Chapter 16
Integration and Prediction in Head-Final Structures

Shravan Vasishth

16.1 Introduction

What determines incremental online processing difficulty? This is a central question in sentence comprehension research. In this paper I will briefly review the history of two phenomena – locality and interference – that together constitute a crucial piece of the incrementality puzzle. The main goal of this paper is to lay out some open research issues regarding these two phenomena.

Incrementality becomes particularly interesting in head-final structures. If a verb occurs after its arguments rather than before, the human sentence parsing mechanism (presumably) faces greater demands than in non-head final structures: it must hold the arguments in memory as well as predictively build syntactic structure until the verb is processed. Indeed, the existence of head-final languages like Japanese has occasionally led researchers (e.g., Pritchett, 1992) to propose a less incremental, head-driven parsing strategy where decisions about structure-building are postponed until the head is encountered.

A great deal of the research on incrementality presupposes a universal parsing mechanism that applies equally to head-final and non-head final languages. However, there is no reason for this to be necessarily true. One alternative is that "the" human parsing mechanism has fundamentally different properties depending on the language being parsed. A third, intermediate position (which could turn out to be the correct one) is also possible: a universally applicable core parsing architecture exists but the processing constraints are conditioned by underlying properties (such as head-finality) of a language. Such conditioning through word order constraints (which derive from grammar) could result in quite different parsing events in head-final languages compared to non-head final ones. It is in this logical space of possibilities that locality and similarity-based interference become relevant.

S. Vasishth (✉)
AQ1 University of Potsdam
e-mail: vasishth@rz.uni-potsdam.de

H. Yamashita et al. (eds.), *Processing and Producing Head-Final Structures*,
Studies in Theoretical Psycholinguistics 38, DOI 10.1007/978-90-481-9213-7_16,

My goal is to discuss some of the recent theories that have been proposed to explain locality and interference effects, and to summarize the cross-linguistic empirical base of these theories. In doing so, I hope to lay out the state of the art on the issue, and outline some of the important open empirical and theoretical issues, especially in relation to the processing of head-final languages. Although the facts I present are well known, the gaps in the theoretical debate on locality and interference effects may not be.

Throughout this paper I use grayed-out boxes to provide additional discussions and more detailed definitions of concepts discussed in the text. These can be skipped by the reader without loss of flow; they are included merely to allow readers new to the area to obtain a quick overview of the theoretical claims mentioned in the text.

16.2 Locality in Sentence Comprehension

Locality is the claim that the distance – however quantified – between a dependent and a head determines integration difficulty at the head. An example is the self-paced reading study by Grodner and Gibson (2005), which showed monotonically increasing reading time at the verb supervised as a function of the distance between the subject nurse and the verb:

(1) a. The nurse supervised the administrator while . . .
 b. The nurse from the clinic supervised the administrator while . . .
 c. The nurse who was from the clinic supervised the administrator while . . .

Chomsky (1965, pp. 13–14) was perhaps the first to propose that the reduced acceptability of sentences containing a “nesting of a long and complex element” arises from “decay of memory.” In related work, Just and Carpenter (1980, 1992) directly address dependency resolution in sentence comprehension in terms of memory retrieval (similar early approaches are the production-system based models of Anderson, Kline, and Lewis (1977)). Just and Carpenter developed a model of integration that involved activation decay (as a side-effect of capacity limitations) as a key determinant of processing difficulty. For example, under the rubric of distance effects, they describe the constraints on dependency resolution as follows (1992, 133):

> The greater the distance between the two constituents to be related, the larger the probability of error and the longer the duration of the integration process.

The explanation for the distance effect in terms of activation decay was taken a great deal further in the Syntactic Prediction Locality Theory or SPLT (see Gibson, 1998, 9 for a historical overview of the connection between decay and

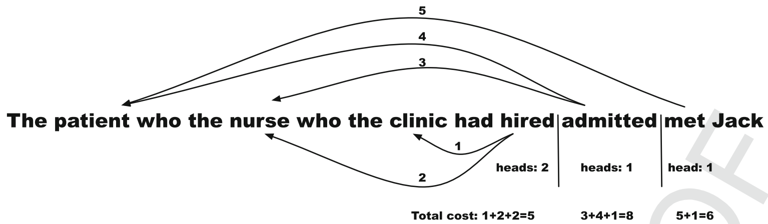

Fig. 16.1 A schematic illustration of DLT's predictions for multiply embedded structures. Integration costs are labeled along the arcs that define the argument-head dependencies, computed by counting the number of intervening discourse referents. Another component of the theory is storage cost; these costs are presented under each verb for illustration. The storage costs is computed by counting the number of heads predicted at each point

distance) and, more recently, the Dependency Locality Theory or DLT (Gibson, 2000). The DLT proposes (among other things) that the cognitive cost of assembling a dependent with a head is partly a function of the number of new intervening discourse referents that are introduced between the dependent and the head; see Fig. 16.1 for an example. In effect, the DLT discretizes the concept of activation decay in the DLT complexity metric (Gibson, 2000, p. 103). The predictions of SPLT and DLT find quite good empirical support from online experiments involving English (e.g., Gibson & Thomas, 1999; Grodner & Gibson, 2005; Warren & Gibson, 2005) and also Chinese (Hsiao & Gibson, 2003) (see the papers on Chinese in this volume). At least one offline study involving Japanese is also consistent with the SPLT's (the precursor of DLT) predictions (Babyonlyshev & Gibson, 1999).

