

## The role of clefting, word order and given-new ordering in sentence comprehension: Evidence from Hindi

SHRAVAN VASISHTH, RUKSHIN SHAHER, AND NARAYANAN SRINIVASAN,  
*University of Potsdam; University of Potsdam; Centre of Behavioural and Cognitive Sciences, India*

Received February 22, 2010; Revised December 31, 2012

### ABSTRACT

Two Hindi eyetracking studies show that clefting a noun results in greater processing difficulty initially, due to the extra processing steps involved in encoding a clefted noun (e.g., for computing the exhaustiveness interpretation). However, this extra difficulty in encoding a clefted noun results in a processing advantage when the clefted noun needs to be retrieved later on in the sentence – the clefted noun is retrieved faster in subsequent processing compared to its non-clefted counterpart. We also show that given-new ordering yields a processing advantage over new-given order, but this is only seen after the whole sentence is processed, i.e., it may be a late effect that occurs after syntactic processing is completed. Finally, following up on work on German by Hörnig et al. (2005), we present evidence that non-canonical order can be processed more easily than canonical order given appropriate context.

### 1 Introduction

What is the functional motivation for elaborate devices such as clefting and given-new ordering? Although it is clear that they serve as information-structuring devices for effectively communicating a message to the hearer/reader, it is less clear how exactly this kind of restructuring impacts processing in real-time sentence comprehension. In this paper, we investigate the role that clefting, given-new ordering and word order play in facilitating online sentence comprehension. We investigate these issues in Hindi, a language that has not been studied extensively in psycholinguistics.

We begin by surveying some of the relevant literature relating to clefting, given-new ordering and word order variation, and then present two Hindi eyetracking studies that address several open issues relating to these information-structuring devices.

#### 1.1 Given-new ordering and word order variation

Psycholinguistic research on language comprehension has shown that given-new ordering is a major information-structure (IS) device. A widely accepted finding is that ordering a given (e.g., previously mentioned) referent before a new referent facilitates processing (Clark and Haviland 1977, Yekovich et al. 1979). According to Clark and Haviland, listeners and speakers adhere to a *given-new contract*; when presented with a sentence, listeners compute what part of the sentence is the given information in order to find an antecedent for it in memory before integrating the new information into memory. The computation of given information is aided by cues such as focal stress, type of construction (passive, cleft, pseudocleft), etc.

Yekovich et al. (1979) demonstrated the given-new processing advantage using a comprehension time task. They showed that in sentences like *The shark attacked the diver near the reef*, the reading time for the entire

sentence was faster when the subject (*shark*), as opposed to the object (*diver*) was given in context (i.e., was previously mentioned).<sup>1</sup>

In a series of studies, Hörnig and colleagues uncovered several subtleties that extend the given-new generalization (see, e.g., Hörnig et al. 2005, Hörnig et al. 2006). They carried out a psycholinguistic investigation of spatial assertions; in a picture-verification task, participants first read two propositions P1 and P2 in sequence, and in doing so had to build a mental model of the relative spatial locations of three objects — call them A, B, and C (the propositions described the relative locations of these objects).<sup>2</sup> After participants signaled with a button-press that the two propositions had been processed, a picture containing A was shown, and after 1000 milliseconds another picture containing C was added to the display (in half the trials C was shown before A). Participants had to verify whether the objects' locations matched the descriptions in the two propositions.

A surprising finding was that if a proposition P1 like *A is to the left of B* is followed by another proposition P2 that describes the location of a third, new element *C* relative to the spatial location of the given element *B*, then P2 is processed faster if the locative expression containing *B* is topicalized, i.e., occurs first in the sentence. That is, processing is *easier* if P2 has a topicalized locative, *To the left of B is C*, and more difficult if the order is canonical, *C is to the left of B*.

Following other work, Hörnig et al. (2005) refer to the locative expression as the *relatum*, since the locative phrase relates the target referent (which in the above example is *C*) in the mental model (Johnson-Laird 1983) of the comprehender. They hypothesized that two constraints operate independently and additively to determine integration speed in comprehension: the *relatum* should be given (*relatum=given*), and the order of elements should be: given, then new (*given-new*). These hypotheses were confirmed in their picture verification experiments.

The work by Hörnig and colleagues motivates one of our research questions: does the given-new preference hold in a language like Hindi, and can we find evidence for the *relatum=given* constraint that Hörnig and colleagues found in German, which, like Hindi, is also a relatively free word-order language? These constraints are expected to be cross-linguistically applicable, but there appears to be little cross-linguistic evidence available beyond relatively well-studied languages like English and German.

## 1.2 Clefting

Apart from the given-new ordering and *relatum=given* constraints, focus also plays an important role in information structuring. It is well-known that listeners detect focused information more quickly than non-focused information (Cutler and Fodor 1979) and can also remember it better than non-focused information (Singer 1976). A dramatic example of the effect of focus is the disappearance of the Moses illusion under focus. The Moses illusion refers to the following phenomenon. When participants are asked, *How many animals of each kind did Moses take on the Ark?*, most respond *two*, even though they know that the person concerned here is Noah and not Moses. This illusion disappears when the incorrect information is focused by clefting the critical noun; e.g., when they are asked to judge the truth of a sentence like *It was Moses who took two animals of each kind on the Ark*. This suggests that it is easier to detect errors in focused information as opposed to non-focused information (Bredart and Modolo 1988, Bredart and Docquier 1989).

Like other types of focus, *it*-clefts also facilitate processing in online sentence comprehension tasks. In a probe recognition and naming task, Birch and Garnsey (1995) observed that clefted nouns are named faster and more accurately than non-clefted nouns, suggesting that it is easier to access the clefted material. Similarly, in an eyetracking study on pronoun resolution, Foraker and McElree (2007) found that clefted noun phrases were more available as antecedents for anaphoric pronouns than their non-clefted counterparts; they argued that this clefting advantage comes from activation boosts rather than special buffers for clefted elements.

<sup>1</sup>Note, however, that the given-new preference is not a global constraint on processing. For example, Frazier and Clifton (2004) showed that the given-new ordering preference is limited to certain constructions. Furthermore, several studies have shown that the given-new preference does not hold for premise integration in a spatial reasoning task (Ehrlich and Johnson-Laird 1982, Johnson-Laird 1983, Baguley and Payne 1998).

<sup>2</sup>In their experiments, the letters A, B, and C stand for names of animals.

Interestingly, *it*-clefts interact with the given-new ordering constraint: so-called focus-background clefts (Delin and Oberlander 1995) like *It was John who bought a book* generally partition the sentence into a new-information segment (the clefted noun) and a given-information segment (the predicate), resulting in a new-given ordering (cf. Delin and Oberlander 1995; for so-called topic-comment *it*-clefts, this generalization does not hold). As a consequence, in focus-background *it*-clefts the default ordering is not given-new but rather new-given. In such clefts, it is possible that the preferred order is new-given, not given-new.

One goal of the present study was to investigate whether the given-new ordering preference is reversed in cleft constructions. Another goal was to further investigate the effect that clefting has on processing, independent of the given-new ordering constraint. We discuss the background for this second goal next.

As mentioned above, clefted material has been argued to be more accessible in memory. One reason for this greater accessibility may be that the focused item achieves a higher activation in memory because the clefted item ends up being encoded more richly. In a clefted structure like *It was John who bought a book*, an exhaustiveness interpretation is implied: no one but John bought a book.<sup>3</sup> Processing the clefted noun would at least involve generating the exhaustiveness reading and associating it with the noun; these are probably the underlying processing steps that result in a richer encoding of the clefted noun in memory. Indeed, recent work by Hofmeister (2009) shows that this extra processing increases the activation of the clefted noun in memory, which in turn results in easier retrieval of the noun at a subsequent stage.

Thus, the evidence shows that clefting a noun makes it more accessible in memory at a subsequent stage. In addition, evidence also exists for increased encoding cost due to clefting; this comes from an eyetracking study conducted by Birch and Rayner (1997). They compared clefted sentences such as (1a) with a non-clefted sentence (1b).

- (1) a. It was the *suburb* that received the most damage from the ice storm.  
 b. Workers in the *suburb* hurried to restore power after the ice storm.

They found twice as many regressive eye movements<sup>4</sup> to the first three words of the sentence in the clefted condition compared to the non-clefted one. In addition, the probability of regressing out of the clefted word was higher but did not reach significance (note however that in a similar experiment (their 3B) presented in Birch and Rayner (2010), regression probability did reach significance). Finally, longer re-reading times were seen on the clefted versus non-clefted word. Birch and Rayner suggest that this is evidence for the encoding cost arising from the processing events triggered by a cleft structure. As an aside, note that the sentences in (1) are not really minimal pairs: the clefted noun in (1a) is a subject, whereas the same noun in (1b) occurs inside a prepositional phrase adjunct of a subject. For this reason, it is reasonable to question whether reading times can be compared in this pair of sentences. In fact, in a subsequent study, Birch and Rayner (2010) found shorter re-reading times in the clefted condition (their Experiment 3B).

