive participle (e.g., *maarnaa* 'hitting') due to the presence of *abhay ko* 'Abhay ACC' which is

[8]Non-finite category prediction was coded as 1, while wrong category prediction was coded as 0. Data from two subjects were removed during the analysis as they did not understand the task.

not easily incorporated with the contextual prediction of intransitive *haMsnaa* 'laughing'. Using *haMsnaa* after seeing an accusative case-marker is only possible by positing a non-projective dependency shown in example 2(a), i.e. *abhay ko* → *lagaa* makes *raama kaa* → *haMsnaa* dependency non-projective. On the other hand, in the AttachNFV (projective dependency) condition, the response was *haMsnaa* 'laughing', i.e. participants did not deviate from the verb that was provided in the context. This is because the case-marker on the phrase in the AttachNFV condition *abhay par* 'Abhay LOC' can easily be incorporated with an intransitive verb like *haMsnaa* 'laughing'.

Given these results, it is reasonable to assume that, in an online study, when subjects will hear/read *haMsnaa* 'laughing' in 2(a), they would be surprised (as they are expecting *maarnaa* 'hitting') leading to additional processing cost as a result of dashed expectation. Note that, surprisal will correctly predict that reading time at *haMsnaa* in sentence 2(a) will be higher than 2(b) because P(*haMsnaa*|Noun-ACC) will be lower than P(*haMsnaa*|Noun-LOC)[9]. However, it is important to stress that this cost does not reflect prediction maintenance per se (as is argued by Levy et al. (2012)), rather it is prediction revision that eventually gets reflected as additional processing cost.

4 General Discussion

Experiment 1 shows that for a Hindi participle clause construction involving a non-projective dependency, expectation in the context condition could be playing a role in eliminating any effect of reactivation between the two attachment types; recall that in the no-context condition, reactivation effect was seen in the projective dependency conditions while non-projective processing seemed to attenuate reactivation facilitation in the non-projective conditions. This shows that a non-projective structure might not be inherently difficult to process, a claim also made in Levy et al. (2012). Levy et al. (2012) essentially cast the problem of processing a non-projective dependency as maintenance of such syntactic expectation. While such a formalization does account for the processing difficulty in their experiments, it fails to explain the results discussed in section 3.2. Basically, Levy et al. (2012) do not explore processes that are orthogonal to surprisal but have relevance for non-projective dependency processing. One such process is **memory activation** discussed in Experiment 1.

Another factor, **prediction revision**, was illustrated in Experiment 2 where although surprisal does correctly predict the results, it does not flesh out the source of the processing cost. As shown in figure 5, we argue that the processing cost at a head depends on the compatibility of intervening material with the predicted head. Closely related to this is the issue of **dependency type**. While certain dependencies are more inert (e.g., Adj ← Noun), others are less so (e.g., Noun ← Verb). This has the effect of making a prediction more immune to the influence of other dependencies in some cases. For example, once a prediction for an extraposed RC is made, following material has little influence over the validity of the prediction. On the other hand, a prediction of a verb at an argument is susceptible to revisions once additional arguments are encountered. This means that together the dependency type and the intervening material influence the longevity of a prediction.

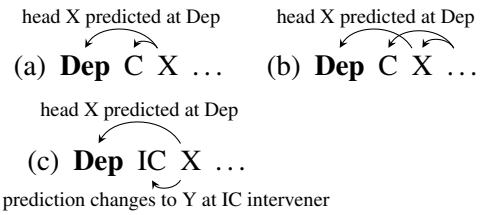

Figure 5: Incompatible (IC) vs compatible (C) intervener. Only when the intervener is compatible will the original prediction triggered at the dependent (Dep) be maintained. The compatible intervener can either cause the predicted dependency to be projective or non-projective. (a) was seen in example 2(b), (b) was seen in example 1(a), and (c) was seen in example 2(a).