As mentioned above, locality cost is characterized by the DLT in terms of the number of discourse referents intervening between the dependent and the head. One may ask: what is so special about the number of new discourse referents? Why not count the number of intervening syntactic nodes, words, letters, syllables, etc.? The rationale within the DLT is that building discourse referents is computationally costly; independent evidence for this idea comes from studies showing that the accessibility of the intervening discourse referent (as defined by the accessibility hierarchy) can modulate retrieval difficulty (Warren, 2001; Warren & Gibson, 2005).

In direct opposition to the locality hypothesis, Lewis (1996, p. 15) proposed that increased difficulty in resolving a long-distance dependency could at least in certain cases be attributable to syntactic interference. His proposal was that dependency resolution would become difficult if there are multiple intervening potential filler sites that correspond to the gap. This is how he explains the existence of the wh-island constraint (Ross, 1967): **Who does Phineas know a boy who hates the man who saw* _? The sentence is perceived to be ungrammatical because of an upper bound on the number of similar filler sites for the wh-gap. In later work, Lewis and colleagues generalized the interference idea beyond structure-based interference and suggest that the similarity of any kind of feature (not just syntactic) can make processing more difficult. For example, Lee, Nakayama, And Lewis (2005) found evidence for interference due to phonological similarity of case markers in Korean (cf. Vasishth 2003, which failed to find evidence for this idea in Hindi case-marking). The interference idea now exists in four distinct variants: the original conception by Lewis (1996), Van Dyke's retrieval interference model based on Search of Associative Memory or SAM (Van Dyke, 2002) (discussed below), the cue-based retrieval model's interference theory (Lewis & Vasishth, 2005) (which subsumes but goes beyond Lewis' (1996) proposals), and Gordon and colleagues' idea of interference due to feature-similarity of noun phrases (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, Johnson, & Lee, 2006; Gordon, Hendrick, & Levine, 2002). Are locality and interference two alternative explanations or do both the factors operate independently? It is plausible to assume that they are two independent factors, but I will return to this question in the next section.

At this juncture I discuss some well-known but still open empirical problems with the locality hypothesis. A major issue is that locality does not seem to have much empirical support when we look beyond head-final structures in English (indeed, Jaeger, Fedorenko, Hofmeister, and Gibson (2008) have recently presented evidence that the locality constraint may not apply even in English, the language that has the most-attested instances of locality effects).

16.2.1 Counterexamples to Locality: Antilocality

Konieczny (2000) presented an important counterexample to the locality hypothesis. In a self-paced reading study involving German center-embedded relative clauses, he showed that increasing argument-head distance, analogous to example 1 above, resulted in faster reading time at the verb, not slower, as predicted by locality. Konieczny's explanation for the result was that the strength of prediction for the upcoming verb increases if more intervening material is present between the dependent and the head (he calls this the anticipation hypothesis).

Konieczny's finding would not necessarily have been so damaging to the locality idea: the effect he found could have been a consequence of confounding factors such as spillover from the region(s) preceding the verb, or due to differences in word position that result from the locality manipulation. Here, it is worth briefly considering these possibilities.

Mitchell and Green (1978, p. 632) appear to have been the first to discuss the issue of spillover in print; they do so in the context of sentence-final wrap-up effects:

> One further possibility is that the pauses occurred because certain syntactic processes lagged behind the serial presentation of the material. For example, the subjects might have left some of all of the parsing to be carried out at the end of the clause.

Similarly, in later work Mitchell (1984, p. 76) writes:

> In most immediate processing tasks the end of one response measure is immediately followed by the beginning of another, together with a new portion of text. In this situation any uncompleted processing will spill over from one response measure to the next. In others words, certain aspects of processing will be postponed and join a queue or buffer so that they can be dealt with later… Here, the response measure will be influenced not only by the problems in the current display but also by any backlog or processing that may have built up in the buffer.

Spillover is a legitimate concern in Konieczny's study, where the regions preceding the verb were not identical, and therefore differential amounts of spillover from the preverbal regions could well have been responsible for the antilocality effect.

Another possible confound in Konieczny's finding was that the location of the critical word was different in the local versus non-local condition. It could be argued as one reads a sentence from the left to right, reading time becomes faster and faster; this might be responsible for the facilitation Konieczny observed at the verb (note, however, that if this were a systematic speedup in reading regardless of language, one might ask why a speedup was not observed in the Grodner and Gibson (2005) study). The evidence for systematic speedups as a function of word position are based, however, on a misreading of the literature. The key finding was not, as is often claimed, that reading time steadily increases as one proceeds through the sentence but that average reading time is faster in longer sentences. The history of this idea is discussed elsewhere (Vasishth, 2003, pp. 170–185) so I will not repeat it here; rather, I simply note that although the alleged speedup-by-word-position is invoked as fact, as far as I know there exists no empirical demonstration in the literature that speedup generally occurs as one proceeds through a sentence, and there is at least one demonstration (Vasishth, 2003) that shows an absence of such a tendency. In particular, there is currently no evidence that the Konieczny result is a consequence of the alleged speedup-by-word-position. Such a demonstration would also have to explain why the locality manipulation of Grodner and Gibson (2005) is not subject to a word-position effect (i.e., why their results showed a locality effect).