The above review of the literature shows that clefting induces an encoding cost at the clefted word, but also results in facilitation later on in the sentence. However, to our knowledge the presence of both effects — increased encoding cost followed by facilitation during subsequent processing — has only been found in a self-paced reading study conducted by Hofmeister (2009) and in an eye-tracking study by Birch and Rayner (2010) (although Birch and Rayner argue in this paper that clefting only facilitates processing, their Experiment 3B shows increased regression probability at the clefted noun versus the non-clefted one). Indeed, recent work by Morris and Folk (1998) and Ward and Sturt (2007) argues that clefting results only in an integration advantage but (contrary to the claims by Birch and Rayner 1997, and contrary to Hofmeister's 2009 and Birch and Rayner's 2010 findings) not a greater encoding cost.

These disagreements in the literature lead to the second goal of the research presented here. We set out to establish whether (i) only encoding cost is seen with clefts, consistent with Birch and Rayner (1997); (ii) only facilitation effects are seen, consistent with Ward and Sturt (2007) and Morris and Folk (1998); or (iii) both encoding cost and subsequent facilitation occur in cleft constructions, as suggested by Hofmeister (2009) and

<sup>3</sup>For online evidence from event-related potentials that the exhaustiveness reading associated with clefts is not truth-conditional, see Drenhaus et al. (2011).

<sup>4</sup>Regressions, or leftward eye-movements while reading, have been associated with increased processing effort; see, e.g., Frazier and Rayner (1982).

Birch and Rayner (2010). Our results turn out to be consistent only with Hofmeister's 2009 and Birch and Rayner 2010's findings.

### 1.3 Some facts about Hindi word order and clefts

As mentioned above, our major research questions for Hindi are the following: (i) is given-new order the preferred order; (ii) in focus-background clefts, is new-given order the preferred order; (iii) does the *relatum=given* constraint apply in Hindi; and (iv) what is the effect of clefting on sentence comprehension? Because it is difficult to conduct multiple eyetracking studies in India, we took the unusual step of including all these questions into one experiment. This admittedly results in a more complex design that is generally deprecated in psycholinguistic research; however, for logistical reasons it was necessary to include as many of these questions as possible in one experiment.

Before we turn to the experiments, it may be useful to summarize some relevant aspects of Hindi syntax. We consider cleft structures first and then Hindi word order as it relates to our experimental design.

As Srivastav (1991) points out, there are essentially three different ways to realize relative clauses in Hindi: left-adjoined relative clauses, embedded relatives, and right-adjoined relative clauses.<sup>5</sup>

- (2) a. Left-adjoined relative clause  
 jo laRkii khaRii hai vo lambii hai  
 who girl.F.Nom standing is she tall is  
 'The girl who is standing is tall.'
- b. Embedded relative clause  
 vo laRkii jo khaRii hai lambii hai  
 that girl.F.Nom who standing is tall is  
 'The girl who is standing is tall.'
- c. Right-adjoined relative clause  
 vo laRkii lambii hai jo khaRii hai  
 that girl.F.Nom tall is who standing is  
 'The girl who is standing is tall.'

Hindi clefts superficially look like right-adjoined relative clauses. An example of a cleft is shown in (3a). However, it is clear that there is a difference between the two structures. For example, the right-adjoined version (2c) derives from the center-embedded relative clause (2b), but the cleft has no center-embedded correlate, see (3b).

- (3) a. vo laRkii hai jo khaRii hai  
 that girl.F.Nom is who standing is  
 'It is that girl who is standing.'
- b. \*vo laRkii jo khaRii hai hai  
 that girl.F.Nom who standing is is  
 '(Intended) It is that girl who is standing.'

Moreover, as mentioned above, the cleft construction induces an exhaustiveness reading (accompanied by narrow focus on the clefted noun) that is not present in the right-adjoined relative clause shown in (2c). This suggests that the cleft should be analyzed as a construction with the fixed form *N hai jo predicate*, rather than a right-adjoined relative clause derived from a center embedded relative clause. For our purposes, it is only important to note that the cleft construction in Hindi induces the same exhaustiveness interpretation that *it*-clefts do in English and other languages, and that unlike English the segment signaling clefting (*hai jo*, 'is who') occurs after the head noun.

Regarding word order, Hindi is well-known as a head-final language (see Gambhir 1981, Kidwai 2000, a.o.). As in many other languages, there is a subject-first preference; this fact is relevant for our studies, where we look at locative expressions like *X is to the left of Y*. The canonical word order in a locative expression

<sup>5</sup>The glosses used for Hindi-Urdu are follows: M=Masculine; F=Feminine; Nom=Nominative; Obl=Oblique; Gen=Genitive; PROG=Progressive; EMPH=Emphatic.

is grammatical subject, locative expression, and copula, see (4a). In the locative construction, several non-canonical orders are possible, such as locative inversion with canonical (4b) and non-canonical (4c) word order inside the locative phrase. In this paper we focus on the locative inversion shown in (4b); this is because one of the goals of the present paper is to compare our findings for the processing of non-canonical order with those of Hörnig et al. (2005). In this paper, we will refer to the first noun phrase in the canonical sentence (4a) as the subject noun, and the noun in the locative phrase as the non-subject noun.

- (4) a. *duurbiin banduuk=kii daayii taraf hai*  
 binoculars.F.Nom gun.F=F.Gen right side is  
 ‘The binoculars are to the right of the gun.’
- b. *banduuk=kii daayii taraf duurbiin hai*  
 gun.F=F.Gen right side binoculars.F.Nom is  
 ‘To the right of the gun are the binoculars.’
- c. *daayii taraf banduuk=kii duurbiin hai*  
 right side gun.F=F.Gen binoculars.F.Nom is  
 ‘To the right of the gun are the binoculars.’

We describe next two eyetracking experiments using two different designs that address the major goals of this study.

## 2 Experiment 1

This experiment was an eyetracking study that followed the design of Hörnig et al. (2005): participants were asked to mentally visualize the relative positions of three objects as they read two successive sentences that described the relative position of these objects. Participants’ fixations on the words in the sentences were recorded as they carried out the task. The details of the experiment are described next.

### 2.1 Participants

Thirty-two students at the Centre of Behavioural and Cognitive Sciences at Allahabad, India, participated for payment.

### 2.2 Design

The experiment design is summarized schematically in Table 1. As mentioned above, participants were shown two sentences successively and then shown a picture. The two sentences described the relative position of three objects; participants were asked to build a mental representation of the relative position of three objects as they read the sentences. After both sentences had been read, a picture showing two objects appeared on the screen after a latency of 1000 ms, and the participants had to decide, by pressing a button, whether the layout of these objects matched the description provided. An example picture is shown in Figure 1. The 1000 ms latency was used in order to stay as close to the original Hörnig et al. design as possible; it does not affect the main findings in our paper since we analyze the reading times preceding the picture verification task.

The first sentence was either in canonical order or non-canonical order (where non-canonical order refers to the locative inversion shown in (4b)), and the second sentence had the same word order as the first sentence. This parallelism in word order was maintained in order to prevent the task from becoming excessively difficult. Another common factor between the two sentences was that either the subject noun or the non-subject noun in the first sentence was repeated in the second sentence; this repetition of one of the nouns in the first sentence ensured that one noun (either the subject or the non-subject) in the second sentence was previously mentioned or given; the other noun in sentence 2 was therefore new, i.e., not previously mentioned.

We provide examples of the first sentence next. Consider the canonical order sentences first. As shown in (5), the canonical order sentences had two versions; in one version (5a) the target noun that would be repeated in the second sentence (in clefted or non-clefted form) was in subject position, and in the second version it was in non-subject position (5b). In the present example this word is *duurbiin*, ‘binoculars’; hereafter, we will refer to this word in the first sentence as the antecedent of the given word in the second sentence.

**Canonical Sentences**

Sentence 1:

Subject Antecedent - *The binoculars are to the right of the gun.*

Non-Subject Antecedent - *The gun is to the left of the binoculars.*

Sentence 2:

**Given NPI**

**Clefted**  
 There is a chair  
 but it is the binoculars  
 that are to the left of the flag.

**Non-clefted**

There is a chair  
 and the binoculars are  
 to the left of the flag.

**New NPI**

There is a chair  
 but it is the flag  
 that is to the right of the binoculars

There is a chair  
 and the flag is  
 to the right of the binoculars.

**Non-canonical Sentences**

Sentence 1:

Subject Antecedent - *To the right of the gun are the binoculars.*

Non-Subject Antecedent - *To the left of the binoculars is the gun.*

Sentence 2:

**Given NPI**

**Clefted**  
 There is a chair  
 but it is the binoculars  
 to whose right is the flag.

**Non-clefted**

There is a chair  
 and to the right of the binoculars  
 is the flag.