We have so far discussed two factors (other than expectation strength) that can account for processing cost in non-projective structures, these are (a) memory activation, (b) prediction revision due to intervening material and dependency type. In addition to these one can posit some more factors.

One such factor is **prediction decay**. While keeping the prediction strength constant, a prediction can suffer memory decay due to the complexity of the intervening material. Such effects

[9] *haMsnaa* is an intransitive verb and in its non-finite form can only take a subject with a genitive case marker. It can easily take a locative adjunct however.

can arise due to limited working memory constraints. There is a large body of work that supports the role of working memory in sentence comprehension (e.g., Gibson (1998); Grodner and Gibson (2005)). Expectation-based theories such as surprisal do not make any predictions about such effects. Indeed, recent work has argued for a more unified approach to sentence processing where both expectation and working memory play a role (e.g., Vasishth and Drenhaus (2011); Levy and Keller (2012)).What concerns us here is the issue of expectation maintenance and how it interacts with working memory. Two recent results need to be mentioned here. For German, Levy and Keller (2012) show that the benefits of predictive processing can be attenuated (and be reversed) if the complexity of the phrases before the predicted head is high. Similary, Safavi et al. (2015) show that in Persian separable complex predicate, processing time at the light verb can be high in spite of it being highly predictable if the precritical phrase is a complex NP. Both works point to the possibility that even for a highly predictable non-projective dependency, processing cost can be influenced by the complexity of the intervening material. If this complexity is high, it will affect the prediction adversely and lead to higher processing cost of the non-projective dependency.

Another important factor is the **frequency of a dependency**. It is quite well known that non-projective dependencies are infrequent compared to their projective counterparts, for example, in English the right-extraposed RC is less frequent compared to the embedded RC[10] (Levy et al., 2012). Two related questions need to be asked here: (a) Will a dependency that is non-projective but highly frequent be easy to process? An interesting case in point is the relative clause in Hindi. Unlike English, the right-extraposed RC in Hindi is more frequent than the embedded RC. (b) Similarly, certain heads are always triggered due to the specific dependents, e.g., relative-correlative dependency and paired discourse connectives in Hindi. Many of these dependencies are non-projective (and are also long distance dependencies). Given their high collocational frequency, will they still be difficult to process? Surprisal will predict that, in Hindi, right-extraposed RC should be easier to process than the embedded counterpart. This needs to verified experimentally.

Finally, the processing cost of a non-projective dependency could also reflect certain **parsing heuristics/strategies**. For example, it is possible that when the expectation is weak (i.e. when the head of the dependency cannot be predicted with high certainty), cases like Figure 3(a) are costly due to incorrect dependency attachment. In particular, the phrase *according to me* is incorrectly attached to the upcoming unknown verb. After encountering the non-finite verb the attachment has to be revised leading to additional processing cost. Such a strategy implies that when expectation is weak and therefore prebuilding of structures is not possible, the parser employs a conservative projective attachment heuristic. The parser pursues and maintains a non-projective dependency only when the expectation strength is strong.

More recent developments in transition-based incremental parsing (Nivre, 2009) introduce special transitions to handle non-projectivity. Such transitions can only be employed in cases where expectation of a non-projective dependency is high, in all other cases a projective parsing algorithm could be pursued. In this context, the parsing strategies proposed by Joshi (1990)[11] to account for the results of Bach et al. (1986) are relevant. The ease of processing cross-serial dependency and the use of embedded push-down automata to process them could be understood as the parser adapting to a specific property of a language.

Processing cost of a non-projective dependency can therefore arise as a result of variety of factors. This could be either structural or non-structural. Structural factors include syntactic expectation, its revision and frequency. Non-structural factors include expectation decay, memory activation and parsing heuristics.

The factors mentioned above might interact in interesting ways and such interaction can form the focus of future investigations. In addition, as mentioned by Levy et al. (2012), information structure and grammatical weights might also have some role to play in determining processing cost in such syntactic configurations. In addition, it is an open question whether the processing patterns observed for non-projective dependency also hold true for other dependency configurations such as well-nestedness, etc. (Bodirsky et al., 2005).