Following Konieczny's work, Vasishth (2003) and Vasishth and Lewis (2006) presented further evidence from Hindi that increasing argument-head distance resulted in faster reading times at the verb; the latter work showed that such a speedup was seen even after spillover was taken into account as a covariate in the data analysis. They proposed an explanation for both locality and antilocality effects based on very general assumptions about forgetting (decay) and reactivation effects in working memory as derived from the ACT-R cognitive architecture (Anderson et al., 2004). Under this view, which is spelled out in a series of articles (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Vasishth, Bruessow, Lewis, & Drenhaus, 2008; Vasishth & Lewis, 2006), instead of defining constraints on retrieval in terms of the number of intervening new discourse referents (DLT's proposal), the cognitive costs of dependency resolution are derived from an independently motivated theory of working memory retrieval: Dependents are retrieved through a content-based retrieval process that relies on cues expressed as feature-value specifications, and probability of correct retrieval and retrieval difficulty are dependent on the dynamic interaction of constraints on processes such as decay, reactivation, interference, partial-cue matching and stochastic noise. Locality effects arise if the dependent decays over time, but antilocality effects can arise if a to-be-retrieved element is reactivated prior to its retrieval at the head (for details, see the articles cited above).

Differences Between DLT and the ACT-R Model

One might ask whether the DLT and ACT-R based model are identical theories. I point out some differences here. First, the ACT-R based model relies on a generally applicable theory of working memory (Anderson et al., 2004); by contrast, DLT instantiates a theory of decay specific to sentence processing – the number of discourse referents are counted to quantify distance. The decay equation in ACT-R would apply equally to (for example) list recall tasks, whereas the DLT instantiation of decay does not have a correlate in working memory tasks outside of sentence comprehension. The second difference is that the ACT-R theory of decay subsumes a reactivation component that is missing in the DLT. The ACT-R decay equation says that an item will decay in memory unless it is reused (reactivated), and the more often it is reactivated, the greater the boost in activation. The reactivation through reuse could account for the antilocality effects that serve as a counterexample to the DLT's predictions (however, not all antilocality effects can be so explained, as discussed below). Independent evidence for reactivation effects comes from Hofmeister (2009). Hofmeister presents evidence from self-paced reading suggesting that in clefting constructions such as *It was John who bought a book*, the noun is retrieved faster at the verb when it is clefted versus non-clefted. He argues that the clefting boosts activation of the noun, resulting in faster retrieval at the verb. The third difference between the DLT and the ACT-R model is that the latter includes an interference theory; by contrast, DLT has no explanation for interference effects.

16.2.2 Surprisal: A New Explanation for Antilocality Effects

An interesting theoretical development in the locality debate was a paper by Levy (2008). He proposed that antilocality effects could be explained by assuming that the material intervening between the dependent and head could serve to sharpen the expectation for the upcoming verb. This sharpened expectation emerges from the elimination of alternative possible parses. The expectation hypothesis – which Levy argues is related to surprisal (Hale, 2001) – was argued to be the explanation for the antilocality effects seen in German (Koniechzny, 2000) and Hindi Vasishth & Lewis, 2006.

Surprisal Theory (Quoted from Boston, Hale, Patil, Kliegl, & Vasishth, 2008)

The idea of surprisal is to model processing difficulty as a logarithmic function of the probability mass eliminated by the most recently added word. This number is a measure of the information value of the word just seen as rated by the grammar's probability model; it is nonnegative and unbounded. More formally, define the *prefix probability* of an initial substring to be the total probability of all grammatical analyses that derive $w = w_1 \cdots w_n$ as a left-prefix (definition 1). Where the grammar G and prefix string w (but not w's length, n) are understood, this quantity is abbreviated by the forward probability symbol, α_n.

$$\text{prefix-probability}(w, G) = \sum_{d \in \mathcal{D}(G, wv)} \text{Prob}(d) = \alpha_n \tag{16.1}$$

(Note: In this definition, G is a probabilistic grammar; the only restriction on G is that it provide a set of derivations, $\mathcal{D}$ that assign a probability to particular strings. When $\mathcal{D}(G, u) = \emptyset$ we say that G does not derive the string u. The expression $\mathcal{D}(G, wv)$ denotes the set of derivations on G that derive w as the initial part of larger string, the rest of which is v. (See Charniak, 1993; Jurfsky Martin, 2000; or Manning & Schütze, 2000 for more details on probabilistic grammars.)

Then the surprisal of the n^{th} word is the log-ratio of the prefix probability before seeing the word, compared to the prefix probability after seeing it (definition 2).

$$\text{surprisal}(n) = \log_2 \left(\frac{\alpha_{n-1}}{\alpha_n} \right) \tag{16.2}$$

As the logarithm of a probability, this quantity is measured in bits.