**New NPI**

There is a chair  
 but it is the flag  
 to whose left are the binoculars.

There is a chair  
 and to the left of the flag  
 are the binoculars.

TABLE 1 Schematic design of Experiment 1.

The antecedent was systematically controlled to be the grammatical subject or object in order to balance the design; i.e., we wanted to avoid the asymmetry whereby only subject antecedents or only object antecedents would be used, since such an asymmetry could lead to participants developing a strategy for processing the sentences. There was no theoretical question associated with this manipulation.



FIGURE 1 Example of picture shown on screen in Experiment 1.

- (5) a. Canonical order, Subject antecedent  
 duurbiin            banduuk=kii    daayii taraf hai  
 binoculars.F.Nom gun.F=F.Gen right side is  
 ‘The binoculars are to the right of the gun.’
- b. Canonical order, Non-Subject antecedent  
 banduuk    duurbiin=kii            baayii taraf hai  
 gun.F.Nom binoculars.F=F.Gen left side is  
 ‘The gun is to the left of the binoculars.’

After participants finished reading the first sentence (they indicated this by means of a button press), they were shown the second sentence, which was either clefted or non-clefted, and the first noun phrase was either given or new. The second sentence was the target sentence, i.e., we analyzed reading times only for this sentence.

Each target sentence began by introducing a distractor item (in the example below, this phrase was *There is a chair*); the purpose of this distractor was to set up a contrast that made the cleft sound more natural. Thus, in the cleft condition the distractor phrase was *There is a chair but...*, and in the non-cleft condition (which did not require an explicit contrast set) the distractor phrase was *There is a chair and...* The purpose of providing the distractor phrase in the non-cleft sentence was simply to create as similar a sentence as possible to the clefted sentence. The noun mentioned in the distractor phrase never appeared in the picture shown at the end of a trial.

Shown below are example target sentences for the canonical word order case; the first NP (NP1) following the distractor phrase (*There is an X and/but...*) is either clefted or non-clefted, and is either given or new. The front-slashes (/) indicate the partitioning of regions of interest for purposes of data analysis (the sentences were presented to participants without such a marker).

- (6) a. Canonical order, Non-Clefted, NP1 given  
 [ek kursii            bhii hai aur] / **duurbiin** /            jhanDee=/kii    baayii taraf / hai  
 one chair.F.Nom also is and binoculars.F.Nom flag.M=F.Gen left side is  
 ‘There is a chair and the binoculars are to the left of the flag.’
- b. Canonical order, Non-Clefted, NP1 new  
 [ek kursii            bhii hai aur] / **jhanDaa** /    duurbiin=/kii    daayii taraf / hai  
 one chair.F.Nom also is and flag.M.Nom binoculars=F.Gen right side is  
 ‘There is a chair and the flag is to the right of the binoculars.’

- c. Canonical order, Cleft, NP1 given  
 [ek kursii bhii hai lekin] / **duurbiin** / hai jo / jhanDe=/kii baayii taraf / hai  
 one chair.F.Nom also is but binoculars.F.Nom is that flag.M=F.Gen left side is  
 ‘There is a chair but it is the binoculars that are to the left of the flag.’
- d. Canonical order, Cleft, NP1 new  
 [ek kursii bhii hai lekin] / **jhanDaa** / hai jo / duurbiin=/kii daayii taraf / hai  
 one chair.F.Nom also is but flag.M.Nom is that binoculars.F=F.Gen right side is  
 ‘There is a chair but it is the flag that is to the right of the binoculars.’

The non-canonical counterparts of the first sentence are shown in (7), and the corresponding target sentences are shown in (8).

- (7) a. Non-canonical order, Subject antecedent  
 banduuk=kii daayii taraf duurbiin hai  
 gun.F=F.Gen right side binoculars.F.Nom is  
 ‘To the right of the gun are the binoculars.’
- b. Non-canonical order, Non-Subject antecedent  
 duurbiin=kii baayii taraf banduuk hai  
 binoculars.F=F.Gen left side gun.F.Nom is  
 ‘To the left of the binoculars is the gun.’
- (8) a. Non-canonical order, Non-Clefted, NP1 given  
 [ek kursii bhii hai aur] / duurbiin=/kii daayii taraf / jhanDaa / hai  
 one chair.F.Nom also is and binoculars.F=F.Gen right side flag.M.Nom is  
 ‘There is a chair and to the right of the binoculars is the flag.’
- b. Non-canonical order, Non-Clefted, NP1 new  
 [ek kursii bhii hai aur] / jhanDee=/kii baayii taraf / duurbiin / hai  
 one chair also is and flag.M=F.Gen left side binoculars.F.Nom is  
 ‘There is a chair and to the left of the flag are the binoculars.’
- c. Non-canonical order, Cleft, NP1 given  
 [ek kursii bhii hai lekin] / duurbiin / hai jis=kii / daayii taraf / jhanDaa / hai  
 one chair.F.Nom also is but binoculars.F.Nom is that=F.Gen right side flag.M.Nom is  
 ‘There is a chair but it is the binoculars to whose right is the flag.’
- d. Non-canonical order, Cleft, NP1 new  
 [ek kursii bhii hai lekin] / jhanDaa / hai jis=kii / baayii taraf / duurbiin / hai  
 one chair.F.Nom also is but flag.M.Nom is that=F.Gen left side binoculars.F.Nom is  
 ‘There is a chair but it is the flag to whose left are the binoculars.’

After participants had indicated that they had finished reading the second sentence, the picture shown in Figure 1 was presented after an interval of one second. Once the participant had indicated that they had finished viewing the picture, the picture disappeared and a question such as (9) appeared on the screen.

- (9) dikhaye gaye chitron=kii aapsii sthiti varnan=ke anusaar hai  
 shown were images.M=F.Gen respective position.F.Nom description.M=M.Obl.Gen according is  
 ya nahiiN?  
 or not  
 ‘Does the layout provided here correspond to the description provided?’

If the layout shown was incorrect given the sentences, the participant was expected to respond to the question with a *No*. There were an equal number of correct no and yes responses.

The noun phrases used in the experiment were common nouns such as names of fruits, animals, musical instruments and everyday objects. The positions of these objects were described in terms of prepositions such as ‘to the left of’, ‘to the right of’, ‘above’, ‘below’ and combinations of these. The prepositions were balanced across items. The full list of items is available from the authors.



To summarize, the experiment had three factors, each with two levels: (a) Clefting of first NP; (b) Givenness of first NP (given/new); and (c) Word order (canonical or non-canonical). In addition, although this was not theoretically interesting, in order to maintain balance in the items, the antecedent of the given NP was either the subject or non-subject of the second sentence. 48 sets of items were constructed. There were no fillers because (a) Hörnig et al. (2005) also used no fillers and we wanted to stay close to their design, and (b) pre-testing showed that participants found the task quite difficult, so including fillers would have made the task intractably hard for them (it was probably for this reason that that Hörnig and colleagues also did not use fillers in their study).

### 2.3 Method

Participants were seated 55 cm from a 17" color monitor with  $1024 \times 768$  pixel resolution. The eyetracker used was an IView-X (SensoMotoric Instruments) running at 500 Hz sampling rate, 0.025 degree tracking resolution,  $< 0.5$  degree gaze position accuracy. Participants were asked to place their head in a frame and to position their chin on a chin-rest for stability. The angle per character was 0.26 degrees (3.84 characters per degree of visual angle).

Participants were asked to avoid large head movements throughout the experiment. A keyboard was used to record button responses. The presentation of the materials and the recording of responses was controlled by two PCs running proprietary software (the software used was Presentation, and SensoMotoric Instruments' own software for eyetracker control).

At the start of the experiment the experimenter performed a standard calibration procedure, which involves participants looking at a grid of thirteen fixation targets in random succession in order to validate their gazes. Calibration and validation were repeated after every 10–15 trials throughout the experiment, or if the experimenter noticed that measurement accuracy was poor (e.g., after large head movements or a change in the participant's posture).