[10] Table 1 in Levy et al. (2012), P(extraposedRC|context) is 0.00004, while P(RC|context) is 0.00561.

[11] Also see Rambow and Joshi (1994)

5 Conclusion

Current evidence suggests that human sentence processing is sensitive to non-projective dependencies. The increased processing cost could be a result of either structural or non-structural factors. It is unclear if these varied factors interact and if so under what circumstances. Current experimental research provides us with means to investigate these important questions along with investigating processing cost of other types of dependency configurations such as well-nestedness. Such investigations are critical and will constructively inform both theoretical work as well as parsing approaches in the dependency linguistics framework.

6 Acknowledgements

We would like to thank Dr. Ayesha Kidwai for helping with logistics to run the experiments at Jawaharlal Nehru University, Delhi.

References

E. Bach, C. Brown, and W. Marslen-Wilson. 1986. Crossed and nested dependencies in german and dutch: A psycholinguistic study. *Language and Cognitive Processes*, 1:249–262.

D. Bates, M. Maechler, B. M. Bolker, and S. Walker. 2014. lme4: Linear mixed-effects models using eigen and s4. ArXiv e-print; submitted to *Journal of Statistical Software*.

M. Bodirsky, M. Kuhlmann, and M. Möhl. 2005. Well-nested drawings as models of syntactic structure. In *In Tenth Conference on Formal Grammar and Ninth Meeting on Mathematics of Language*, pages 88–1. University Press.

N. Chomsky. 1981. *Lectures on government and binding*. Dordrecht: Foris.

E. Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76.

D. Grodner and E. Gibson. 2005. Consequences of the serial nature of linguistic input. *Cognitive Science*, 29:261–290.

A. K. Joshi. 1990. Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes*, 5:1–27.

M. A. Just, P. A. Carpenter, and J. D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2):228–238.

Y. Kachru. 2006. *Hindi*. John Benjamins Publishing Company, Philadelphia.

M. Kuhlmann and J. Nivre. 2010. Transition-based techniques for non-projective dependency parsing. *Northern European Journal of Language Technology*, 2(1):1–19.

R. Levy and F. Keller. 2012. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*.

R. Levy, E. Fedorenko, M. Breen, and E. Gibson. 2012. The processing of extraposed structures in English. *Cognition*, 122(1):12–36.

R. Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

R. L. Lewis and S. Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29:1–45, May.

P. Mannem, H. Chaudhry, and A. Bharati. 2009. Insights into non-projectivity in Hindi. In *ACL-IJCNLP Student Research Workshop*.

G. A. Miller. 1962. Some psychological studies of grammar. *American Psychologist*, 17:748–762.

J. Nivre and J. Nilsson. 2005. Pseudo-projective dependency parsing. In *Proceedings Of ACL 2005*.

J. Nivre. 2009. Non-Projective Dependency Parsing in Expected Linear Time. In *Proceedings of the Joint Conference of the 47th ACL and the 4th IJCNLP*, pages 351–359.

C. Pollard and I. A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago.

R Development Core Team, 2006. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

O. Rambow and A. Joshi. 1994. A Processing Model for Free Word Order Languages. In C. Clifton Jr., L. Frazier, and K.Rayner, editors, *Perspective on Sentence Processing*, pages 267–301. Erlbaum, Hillsdale, NJ.

M. S. Safavi, S. Vasishth, and S. Husain. 2015. Locality and expectation in Persian separable complex predicates. In *Proceedings of the 28th CUNY Sentence Processing Conference*, Los Angeles, CA.

M. J. Traxler and M. J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: an eye tracking study. *Journal of Memory and Language*, 35:454–475.

S. Vasishth and H. Drenhaus. 2011. Locality in German. *Dialogue and Discourse*, 1:59–82.

S. Vasishth and R. L. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4):767–794.