Consider some consequences of this definition. Using a law of logarithms, one could rewrite definition 2 as

$$\log_2(\alpha_{n-1}) - \log_2(\alpha_n)$$

But on a well-defined probabilistic grammar, the prefix probabilities α are always less than one and strictly nonincreasing from left to right. This implies that the two logarithms are to be subtracted in the opposite order. For instance, if a given word brings the prefix probability down from 0.6 to 0.01, the surprise value is 4.09 bits.

Intuitively, surprisal increases when a parser is required to build some low-probability structure. The key insight is that the relevant structure's size need not be fixed in advance as with Markov models.

An interesting prediction of Levy's view is that English would also show antilocality effects. Indeed, Levy quotes a self-paced reading study by Jaeger and colleagues (also see the further studies in Jaeger et al., 2008) which, contrary to Grodner and Gibson's study, confirmed this prediction. One novel aspect of Jaeger and colleagues' study was that they tried to bring spillover under experimental control (as opposed to statistical control, as Lewis and Vasishth (2005) did). They manipulated dependent-head distance while holding the pre-verbal region constant; i.e., they tried to ensure that differential spillover from the pre-verbal region would be minimized (this assumes that spillover only occurs from the $n-1$th word to the nth word; although this is an assumption that does not hold in general, the design is a significant improvement on earlier work).

Of course, the expectation account fails to explain the Grodner and Gibson locality effects in English; one possibility is that the Grodner and Gibson locality effect could be due to spillover. Thus, the story comes full circle. We began with the assumption that the locality hypothesis holds and that any demonstrations of antilocality could have been due to confounding factors such as spillover; with the Jaeger et al work, it appears that locality effects found in the literature may be due to confounding factors such as spillover and positional differences. Jaeger et al. have made significant advances in bringing the positional confound under experimental control, although they are still of course forced to control for spillover statistically.

A central problem is that we have no theory of spillover. The current assumptions, discussed by Vasishth and Lewis (2006) and Jaeger et al. (2008), are simply that spillover occurs and that different stimulus items and participants may display different amounts of spillover; it is therefore taken into account as a covariate in the statistical data analysis. A better approach would be to flesh out Mitchell and Green's original formulations into a process model of spillover; I believe such a model will help us address its role in reading experiments better than the current methodology. Once such a theory exists, its predictions could be taken into account in a more systematic manner in (implemented) models of parsing.

16.2.3 Is the Locality Effect an Illusion?

Given the recent English findings of Jaeger and colleagues, one might be inclined to reject locality altogether as a constraint. There are at least two problems with dismissing the locality effect. First, Van Dyke and Lewis (2003) demonstrated the existence of a locality effect while experimentally controlling for the spillover confound. They conducted a self-paced study involving sentences such as 2. One factor was ambiguity (presence/absence of the sentential complement that), and another was distance between an argument (here, the noun *student*) and verb (*was standing*).

(2) a. The assistant forgot that the student was standing in the hallway.
 b. The assistant forgot the student was standing in the hallway.
 c. The assistant forgot that the student who knew that the exam was important was standing in the hallway.
 d. The assistant forgot the student who knew that the exam was important was standing in the hallway.

The ambiguity manipulation ensures that reanalysis takes place at *was standing* – the NP *student* must be reanalyzed as the subject of a sentential complement rather than the object of *forgot*. The distance manipulation ensures that the reattachment of the NP as subject of *was standing* is affected by locality. The reanalysis requires an integration between the verb and the argument, which is either near or distant from the verb. Consequently, if a significantly greater reanalysis cost is observed in the intervening-items conditions 2 c,d than in the non-intervening-items conditions 2 a,b, this would be a locality effect, and it would be independent of spillover confounds because the comparison is no longer a direct one between conditions with differing regions preceding the critical verb. The interaction was in fact observed in the Van Dyke and Lewis study, suggesting that locality can affect processing.

Second, a puzzling asymmetry regarding the locality effect was found by Vasishth, Suckow, Lewis, and Kern (submited) between English and German in a series of experiments. They carried out self-paced and eyetracking studies investigating double center-embedding structures in English and German and found that, as predicted by the DLT and its earlier variants (Gibson & Thomas, 1999), omitting the middle verb in a double center embedding can result in facilitation in processing (compare the ungrammatical 3 b with the missing middle verb with its grammatical but harder to process counterpart 3 a). Vasishth et al. (submitted) generally found faster reading times at the final verb and the region following it in the ungrammatical condition.

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

DLT's explanation for the facilitation in the ungrammatical English structure depends on locality assumptions: the long-distance argument-head dependency associated with the middle verb involves the greatest integration cost compared to the other verbs, resulting in the parser forgetting the prediction for the VP corresponding to the second verb. This counterintintuitive result is an important piece in the locality puzzle.

Interestingly, however, contrary to the DLT's prediction German does not present the pattern seen in English: omitting the middle verb renders processing *more* difficult. German examples analogous to the English ones above are shown in 4.