### 2.4 Dependent measures

In reading studies, several dependent measures are usually presented. Some of the commonly discussed dependent measures and their interpretation as regards word and sentence comprehension processes are as follows. *First fixation duration* (FFD) is the time elapsed during the first fixation during first pass (the first encounter with a region of interest as the eye traverses the screen from left to right), and has been argued to reflect lexical access costs (Inhoff 1984). *Gaze duration* or *first pass reading time* (FPRT) is the summed duration of all the contiguous fixations in a region before it is exited to a preceding or subsequent word; Inhoff (1984) has suggested that FPRT reflects text integration processes, although Rayner and Pollatsek (1987) argue that both FFD and FPRT may reflect similar processes and could depend on the speed of the cognitive process. *First-pass regression probability*, the probability of the eye making a leftward saccade to a previous word during first pass, is another measure that is sometimes used as an index of processing difficulty. *Right-bounded reading time* (RBRT) is the summed duration of all the fixations that fall within a region of interest before it is exited to a word downstream; it includes fixations occurring after regressive eye movements from the region, but does not include any regressive fixations on regions outside the region of interest. RBRT may reflect a mix of late and early processes, since it subsumes first-fixation durations. *Re-reading time* (RRT) is the sum of all fixations at a word that occurred after first pass; RRT has been assumed to possibly reflect the costs of late-stage processes (Clifton et al. 2007, 349), and recent work suggests that it may be informative about retrieval costs (Gordon et al. 2006, 1308, Vasishth et al. 2008, Vasishth et al. 2010). Re-reading time is usually computed including zero re-reading times, an issue we return to later. Another measure that has been invoked in connection with late-stage processes is *regression path duration*, which is the sum of all fixations from the first fixation on the region of interest up to, but excluding, the first fixation downstream from the region of interest. Finally, *total reading time* (TRT) is the sum of all fixations on a word.

In Experiment 1, we present results based on regression probabilities and re-reading times because none of the early measures showed any effect; note that previous work (Birch and Rayner 1997) also found effects due to clefting in these measures. We follow the literature in presenting re-reading times including zeros, although the distinction (including or excluding zeros) is moot in the present case because the proportion of

re-reading was very high; the results do not change substantially regardless of whether we include zeros or not.<sup>6</sup>

## 2.5 Statistical analysis

A linear mixed model (LMM) was fitted to the data, with crossed random intercepts for participants and items, and with the factors as orthogonally coded predictors (fixed factors). LMMs have several advantages over repeated-measures ANOVA, one of them being that they allow us to take by-item and by-participant variance into account simultaneously, which is an improvement over separate analyses or the calculation of min-F (Clark 1973, Raaijmakers et al. 1999); Baayen (2008) presents further discussion of this issue. Throughout this paper we present coefficient estimates, their standard errors, t- or z-scores (depending on the dependent measure). An absolute t-score of 2 or greater indicates significance at the  $\alpha$  level 0.05. Note also that the t-score is not accompanied by degrees of freedom or p-values. This is because degrees of freedom cannot be computed precisely in LMMs (Baayen 2008).

The statistical analyses on reading times were carried out on log-transformed values. We carried out log-transformed analyses in order to achieve approximately normal residuals.

## 2.6 Results

Using multiple regression models and orthogonal contrast coding, we examined the following effects: word order, clefting, givenness status of NP1, the interaction of NP1's givenness status with clefting, the interaction of word order with givenness, and the interaction of clefting with word order.

The relevant regions of interest were the first noun phrase in the second sentence (abbreviated below as NP1), and the integration region. Note that the first noun phrase is the one that occurs after the distractor phrase (e.g., after the phrase *There is a chair...* in Table 1); this would be the subject noun in canonical order and the non-subject noun in non-canonical order. The first noun phrase region was chosen for analysis because, as described above, previous work (e.g., Birch and Rayner 1997) has shown that the effect of clefting a noun phrase appears at the NP itself.

The integration region is the locative phrase (e.g., *kii daayii taraf*, 'to the right side of') where the relationship between the two referents becomes clear for the first time. This region is theoretically interesting because it is here that the first noun phrase would need to be retrieved in order to construct the spatial relationship between the two noun phrases.

We present the results by region of interest.

### 2.6.1 Region of interest: NP1

#### *First-pass regression probability*

First-pass regression probability was investigated using a generalized linear mixed model with a binomial link function; participants and items were the crossed random factors. The reading time measure (re-reading time) was also evaluated using linear mixed models, also with participants and items as crossed random factors.

	Estimate	Std. Error	z value	p-value
WO	-0.05	0.06	-0.81	0.41
Cleft	0.20	0.06	3.43	<0.01 *
Given <sub>NP1</sub>	0.05	0.06	0.84	0.40
Cleft × WO	-0.11	0.06	-1.88	0.06
Cleft × Given <sub>NP1</sub>	0.14	0.06	2.35	0.02 *
WO × Given <sub>NP1</sub>	-0.03	0.06	-0.46	0.65

TABLE 2 Summary of statistical analyses for first-pass regression probability at the first NP. The asterisk (\*) signifies statistical significance at  $\alpha = 0.05$ .

<sup>6</sup>In Experiment 2, by contrast, the proportion of re-reading times is relatively low (approximately 20%); there, it is important to consider the consequences of including versus excluding zero re-reading times. We discuss this in the Results section of Experiment 2.

The results are summarized in Table 2. Regression probability at the first noun phrase was higher in the clefted versus the non-clefted condition (32.33% vs. 24.9%); an interaction was found between clefting and NP1's givenness status, such that when NP1 was given, clefting NP1 resulted in higher regression probability (34.12%) compared to the non-clefted case (21.67%), while when NP1 was new, the clefted noun had a regression probability (30.55%) that was comparable to the non-clefted case (28.12%). In addition, a marginal interaction was seen between clefting and word order such that the difference between clefted and non-clefted conditions was smaller in canonical order (clefted 31.07%, non-clefted 27.6%), than in non-canonical order (clefted 33.6%, non-clefted 22.19%).

**Re-reading time** The mean re-reading times are shown in Table 3 and the corresponding statistical analyses are in Table 4. We found:

1. *a word order effect*: re-reading time at the first noun phrase was faster in non-canonical word order than canonical;
2. *an effect of clefting*: the first noun phrase was re-read faster when it was clefted versus non-clefted, consistent with the advantage due to clefting found in previous work on clefting;
3. *no effect of givenness status of the first NP*;
4. *an interaction between clefting and word order*: a cross-over interaction was found between the clefting status of NP1 and word order, such that in canonical word order the clefted noun phrase was read faster than its non-clefted counterpart, whereas in non-canonical word order the clefted noun phrase was read slower than its non-clefted counterpart;
5. *an interaction between givenness status of NP1 and clefting*: a cross-over interaction was found between the givenness status of NP1 and clefting, such that when NP1 was given, clefting resulted in slower reading time, and when NP1 was new, clefting resulted in faster reading time;
6. *an interaction between givenness status of NP1 and word order*: a cross-over interaction was found between the givenness status of NP1 and word order, such that a given NP1 was read faster in non-canonical order (compared to canonical order), and a new NP1 was read slower in non-canonical order.

Main effects of Word Order and Clefting				
	Word order		Clefting	
	Canonical	Non-canonical	Non-clefted	Clefted
	715	685	711	670

Clefting and Word Order interaction		
	Non-clefted	Clefted
Canon.	777	658
Non-canon.	644	725

N1 Givenness status, Clefting and WO		
	Given	New
Cleft	704	676
Non-cleft	682	739
Canon.	758	672
Non-canon.	621	744

TABLE 3 Experiment 1. Summary of effects at NP1 (re-reading times in msec).

### 2.6.2 Region of interest: Integration region

In the integration region, we looked at the dependent variable re-reading time.

As summarized in Tables 5 and 6, we found (i) a main effect of clefting, such that the integration region was read faster in the clefted condition; (ii) a main effect of word order, such that non-canonical order was

	Estimate	Std. Error	t value
WO	-0.14	0.05	-3.04 *
Cleft	-0.20	0.05	-4.27 *
Given <sub>N1</sub>	-0.02	0.05	-0.37
Cleft × WO	-0.13	0.05	-2.80 *
Cleft × Given <sub>N1</sub>	0.12	0.05	2.67 *
WO × Given <sub>N1</sub>	-0.12	0.05	-2.64 *

TABLE 4 Summary of statistical analyses for re-reading time at the first NP. The asterisk (\*) signifies statistical significance at  $\alpha = 0.05$ .

read slower than canonical; and (iii) an interaction between word order and givenness, such that when NP1 was given the integration region was read equally quickly in canonical and non-canonical order, but when NP1 was new, the integration region was read faster in canonical order than non-canonical.

Main effects of word order and clefting			
Word order		Clefting	
Canonical	Non-canonical	Non-clefted	Clefted
759	962	903	854
Clefting and Word Order interaction			
	Non-clefted	Clefted	
Canon.	654	875	
Non-canon.	1077	840	
N1 Givenness status, Clefting and WO			
	Given	New	
Cleft	743	964	
Non-cleft	762	1041	
Canon.	757	761	
Non-canon.	750	1164	

TABLE 5 Summary of re-reading time results at the integration region.

	Estimate	Std. Error	t value
WO	0.90	0.06	14.60 *
Cleft	-0.14	0.06	-2.31 *
Given <sub>N1</sub>	-0.08	0.06	-1.33
Cleft × Given <sub>N1</sub>	0.02	0.06	0.38
WO × Given <sub>N1</sub>	-0.16	0.06	-2.59 *
Cleft × WO	0.08	0.06	1.26

TABLE 6 Re-reading time (log ms) in integration region the effect of word order, clefting, clefting-word order interaction, and clefting-NP1-givenness status interaction.