(4) a. Der Anwohner, den der Wanderer, den der Pförtner suchte,
The resident that the hiker that the doorman searched-for
störte, verarztete den Verletzten.
disturbed tended-to the injured-person
'The resident that the hiker that the doorman was looking-for disturbed tended to the injured person.'

b. *Der Anwohner, den der Wanderer, den der Pförtner suchte,
The resident that the hiker that the doorman searched-for
verarztete den Verletzten.
tended-to the injured-person
'The resident that the hiker that the doorman was looking-for tended to the injured person.'

Here, one might object that the obligatory presence of commas in German relative clauses might provide cues for remembering the serial position of the verb, resulting in facilitation of the grammatical structure, whereas in English the fact that commas did not delimit the relative clauses might be the reason that English speakers cannot process the grammatical structures quite as easily. However, a further English experiment was carried out that included commas in the English sentences; this study showed that English speakers persist in forgetting the middle verb (see Vasishth et al., submitted for details). Based on these studies, Vasishth and colleagues hypothesized that German speakers are able to maintain predictions of upcoming verbs much better than speakers of a non-head-final language like English. In other words, the absence of the locality effect in German seems to have its source in some other factor: The high frequency of verb-final structures in German could result in German speakers being more practised than English in maintaining the prediction of the upcoming verb (see Engelmann & Vasishth, 2009 for a connectionist model correctly predicts the English and German asymmetry). If this language-based difference turns out to be valid, a language-independent account of locality/antilocality effects – in particular, one that applies to both head-final and non-head-final languages – is unlikely to furnish a complete explanation.

16.2.4 Working Memory Structures and Locality: An Unexplored Puzzle

A further open question relating to the locality/antilocality and interference literature is that the working memory mechanisms that have been proposed by models like the DLT and the cue-based retrieval model are at odds with other related work in the working memory and language processing literature. Using speed-accuracy tradeoff and eyetracking studies, McElree and colleagues (McElree, 2000, 2006; McElree, Foraker, & Dyer, 2003; Foraker & McElree, 2007) have argued that increasing dependent-head distance does not affect its accessibility (the speed of the dependency resolution process), but does affect the availability (the probability that the dependent would be retrieved). If McElree and colleagues are right, locality and interference effects (which are reflected in longer reading times) are due to availability, not accessibility. However, none of the current sentence comprehension models draw out the consequences of this presumed difference between availability and accessibility. In fact, the working memory theory of ACT-R, although similar to McElree and colleagues' in assuming a content-addressable architecture, directly contradicts the availability-accessibility distinction because it assumes a close relationship between probability of retrieval and retrieval latency: they are inversely related. If an item has a lower retrieval probability than another item, its activation will on average will be lower; since retrieval latency depends on activation, on average retrieval latency will be higher.

In essence, the account by McElree and colleagues is a fourth theory of dependency resolution, and it predicts neither locality nor antilocality effects. Their account would ascribe locality effects in reading studies to reduced availability as distance is increased (McElree et al., 2003). It is difficult to test their claims relative to the other theories using reading-time and eyetracking studies because the key differences (availability versus accessibility) are not directly measurable in such studies (cf. Foraker & McElree, 2007). However, even in the theoretical literature there are no hypotheses about what the locality and antilocality effects index (availability or accessibility). I believe this is a critical gap in the locality debate.

16.2.5 Locality: A Summing Up

To sum up the discussion so far, both locality and antilocality effects have been attested in English, and mainly antilocality effects have been seen in German and Hindi (cf. (Sommerfeld, Vasishth, Logačev, Baumann, & Drenhaus, 2007) and (Drenhaus, Vasishth, Wittich, & Patil, 2007) for evidence for locality effects in German). The candidate theories explaining locality effects are the DLT and the cue-based retrieval (ACT-R based) model, and at least two competing explanations for antilocality effects have been offered: the ACT-R based model and surprisal. Unfortunately, the empirical data do not match any of the theories: DLT uniformly fails to explain antilocality effects, the ACT-R model fails to explain the antilocality effects Jaeger and colleagues found in English, and all three theories are unable to explain the asymmetry of the missing verb effect (Vasishth et al., submitted), which seems to be grounded in structural frequency differences.

It is important to note here that the predictions of Levy's expectation theory are not so easy to determine, and although Levy relates the idea to surprisal, it is not clear to me that surprisal and expectation theory predict the same things. The predictions of surprisal as a scalable computational theory of sentence processing difficulty have been investigated recently in several recent articles (Boston, Hale, Patil, Kliegl, Vasishth, 2008; Boston, Hale, Kliegl, & Vasishth, 2008; Demberg & Keller, 2008); all these articles rely on treebank-based estimates of surprisal. The predictions of expectation theory are derived quite differently. For example, Jaeger et al. (2008) base their predictions of expectation theory on introspection, not on a computation derived from a probabilistic context-free grammar. There are of course good reasons not to compute surprisal from a treebank corpus: the theoretically interesting structures may occur so rarely in the corpus that a meaningful calculation may not be possible. However, in its present form expectation theory as conceived by Levy does not provide any objective way to derive a prediction (cf. DLT and other computational models).

16.2.6 Solving the Locality Puzzle

What can be done to disentangle the predictions of these competing theories? The most obvious point is that we need to develop non-messy experimental designs that manipulate locality without introducing the word position and spillover confounds; until we can resolve these confounds, we don't really know what the evidence is. Note that bringing position and/or spillover under statistical control is a start but insufficient; these factors need to be brought under experimental control. Second, the debate has centered around only a few languages, a limited range of syntactic structures, and only two methodologies, self-paced reading and eyetracking. Missing is more evidence from a variety of languages, involving different methods, including speed-accuracy tradeoff studies and event-related potentials. It is here that a closer study of head-final languages other than German and Hindi can be informative. Third, in the comparison between DLT, the ACT-R based model and surprisal one detail needs to be noted. The first two models are "backward-looking" theories: they define the constraints on retrieving previously processed or encoded material in working memory. By contrast, surprisal is a "forward-looking" metric: it defines the processing cost of predicting upcoming material (Demberg & Keller, 2008; Levy, 2008). Theoretically, these are two orthogonal classes of

AQ2

explanation; a priori, they are not competing explanations, but rather explain different aspects of processing difficulty. This observation is also discussed in Sommerfeld et al. (2007) and Jaeger et al. (2008). ~~One implication is that the relative contributions of these theories can be evaluated empirically; but this has not yet been done. Such an evaluation will be an important step in improving our understanding of dependency resolution processes~~.

We turn next to another theoretical position that relates closely to the locality issue – similarity-based interference.

16.3 Interference in Sentence Comprehension

Interference is the claim that memory traces with similar feature specifications can render integration more difficult due to the reduced ability to distinguish between the target of integration and the similar non-target items. Extensive evidence exists for similarity-based interference in sentence comprehension: for example, (see Gordon et al., 2001, 2002, 2004, 2006; Lee et al., 2005; Lewis, 1996; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). Interestingly, however, the differences between competing explanations for interference effects have not yet been discussed in the literature, and several misunderstandings about the differences between these interference-accounts persist in published work.

Consider the interference theory proposed by Gordon and colleagues. They assume that the featural properties of noun phrases (NPs) cause interference when the target NP is to be retrieved, e.g., at a verb. For example, consider example 5.

(5) The banker [that praised the barber/Sophie] climbed the mountain just outside of town.

Here, the subject relative clause has a grammatical object that is either a definite description, *the barber* or a proper name, *Sophie*. The claim is that type identity of the object with respect to the subject, *the banker*, renders it more difficult to integrate with the verb. Following up on their previous work, Gordon and colleagues conducted a reading study using eyetracking that provided evidence consistent with this claim: longer reading times (gaze durations, right-bounded reading time and re-reading time) were found in the relative clause (RC) region (marked in square brackets); no effect of similarity was found at the matrix verb climbed. As they put it (Gordon et al., 2006, 1309):

> The finding that interference . . . occurs in close proximity to the embedded and matrix verbs is consistent with the idea that the *similarity-based interference occurs at the time of memory retrieval*, as has also been indicated by work manipulating memory load during self-paced reading.

The emphasis is mine.

The above quote raises the question: what is the precise relationship between the feature specifications of NPs and the *retrieval* process at the verb? Gordon and colleagues assume that featural similarities among the NPs somehow come into play at the moment of retrieval. But it is not clear how or why this happens. In the interference theory proposed by Gordon and colleagues, type similarity of NPs has been invoked as a source of interference during retrieval, but what does type-similarity have to do with retrieval? I argue below that Gordon and colleagues' type-similarity assumption distinguishes their theory from other interference accounts.

An alternative interference theory to the proposal by Gordon and colleagues has been offered in Lewis and Vasishth (2005); Lewis et al. (2006); Van Dyke (2007); Van Dyke and Lewis (2003); Van Dyke and McElree (2006). The proposal, which has been referred to as the cue-based retrieval theory of interference, is that the retrieval event at the verb triggers interference *due to the subcategorization requirements of the verb*. For example, a verb such as *kissed* requires a human-referring agent; consequently, integrating the verb with its subject involves a search for a noun phrase that refers to a human entity and is consistent with the agent role. Under this view, the retrieval cues set by the verb preparatory to completing the integration could be responsible for interference effects.

Van Dyke's Retrieval Interference Theory

In Van Dyke's model, retrieval cues are combined into a retrieval probe, which determines the strength of association between each probe cue and memory traces of items in memory. The probability of retrieving an item I_i given probe cues $(Q_1, \cdots, Q_m)$ is a function of the strength of association S between each probe cue Q_j and the features of the memory trace, denoted by $S(Q_j, I_i)^{w_j}$, where w_j is a weighting factor denoting the relative salience of the different cues. Equation (3) formalizes this:

$$P(I_i \mid Q_1 \cdots Q_m) = \frac{\prod_{j=1}^{m} S(Q_j, I_i)^{w_j}}{\sum_{k=1}^{N} \prod_{j=1}^{m} S(Q_j, I_k)^{w_j}} \tag{16.3}$$

This equation quantifies the probability of retrieval of an item I_i given cues $(Q_1, \cdots, Q_m)$ as the proportion of the total strength of association for the item I_i (computed as the product of the strengths of association of the probe cues with the item), with respect to the sum of *all* such strengths of associations for all items I_k.