## 2.7 Discussion

To summarize the main results at the first noun phrase, (i) regression probability was higher in the clefted versus non-clefted condition; (ii) regression probability was higher in the clefted versus non-clefted condition when the first NP was given but not when it was new; (iii) re-reading time was faster in the clefted condition compared to the non-clefted one; (iv) re-reading time at the first noun phrase was faster in non-canonical versus canonical order; (v) faster re-reading time was seen in the clefting condition in canonical order but not in non-canonical order (where clefted noun phrases were read *slower* than non-clefted ones); (vi) clefted

nouns were read faster when the first noun phrase was new, not when it was given (when the noun phrase was given, clefts were read slower than non-clefts).

The fact that regression probability increases at NP1 when the noun is clefted suggests increased encoding cost, similar to that found by Birch and Rayner (1997) and Experiment 3B of Birch and Rayner (2010).<sup>7</sup> As discussed earlier, the computationally expensive encoding process includes steps such as building the exhaustiveness interpretation associated with clefts. One important consequence of this encoding process is that the clefted noun receives a higher net activation, which renders it easier to process during second-pass. We see this in the faster re-reading time at the clefted noun. A similar result was also found by Birch and Rayner (2010); in their Experiment 3B, they also find higher regression probability at the clefted noun (which we interpret as encoding cost that results from computing the exhaustiveness interpretation) and easier processing of the clefted noun. Note, however, that Birch and Rayner (2010) found a processing advantage due to clefting in first-fixation durations, gaze duration, total number of fixations, and total reading time; no effect was seen in second-pass (re-reading) time. It is worth pointing out here that their data had relatively few cases of re-reading; their re-reading times are quite low on average (23 and 25 ms for non-clefted and clefted nouns, respectively). Since the vast majority of rereading times consists of zeroes, the statistical model is questionable, as discussed in the appendix.

The interaction between clefting of NP1 and givenness status of NP1 is potentially interesting. When NP1 was new, clefting the noun did not result in higher regression probability, whereas when NP1 was given, clefting resulted in higher regression probability. Assuming that processing a given noun induces a reactivation of the previous mention of that noun in the preceding sentence, clefting a given noun might additionally involve accessing the contents of the preceding sentence simultaneously while building the exhaustiveness interpretation. This could explain the greater processing difficulty at the clefted NP1 when NP1 was given. By contrast, when NP1 is new, since the meaning of the preceding sentence need not be accessed, a floor effect might result, i.e., there may be no significant cost to encoding a clefted new NP compared to a non-clefted new NP. An alternative explanation that we cannot rule out is that the interaction is due to the region preceding NP1; here, the clefted noun is preceded by *lekin*, ‘but’, and the non-clefted noun by *aur*, ‘and’. It could well be the case that this difference in the preceding region is responsible for the interaction. We cannot answer this question definitively in the present work.

The re-reading time at NP1 is also revealing. The faster re-reading time at the clefted NP1 versus its non-clefted counterpart suggests easier processing of NP1 during second-pass. Taken together, the regression probability and re-reading time differences at NP1 show clear evidence for an early encoding cost and subsequent retrieval advantage due to clefting that was hypothesized at the beginning of this paper. An encouraging fact is that these findings are consistent with the existing literature on the processing of clefts, as discussed earlier.

An important finding is that NP1 is read faster in non-canonical order than canonical order; this is the reverse of the usual pattern found in the literature on word-order variation. However, this finding is related to that suggested by Hörnig et al. (2005). As discussed earlier, Hörnig and colleagues provide evidence that non-canonical order sentences are read faster than canonical order sentences when the initial part of a non-canonical order sentence provides important linking information to the preceding sentence; in their case this was so when the initial part of the sentence was given. Hörnig and colleagues refer to this as the *relatum=given* principle. Our results suggest an additional constraint independent of givenness: *relatum-first*. When the *relatum* occurs first, as it does in non-canonical word order, processing is easier, regardless of whether the *relatum* contains a given NP or not. In addition to this finding, the word order and givenness interaction provides support for the *relatum=given* principle. At NP1, a given NP1 was read faster in non-canonical order than canonical, as predicted by *relatum=given*. We should point out here that these “principles” are not intended to have the same level of generality as general principles of grammar; they may well be specific to the processing of relational expressions. Another important point, noticed by a reviewer, is that it is quite possible that the faster re-reading time at NP1 in non-canonical order could be due to increased

<sup>7</sup>As a reviewer correctly points out, we cannot rule out the possibility that the increased regression probability at NP1 may be due to differences arising from the preceding word (*lekin* ‘but’ versus *aur* ‘and’). However, even if this finding turns out to be confounded, our main claim is about facilitation due to clefting at the point of retrieval of the clefted noun.

re-reading of the integration region (which showed longer re-reading times in non-canonical order); if that is the reason for the effect, the *relatum-first* effect would be simply epiphenomenal.

An unexpected finding is that the clefting advantage is reversed in non-canonical order — clefted nouns are harder to read than non-clefted ones. We had expected the advantage due to clefting to apply uniformly in both canonical and non-canonical word order. A plausible explanation for the adverse effect of word order is that in canonical order fewer processing steps are involved than in non-canonical order, and clefting the noun in non-canonical order may overload the processing system. In canonical order, only the exhaustiveness interpretation needs to be computed when a noun is clefted. But in non-canonical order the presence of the relational phrase at the beginning of the sentence also reactivates the memory trace of the preceding sentence; although non-canonical order makes processing easier (as discussed above), clefting the noun could disrupt the attempt to connect the content of the current sentence with the previous one, resulting in an interaction between clefting and word order. This is admittedly a post-hoc and speculative explanation and needs to be followed up with planned experiments.

Finally, clefted nouns are read faster than non-clefted nouns when the noun is new, and the opposite pattern is seen when the noun is given. This finding is consistent with the observation, discussed by Delin and Oberlander (1995), that in focus-background clefts the clefted noun is new. When the clefted noun is given, processing is expected to slow down because clefting a given noun is not the normal case in focus-background clefts. If this interpretation is correct, a clear prediction is that in topic-comment clefts the reverse preference should be seen — given nouns should be processed faster when clefted versus non-clefted, and new nouns should be processed slower when clefted. We intend to test this prediction in future work.

Turning next to the results at the integration region, we find (i) faster re-reading times in the clefted versus non-clefted conditions; (ii) faster re-reading time in canonical versus non-canonical order; and (iii) an interaction between word order and givenness, such that when NP1 was given the integration region was read equally fast in canonical and non-canonical order, but when NP1 was new the integration region was read faster in canonical than non-canonical order.

It is interesting that the integration region is read faster in canonical order compared to non-canonical order, especially when one considers the fact that the opposite pattern was found at NP1 (the *relatum-first* effect). In other words, although non-canonical order can facilitate processing at NP1 (due the *relatum-first* effect), processing the locative phrase (which occurs immediately after NP1) is more difficult in non-canonical than canonical order. This pattern is hard to reconcile with the *relatum-first* pattern we saw at NP1.

The interaction between word order and givenness status of NP1 at the integration region is also quite revealing. The well-known preference for canonical word order is seen when we consider sentences with a new NP1: the integration region is read faster in canonical versus non-canonical order. However, when NP1 is given, the canonical order preference is neutralized. This seems to be at least consistent with the *relatum=given* principle that Hörnig and colleagues propose: in non-canonical order the relational information comes first, and when the relational expression includes a given NP, this may render non-canonical order just as easy to process as canonical order.

Our next goal was to attempt to replicate the two effects we found for Hindi clefts: the increased encoding cost followed by facilitation during later processing. One aim of the second experiment was to determine whether the effect could be replicated in a simpler reading task; a second aim was to determine whether the clefting results in a short-lived activation-increase of the clefted noun — i.e., is the increased activation of the noun limited to the current sentence being processed — or does the clefted noun remain in a permanently high-activation state even when the clefted noun is retrieved in a subsequent sentence. A short-lived activation is predicted by decay-based models such as Dependency Locality Theory (Gibson 2000), and decay- and interference-based models of sentence comprehension such as the cue-based retrieval model of Lewis and Vasishth (2005). In both these classes of models, the memory trace for the prominent (clefted) element would degrade over time, either due to decay or due to interference, or perhaps a combination of both. An alternative possibility is that the noun made more prominent by clefting may remain active in memory, perhaps because it is in a focus buffer that is not subject to decay. Of course, evidence for a short-lived effect of clefting would necessarily be a null result, i.e., not much can be concluded from it. The more conclusive finding would be a

long-lasting facilitation due to clefting.

In addition to the above goals, we were also interested in the interaction of clefting with topic status, specifically whether the topic was a continuation or a shift (these terms are defined below) relative to the preceding sentence. It is well-known that topic continuations are easier to process than topic shifts (Grosz et al. 1995); we were interested in determining whether the processing advantage due to clefting is independent of this information-structural factor or whether there is an interaction. If the clefting advantage occurs independently of topic status, then this suggests that there is something about the syntactic structure itself that controls the activation level of the clefted noun, independent of other information structure factors. If an interaction is found, this would suggest that the activation-boost afforded by clefting is not an automatic reflex of the syntactic structure, but rather can be counteracted by information-structural (extra-syntactic) constraints like topic status (because topic status is dependent on preceding context). A priori, we would expect that clefting would facilitate processing in both types of topic, but that processing topic shifts would be more difficult overall; in other words, we expect a main effect of clefting, and a main effect of topic status, but no interaction. We describe the details of this experiment next.

### 3 Experiment 2

#### 3.1 Participants

Thirty two native speakers of Hindi participated for payment at the Centre of Behavioural and Cognitive Sciences, Allahabad, India.

#### 3.2 Design

Each trial consisted of three sentences. The first was a context sentence that introduced two referents; this context sentence always had a masculine and a feminine noun (the order of occurrence of masculine and feminine gender was counterbalanced across items). An example is shown in (10).

- (10) ek aadmii sarak=ke kinaare ek aurat=ke saath kharāa thaa  
 one man.M.Nom road.F=M.Obl.Gen near one woman.F=M.Obl.Gen with standing was  
 'A man was standing by the street corner with a woman.'

Following standard definitions for Hindi and Urdu (e.g., Kidwai (2000), Dwivedi (1994), Butt and King (1997)), the first noun of the context sentence was considered the topic. The context sentence was followed by a second sentence, which was either in canonical (SOV) or non-canonical (OSV) word order, and the first noun phrase in this sentence was either clefted or non-clefted. When the second sentence was in canonical order, the topic was a continuation of the preceding context sentence (a topic continuation); and when the sentence was in non-canonical order, the topic was a considered to have been shifted (a topic shift).

An example of the second sentence in canonical order is shown in (11). The clefting manipulation in the present experiment differs from that in the first experiment in two important respects. First, the noun phrase (which is either clefted or non-clefted) is not preceded by a distractor phrase; this has the consequence that the only way that the reader can infer that a contrast is intended for the first noun is by fixating on the first noun and then parafoveally processing the two words following the noun, or by reading the two words following the clefted noun. Second, in the canonical order sentence shown in (11), the topic of the sentence is the same as the topic of the first context sentence. We will refer to this as a topic continuation (Grosz et al. 1995).

- (11) {aadmi hai jo / aadmi} aurat=ko dekh rahaa thaa  
 man.M.Nom is who man woman.F=Acc see PROG was  
 'It is the man who / The man was looking at the woman.'

By contrast, the non-canonical word order version of the second sentence, shown in (12), has a topic shift: the new topic is the object of the preceding sentence.

- (12) {aurat hai jisko / aurat=ko} aadmi dekh rahaa thaa  
 woman.F.Nom is who.Acc woman.F=Acc man see PROG was  
 (Approximate translation) '(It is) the woman (whom) the man was looking at.'

The second sentence is followed by a third sentence; an example is shown in (13). This sentence begins with a pronoun *vo*, 'he/she'; in Hindi, such pronouns are unspecified for gender. Further on in the sentence, an adjective is introduced that disambiguates the pronoun (adjectives agree with the subject in gender marking); the disambiguation is either to the subject or the object of the second sentence. For example, in (13), masculine marking *-aa* on the adjective *moṭ-aa*, 'fat-masc' and the auxiliary *th-aa*, 'was-masc' signals that the person who being described as fat is the man. The corresponding feminine marker is *-ii*.

- (13) *vo mere vicaar=me bahut=hii moṭ-aa/-ii th-aa/-ii*  
 he/she my opinion=in much=EMPH fat-M/-F was-M/-F  
 (Lit.) 'He/she, in my opinion, is very fat'.

After the triplet of sentences was read, a yes/no question was presented to probe the participants' comprehension of the sentences. There were 48 items, which are available from the authors. Here, too, the experiment items were presented without fillers; although this is not the standard practice in psycholinguistic research, logistical constraints associated with running this experiment in India forced us to keep the experiment as short as possible.

### 3.3 Method

The same method was used as in Experiment 1: fixations were recorded as participants read the sentences. Each sentence was presented separately; i.e., each trial began when the first sentence was presented, and then the participant signaled with a button-press that he/she has finished reading it; the button-press resulted in the first sentence being removed from the screen and the second sentence being presented; and so on. Participants had to fixate on a circle on the left edge of the screen to trigger the display of a sentence on the screen; this ensured that they began reading from the left edge of the sentence.

### 3.4 Predictions

The predictions in this experiment are as follows. First, clefting the noun in the second sentence should result in a processing advantage later in the sentence, i.e., at the verb, when the clefted noun is to be retrieved. This follows from the previous psycholinguistic research on clefting, and the results of experiment 1. Second, if the processing advantage due to clefting interacts with discourse-level factors like topic status, the clefted noun should be easier to process (should be fixated for a shorter duration) when it is a topic continuation than when it is a topic shift; alternatively, if the syntactic marker for clefting induces a processing advantage (by increasing activation of the clefted noun) regardless of discourse factors, no interaction should be seen between topic status and clefting. Third, if clefting status results in permanently prominent status of the clefted noun, fixation durations at the disambiguating region should be shorter when the antecedent is clefted rather than non-clefted; if topic status has a similar effect, we should see a similar facilitation at the disambiguating region when the antecedent is a topic continuation rather than a topic shift (main effects of clefting and topic status when the ambiguous region resolves to the first noun in the preceding sentence). Alternatively, if the advantage due to clefting and topic status is short-lived (i.e., is subject to decay and/or interference processes as some sentence comprehension theories assume), no advantage due to clefting and topic status should be seen in the third sentence.

### 3.5 Results

We present the results for sentence 2, where the regions of interest were the first NP in the sentence and the integration region (the verb); and for sentence 3, where the region of interest was the adjective where the antecedent was disambiguated. There was no effect of word order and no cleft  $\times$  word order interaction; in other words, we failed to find any advantage due to topic continuation at the first noun in sentence 2.

#### 3.5.1 Sentence 2: first NP (re-reading times)

As in Experiment 1, the clefted NP had faster re-reading times (480 ms) than the non-clefted one (630 ms), replicating the processing advantage due to clefting that we describe in Experiment 1. Re-reading times here exclude zero re-reading times (see Appendix) and all reading times greater than 2000 ms; this resulted in removal of 80% of the data (i.e., re-reading probability was only about 20%). The results of the linear mixed model are presented in Table 7.



No difference was found between the clefted and non-clefted conditions when zero re-reading times were included, and an analysis of re-reading probability also showed no effects. However, note that when re-reading times include zeros, the residuals are remarkably non-normal; this is regardless of whether one log-transforms re-reading time (see Figure 2 in Appendix).

	Estimate	Std. Error	t value
Cleft	-0.14322	0.06883	-2.08 *
WO	0.03537	0.06324	0.56
Cleft×WO	-0.04421	0.06129	-0.72

TABLE 7 Experiment 2. Effect of clefting, word order, and the interaction of the two factors on log re-reading times at the first NP. Crossed random intercepts were fit for items and participants in a linear mixed model.

### 3.5.2 Sentence 2, integration region (first-fixation durations)

Unlike Experiment 1, where only re-reading time showed any effect in the integration region, in Experiment 2 only first-fixation durations showed significant effects. The results of the analyses are summarized in Tables 8 and 9. Overall, the integration region was read faster when the first NP was clefted, replicating the effect found in Experiment 1. Interestingly, however, although we found no main effect of word order, an interaction was found between clefting and word order, such that the facilitation due to clefting was seen only in the non-canonical word order condition (the topic-shift condition); in the canonical word order condition (topic continuation) there was no advantage due to clefting in the integration region.

	Canon.	Non-canon.
Non-cleft	526	573
Cleft	532	496

TABLE 8 Experiment 2. Mean first-fixation durations by word order and clefting conditions at the integration region.

	Estimate	Std. Error	t value
Cleft	-0.046132	0.021565	-2.14 *
WO	0.002713	0.021645	0.13
Cleft×WO	-0.050441	0.021617	-2.33 *

TABLE 9 Experiment 2. Effect of clefting, word order, and the interaction of the two factors on log first fixation duration at the integration region. Crossed random intercepts were fit for items and participants in a linear mixed model.