Van Dyke's interference theory is the only one among the candidate theories that defines interference explicitly in terms of retrieval probability (cf. the discussion about availability versus accessibility above).

A self-paced reading study described in (Van Dyke & McElree, 2006) illustrates the key idea. In this study, participants had to memorize triplets of nouns such as *table*, *sink*, *truck* (see 6), and then read target sentences that had a high-interference or low-interference manipulation. The high-interference target sentences had a critical verb, e.g., *fixed*, that could plausibly take as subject any of the three nouns in the memorized triplet, as well as the NP in the target sentence, *boat*, 6b. The low-interference target sentences had a critical verb, e.g., *sailed*, that could plausibly take as its subject none of the three nouns in the memorized triplet, but it could have as subject the NP present in the target sentence, *boat*, 6b.

(6) a. Memorized set: TABLE SINK TRUCK
 b. High interference
 It was the boat that the guy who lived by the sea fixed in two sunny days.
 c. Low interference
 It was the boat that the guy who lived by the sea sailed in two sunny days.

Compared to baseline conditions that required no list memorization, Van Dyke and McElree found longer reading times at the verb *fixed* compared to *sailed*. To summarize, two distinct explanations have been offered for interference effects. The feature-similarity theory of Gordon and colleagues is about the features common to NPs but not necessarily relevant to the retrieval process per se, while the cue-based retrieval theory is about the features directly triggering the retrieval process.

Clearly, the two theories explain qualitatively different empirical phenomena: the feature-similarity based theory cannot explain the Van Dyke and McElree results, while the cue-based retrieval theory cannot explain the interference effect found by Gordon and colleagues. Apart from the orthogonality in empirical coverage, there is a further important difference between the feature-overlap and cue-based retrieval accounts. Interference due to feature-similarity entails that the ease with which an item is encoded and maintained in memory depends on its featural similarities with previously encountered items (Lewis et al., 2006). There are different ways in which such feature-similarity could affect processing adversely: The features representing the new item could partially overwrite features of the old items, rendering the older items harder to maintain in memory; or the new item could compete with old items for common features, making it harder to encode the new item; or the new item could entirely displace older items from memory (Jonides et al., 2008; Lewis et al., 2006) (cf. Oberauer & Kliegl, 2006).

When we consider that the feature-overlap theory implies the existence of interference effects even before retrieval occurs, a natural question arises: can the interference effect observed by Gordon and colleagues plausibly be attributed to the encoding stage (during the processing of the second NP) rather than at the retrieval stage (during processing at the verb)? Gordon and colleagues' findings could equally be explained as interference during encoding: they analyze reading times over the entire relative clause region, which contains both the second NP and the verb, making it impossible to determine whether the effect was occurring at the NP or the verb. Indeed, in earlier work (2004) they considered the possibility that locus of interference effects may occur even earlier than during the retrieval event; however, they concluded that the 2004 experimental manipulations "do not allow us to ascertain definitively the locus of interference within working memory" (2004, p. 112). This difficulty in isolating the locus of interference also holds for their 2006 experiment discussed above.

Thus, an important open question is disentangling the claims of the two classes of interference theory.

Finally, Logačev and Vasishth (2009) show that *all* existing interference theories are currently unable to explain the fact that a match between two noun phrases along two dimensions can facilitate rather than hinder processing – what Logačev and Vasishth refer to as similarity-based facilitation. All existing theories of interference incorrectly predict greater rather than reduced difficulty. Logačev and Vasishth propose an explanation for this puzzle based on an assumption about the nature of memory representations: conflicting bindings (Hommel, 1998). They show that, under this representational assumption, not only can the published similarity-based interference facts be explained but also the similarity-based facilitation findings.

16.4 Concluding Remarks

I have summarized some of the recent work on locality and interference effects, especially as they relate to head-final structures.

Regarding interference theories, there appear to be several distinct theories which have – at least in principle – empirically distinguishable impact on processing difficulty. These differences have apparently not been noted in the literature.

Regarding locality, it is clear that current explanations – I have discussed only three in some detail, DLT, the ACT-R-based model, and surprisal – need to be teased apart empirically. As discussed above, surprisal is a "forward-looking" metric and the other two theories mainly refer to retrieval of already processing material; they propose "backward looking" metrics. In principle, they are orthogonal explanations and therefore their impact should be independent. ~~This is a claim that has yet to be verified.~~ ~~Once the backward and forward looking models are combined, we are likely to obtain a better empirical fit to the data on locality and antilocality effects.~~ One related problem is that we do not yet understand whether locality effects are confounded with position-effects and spillover. Another open issue is that we do not understand the implications for theories of locality (and interference) are of McElree and colleagues' claims about retrieval accuracy versus latency. Thus, many important research questions remain open in the locality and interference debates.

All these issues are of great relevance to research on head-final languages. Head-final languages like German and Hindi have played an important role in discovering the limits of locality theories. However, missing in the literature is a broader study of head-final languages in the context of phenomena like locality and interference.

Acknowledgements I am grateful to Ted Gibson, Florian Jaeger, Roger Levy, Richard L. Lewis, Don Mitchell and Julie Van Dyke for discussions about locality and interference theories over the last few years. I also benefitted from many comments from, among others, Mike Tanenhaus, Florian Jaeger and Jeff Runner. I also thank my students Esther Sommerfeld, Pavel Logačev, Umesh Patil and Titus von der Malsburg for extended discussions regarding the issues presented here. Marisa Ferrara Boston contributed valuable comments to an earlier draft of this paper.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson, J. R., Kline, P., & Lewis, C. (1977). A production system model for language processing. In M. Just, & P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Babyonyshev, M., & Gibson, E. (1999). The complexity of nested structures in Japanese. *Language*, *75(3)*, 423–450.

Boston, M. F., Hale, J. T., Kliegl, R., & Vasishth, S. (2008). Surprising parser actions and reading difficulty. *Proceedings of ACL-08: HLT Short Papers* (pp. 5–8). Columbus, OH: The Ohio State University.

Boston, M. F., Hale, J. T., Patil, U., Kliegl, R., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1), 1–12.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

AQ3 Demberg, V., & Keller, F. (2008). ~~*Eye-tracking corpora as evidence for theories of syntactic processing complexity* (Submitted to Cognition).~~

Drenhaus, H., Vasishth, S., Wittich, K., & Patil, U. (2007). Locality and working memory capacity: An ERP study of German. *Proceedings of the AMLaP conference*, Turku, Finland.

Engelmann, F., & Vasishth, S. (2009). Processing grammatical and ungrammatical center embeddings in English and German: A computational model. *Proceedings of international conference on cognitive modeling*, Submitted to ICCM 2009.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357–383.

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

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(6), 1411–1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*, 97–104.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (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.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, 425–430.

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

Hale, J. T. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA.

Hofmeister, P. (2009). Encoding effects on memory retrieval in language comprehension. *Proceedings of cuny conference*. Davis, CA: University of Davis.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulusresponse episodes. *Visual Cognition*, *5*, 183–216.

Hsiao, F. P.-F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*, 3–27.

Jaeger, F. T., Fedorenko, E., Hofmeister, P., & Gibson, E. (2008). Expectationbased syntactic processing: Antilocality outside of head-final languages. *Cuny sentence processing conference*, NC.

AQ4 Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., et al. (2008). The mind and brain of short-term memory. *The Annual Review of Psychology*.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Just,M., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645.

Lee, S.-H., Nakayama,M., & Lewis, R. L. (2005). Difficulty of processing Japanese and Korean center-embedding constructions. In M. Minami, H. Kobayashi, M. Nakayama, & H.Sirai (Eds.), *Studies in language science* (Vol. 4, pp. 99–118). Tokyo: Kurosio Publishers.

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

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*(1), 93–115.

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

Lewis, R. L., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454

Logačev, P., & Vasishth, S. ~~(2009~~). Morphological ambiguity and working memory. In P. de Swart, & M. Lamers (Eds.), *Case, word order, and prominence: Psycholinguistic and theoretical approaches to argument structure*. Berlin: Springer.

Manning, C. D., & Schütze, H. (2000). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111–123.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46). San Diego, CA: Academic Press.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91.

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras, & M. Just (Eds.), *New methods in reading comprehension research*. Hillsdale, NJ: Erlbaum.

Mitchell, D. C., & Green, D.W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of experimental Psychology*.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, MIT, Cambridge, MA.

Sommerfeld, E., Vasishth, S., Logačev, P., Baumann, M., & Drenhaus, H. (2007). A two-phase model of integration processes in sentence parsing: Locality and antilocality effects in german. *Proceedings of the CUNY sentence processing conference*, La Jolla, CA.

Van Dyke, J. (2002). *Parsing as working memory retrieval: Interference, decay, and priming effects in long distance attachment*. Unpublished doctoral dissertation, University of Pittsburgh, PA.

Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology Learning Memory Cognition*, *33*(2), 407–30.

Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285–316.

Van Dyke, J., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*, 157–166.

Vasishth, S. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings*. New York: Garland Press. (Published in the Garland series Outstanding Dissertations in Linguistics, edited by Laurence Horn)

AQ5 Vasishth, S., Bruessow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4)

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

Vasishth, S., Suckow, K., Lewis, R., & Kern, S. (~~submitted~~). *Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures*. ~~(Submitted to~~ Language and Cognitive Processes~~, Oct. 2008)~~

Warren, T. C. (2001). *Understanding the role of referential processing in sentence complexity*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

AQ6 Warren, T. C., & Gibson, E. (2005). ~~Effects of NP-type on reading English cleft~~ *nguage and Cognitive Processes*, ~~89–104.~~

Chapter 16

Query No.	Line No.	Query
AQ1	44	Please provide the complete affiliation details for the corresponding author.
AQ2	539	Please include the reference "Levy (2008)" to the reference list.
AQ3	754	Please update the Ref. Denberg & Keller (2008)"
AQ4	790	Please provide the complete details in "Jonides (2008)".
AQ5	842	Please provide the page range in "Vasishth et al. (2008)"
AQ6	851	Please provide the volume number in "Warren & Gibson (2005)."