### 3.5.3 Sentence 3, integration region (re-reading times and re-reading probability)

We analyzed the re-reading times at the adjective separately for canonical and non-canonical order because this allows us to investigate the effect of antecedent resolution while holding word order constant. In the non-canonical word order condition no effect was statistically significant. In canonical order, however, a main effect was found of antecedent resolution, such that the integration region was processed faster when the antecedent was the object rather than the subject. No effect of clefting was found, and no interaction was found between clefting and antecedent.

Here, re-reading times exclude zero re-reading times, that is, we are looking at pure second-pass re-reading time. Excluding zeros and all re-reading times greater than 2000 ms resulted in the removal of 85% of the data (i.e., re-reading occurred only in 15% of trials). Including zeros in re-reading times showed no effects, but, as discussed earlier, the residuals were remarkably non-normal, leading us to question whether the coefficients are interpretable. Re-reading probability at the disambiguating region for canonical and non-canonical word order (analyzed separately) showed no effects of clefting or antecedent type.

	Subj	Obj
Non-cleft	892	677
Cleft	735	565

TABLE 10 Experiment 2. Mean re-reading time by antecedent type and clefting conditions in the integration region (canonical word order).

	Estimate	Std. Error	t value
Cleft	-0.071756	0.065306	-1.10
Ant	0.144902	0.062597	2.31 *
Cleft×Ant	-0.004072	0.062283	-0.07

TABLE 11 Experiment 2. Effect of clefting, antecedent, and the interaction of the two factors on log re-reading time at the integration region in sentence 3, canonical word order. Crossed random intercepts were fit for items and participants in a linear mixed model.

### 3.6 Discussion

We present the experiment design below to save the reader the effort of revisiting the earlier sections of the paper:

(14) Sentence 1:

ek aadmii sarak=ke kinaare ek aurat=ke saath kharāa thaa  
 one man.M.Nom road.F=M.Obl.Gen near one woman.F=M.Obl.Gen with standing was  
 ‘A man was standing by the street corner with a woman.’

Sentence 2 (canonical):

{aadmī hai jo / aadmī} aurat=ko dekh rahaa thaa  
 man.M.Nom is who man woman.F=Acc see PROG was  
 ‘It is the man who / The man was looking at the woman.’

Sentence 2 (non-canonical):

{aurat hai jisko / aurat=ko} aadmī dekh rahaa thaa  
 woman.F.Nom is who.Acc woman.F=Acc man see PROG was  
 (Approximate translation) ‘(It is) the woman (whom) the man was looking at.’

Sentence 3:

vo mere vicaar=me bahut=hii mot-aa/-ii th-aa/-ii  
 he/she my opinion=in much=EMPH fat-M/-F was-M/-F  
 (Lit.) ‘He/she, in my opinion, is very fat.’

To summarize the results of Experiment 2, we find (i) a facilitatory effect of clefting on the first noun phrase in sentence 2, replicating the results of Experiment 1; (ii) an early facilitatory effect of clefting in the integration region; (iii) an interaction between word order and clefting in the integration region, such that the advantage due to clefting was only seen in non-canonical word order (the topic-shift condition); and (iv) in sentence 3, it is easier to resolve the antecedent to the object in sentence 2 than to the subject, when sentence 2 is in canonical order.

This experiment thus replicates our finding in Experiment 1 that clefting yields a processing advantage when the clefted noun needs to be retrieved later in the sentence: the re-reading time on the clefted noun is faster (result (i)) and the processing time at the integration region in sentence 2 is faster when the first NP is clefted (result (ii)). The interaction between clefting and word order (result (iii) above) suggests that in topic continuations clefting yields no processing advantage but in topic shifts, where processing is more difficult, clefting facilitates comprehension. Topic-continuation sentences may be relatively easy to process compared to topic shifts (Grosz et al. 1995); this could have the consequence that any processing advantage due to clefting may be masked the low overall processing difficulty.

The finding that the antecedent is preferentially resolved to the object in canonical word order (result (iv)) suggests a recency preference. This finding runs counter to the claims of Prasad and Strube (2000),

who found an overwhelming preference for resolution to subjects. However, their claims were based on production data (corpus data), and it is quite possible that the constraints on antecedent resolution differ in real-time comprehension.

Finally, since in sentence 3 clefting and antecedent-type did not interact, and since no effect of clefting was found, it appears that the facilitatory effect of clefting may not extend beyond the sentence in which the cleft appears. The facilitatory effect due to clefting appears to be short-lived, as predicted by decay and/or interference accounts of parsing. Note, however, that there was a numerical tendency towards faster reading time at the disambiguating region when the subject or object was clefted (see Table 10). This is suggestive of a slight facilitation in resolving to the subject or object antecedent when the antecedent is clefted. A cautionary note is in order here: since the findings in sentence 3 are null results, we cannot conclude much from them.

#### 4 General Discussion

The two experiments presented here provide several new results regarding the effects of clefting on sentence comprehension. Encoding a clefted noun consumes processing resources, but such deeper encoding renders subsequent processing of the clefted noun easier, e.g., when it is retrieved later on in the sentence. The increased encoding cost due to clefting, as expressed by the higher regression probability in clefted versus non-clefted nouns in Experiment 1, appears to be restricted to cases where the NP is given, not when it is new.<sup>8</sup> We believe that this difference between given and new NPs with respect to clefting may be due to the particulars of the experiment design — when the NP is given, extra processing load is presumably incurred because the relational information described in the preceding sentence must be recalled, and it is in this high-load condition that we observe the greater processing cost associated with clefting an NP. Evidence for this conclusion comes from other work by Birch and Rayner (2010), who have shown similar increases in processing difficulty at the clefted noun (increased regression probability) even when the clefted noun was not explicitly mentioned in the preceding context (their Experiment 3B; although in their experiment, the clefted noun was contextually implied, e.g., referring to a waitress in a context where a restaurant is being discussed).

In addition to encoding cost, clefting also delivers a processing advantage. This is clear from the faster re-reading times at the clefted versus non-clefted noun<sup>9</sup> and the faster re-reading time at the integration region, where the clefted noun must be retrieved to construct relational information (i.e., what is to the left/right of what). Experiment 2 replicated our main findings for the clefting manipulation and demonstrated that even in simpler reading tasks the processing advantage due to clefting can be observed.

Experiment 1 also provides evidence consistent with the *relatum=given* principle that Hörnig and colleagues propose for German: the first NP is read faster in non-canonical order when it is given than when it is new. In addition, at NP1 we also find evidence for a new principle that we call *relatum-first*: NP1 is read faster in non-canonical order versus canonical order (however, as mentioned earlier, this result may be just epiphenomenal). These results suggest that the *relatum=given* principle may indeed be cross-linguistically applicable, and that the default canonical word order preference can be counteracted and even reversed given appropriate context.

Interestingly, we did not find any evidence for the given-new principle in any region of interest (the first NP and the integration region). This may be the case because the effect of given-new ordering emerges only after the entire sentence has been processed. In fact, all the studies mentioned in the introduction that found evidence for given-new ordering looked at total sentence reading times. A clear prediction is that we should also find evidence for a given-new ordering preference if we look at whole sentence reading times in Experiment 1. Indeed, when we considered total reading times on the whole second sentence in Experiment 1, we found a main effect of givenness, such that sentences with given-new ordering were read faster than new-given ( $t=2.14$ ,  $p<0.05$ ). As a reviewer notes, however, it may well be that the effect of given-new ordering occurs as the sentence is being read but is spread out over several regions, with the result that a region-by-

<sup>8</sup>Note that we found no effect of encoding cost in Experiment 2, but this is a null result and as a consequence we cannot conclude anything from it, especially given the relatively small sample size of the study.

<sup>9</sup>A reviewer insightfully notes that we cannot be sure that the re-reading of the noun is a consequence of having started integration; this is a reasonable objection but cannot be investigated with our current data.

region analysis cannot detect it.

Experiment 2 served to replicate our main findings for the effect of clefting in sentence comprehension.

To conclude, the two experiments presented here provide new evidence regarding the role of clefting, word order and given-new order in online sentence comprehension. The present work contributes toward our understanding of how exactly information structure serves to facilitate comprehension in language. Regarding given-new ordering, the present work suggests that the given-new preference is a rather late effect in processing that comes into play after the whole sentence has been processed; variation in given-new ordering does not seem to make a difference during incremental processing. Regarding word order, the experiments show that despite the fact that non-canonical word order is more complex syntactically, it can be easier to comprehend than canonical order if the non-canonical order makes it easier to connect sentences in context. Finally, regarding clefting, although clefting initially involves a greater processing cost, this has the important consequence that the clefted noun is easier to retrieve from memory at a later stage due to its heightened activation.

### **Acknowledgments**

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) through project C6 in the Collaborative Research Center (Sonderforschungsbereich) 632 on Information Structure (2007-2011). We are grateful to Pavel Logačev for carrying out the experiments in India and for significant assistance with preprocessing the raw data. Felix Engelmann and Titus von der Malsburg also provided considerable technical assistance. Thanks to Sabrina Gerth for detailed comments. Finally, we would like to thank the two anonymous reviewers for their constructive, detailed, and helpful reviews. Correspondence should be addressed to vasisht@uni-potsdam.de.

### **5 Appendix: A note on re-reading time**

Most eyetracking studies present re-reading times (e.g., Sturt (2003), Birch and Rayner (2010)) including zero re-reading time. However, when (as in the present experiment) re-reading occurs in a low proportion of the trials, it is worth considering what it means to include zero re-reading time. One problem that arises in such cases is that if the dependent measure contains a large proportion of zeros, the statistical model will have extremely non-normal residuals in the model fit, rendering model assumptions invalid. We illustrate this problem in Figure 2.

A second reason to take re-reading time excluding zeros seriously is that it may reflect later integration processes. Evidence consistent with this suggestion comes from other work (Vasishth et al. 2010) where we have systematically compared re-reading time excluding zeros with self-paced reading data and found that this measure yields similar results to the self-paced reading data.

In cases (such as Experiment 2) where a low proportion of trials contains re-reading events, one issue is the loss of statistical power: with fewer data points it becomes less likely that a true effect would be discovered. Here, re-reading probability could be a useful measure. Although not used standardly, this could be a very informative measure. We therefore include this measure in our analyses where we look at re-reading time.

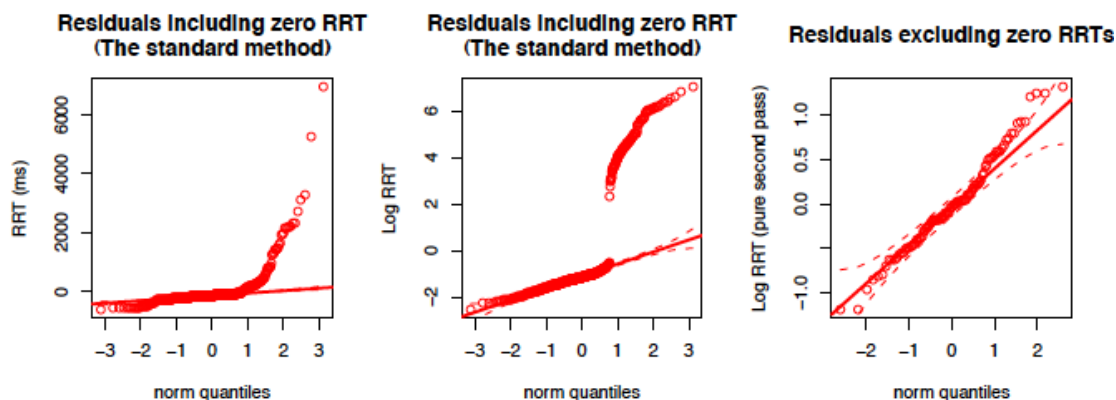


FIGURE 2 Residuals in the linear mixed model when the dependent measure is the standardly used re-reading time with zero re-reading times (raw re-reading time or log-transformed), versus pure second-pass time (log re-reading time excluding zeros).

## References

- Baayen, R. Harald. 2008. *Practical data analysis for the language sciences*. Cambridge University Press.
- Baguley, Thom and Stephen J. Payne. 1998. Given-new versus new-given? An analysis of comprehension times for spatial descriptions. In S. O'Nualláin, ed., *Spatial cognition: Foundations and applications*, pages 317–328. Amsterdam: Benjamins.
- Birch, Stacy and Keith Rayner. 2010. Effects of syntactic prominence on eye movements during reading. *Memory & Cognition* 38(6):740.
- Birch, Stacy L. and Susan M. Garnsey. 1995. The effect of focus on memory for words in sentences. *Journal of Memory and Language* 34(2):232–267.
- Birch, Stacy L. and Keith Rayner. 1997. Linguistic focus affects eye movements during reading. *Memory and Cognition* 25(5):653–60.
- Bredart, Serge and Frédéric Docquier. 1989. The Moses illusion: A follow-up on the focalization effect. *Cahiers de Psychologie Cognitive, European Bulletin of Cognitive Psychology* 9:357–362.
- Bredart, Serge and Karin Modolo. 1988. Moses strikes again: Focalization effects on a semantic illusion. *Acta Psychologica* 67:135–144.
- Butt, Miriam and Tracy H. King. 1997. Structural topic and focus without movement. In *Proceedings of the LFG97 Conference*. Stanford: CSLI Publications.
- Clark, Herbert H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12(4):335–59.
- Clark, Herbert H. and Susan E. Haviland. 1977. Comprehension and the given-new contract. In *Discourse processes: Advances in research and theory*, vol. I. Norwood, NJ.
- Clifton, Charles, Adrian Staub, and Keith Rayner. 2007. Eye Movements in Reading Words and Sentences. In R. V. Gompel, M. Fisher, W. Murray, and R. L. Hill, eds., *Eye movements: A window on mind and brain*, chap. 15. Elsevier.
- Cutler, E. Anne and Jerry A. Fodor. 1979. Semantic focus and sentence comprehension. *Cognition* 7:49.
- Delin, Judy and Jon Oberlander. 1995. Syntactic constraints on discourse structure: The case of it-clefts. *Linguistics* 33(4):465–500.
- Drenhaus, Heiner, Malte Zimmermann, and Shravan Vasishth. 2011. Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics* 24:320–337.
- Dwivedi, Veena. 1994. *Syntactic Dependencies and Relative Phrases in Hindi*. Ph.D. thesis, University of Massachusetts, Amherst, MA.
- Ehrlich, Kate and Phillip N. Johnson-Laird. 1982. Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior* 21:296–306.
- Foraker, Stephani. and Brian McElree. 2007. The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language* 56(3):357–383.

- Frazier, Lyn and Charles Clifton. 2004. Should given information appear before new? Yes and no. *Memory and Cognition* 32:886–895.
- Frazier, L. and K. Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14:178–210.
- Gambhir, Vijay. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi*. Ph.D. thesis, University of Pennsylvania, Philadelphia.
- Gibson, Edward. 2000. Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O’Neil, eds., *Image, Language, brain: Papers from the First Mind Articulation Project Symposium*. Cambridge, MA: MIT Press.
- Gordon, Peter C., Randall Hendrick, Marcus Johnson, and Yoonhyoung Lee. 2006. Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning Memory and Cognition* 32(6):1304–1321.
- Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21:203–225.
- Hofmeister, Phillip. 2009. Encoding effects on memory retrieval in language comprehension. In *Proceedings of CUNY conference*. Davis, CA: University of Davis.
- Hörnig, R., K. Oberauer, and A. Weidenfeld. 2005. Two principles of premise integration in spatial reasoning. *Memory and Cognition* 33(1):131–139.
- Hörnig, R., T. Weskott, R. Kliegl, and G. Fanselow. 2006. Word order variation in spatial description with adverbs. *Memory & Cognition* 34(5):1183–1192.
- Inhoff, Albrecht W. 1984. Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior* 23(5):612–624.
- Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Kidwai, Aeysha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. New York: OUP.
- Lewis, Richard L. and Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29:1–45.
- Morris, Robin K. and Jocelyn R. Folk. 1998. Focus as a contextual priming mechanism in reading. *Memory and Cognition* 26(6):1313–22.
- Prasad, Rashmi and Michael Strube. 2000. Discourse salience and pronoun resolution in Hindi. In A. Williams and E. Kaiser, eds., *Penn Working Papers in Linguistics, Volume 6.3*, pages 189–208. University of Pennsylvania.
- Raaijmakers, Jeroen G.W., Joseph M.C. Schrijnemakers, and Frans Gremmen. 1999. How to deal with the “language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language* 41(3):416–426.
- Rayner, Keith and Alexander Pollatsek. 1987. Eye movements in reading: A tutorial review. *Attention and performance XII: The psychology of reading* pages 327–362.
- Singer, Murray. 1976. Thematic structure and the integration of linguistic information. *Journal of Verbal Learning and Verbal Behavior* 15(5):549–558.
- Srivastav, Veneeta. 1991. The syntax and semantics of correlatives. *Natural Language & Linguistic Theory* 9(4):637–686.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48:542–562.
- Vasishth, Shravan, Sven Bruessow, Richard L. Lewis, and Heiner Drenhaus. 2008. Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science* 32(4).
- Vasishth, Shravan, Katja Suckow, Richard L. Lewis, and Sophie Kern. 2010. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures. *Language and Cognitive Processes* 25(4):533–567.
- Ward, Peter and Patrick Sturt. 2007. Linguistic focus and memory: An eye movement study. *Memory and Cognition* 35(1):73–86.
- Yekovich, Frank R., Carol H Walker, and Harold S Blackman. 1979. The role of presupposed and focal information in integrating sentences. *Journal of Verbal Learning and Verbal Behavior* 18(5):535–548.