Locality and Parsing Complexity

Lars Konieczny[1]

Two experiments, an off-line acceptability judgment study and an on-line self-paced reading experiment, were conducted to tackle the question of locality-based preferences in sentence processing. The material consisted of German verb-final sentences containing a relative clause that was either host adjacent or extraposed. While the off-line data seem to reflect locality based integration cost, the on-line data revealed a highly different picture. It is argued that, in the present constructions at least, locality is relevant for production rather than perception. In perception, heads can be anticipated through incremental integration of their arguments.

KEY WORDS: word-order; extraposition; anticipation; locality; parsiyol.

INTRODUCTION

Parsing performance is widely recognized as being determined by the distance or locality of linguistic entities and their dependents. Virtually every parsing model in the last decade employs some form of locality preference [like *right association* (Kimball, 1973), *local attachment* (Fodor and Frazier, 1980), *late closure* (Frazier, 1987), and *activation decay* (Stevenson, 1993, Vosse and Kempen, 2000)]. Among these, there are at least two recent proposals that

[1] Center for Cognitive Science Institute of Computer Science and Social Research, Freiburg University Friedrichstr. 50 D-79098 Freiburg, Germany and SFB "Resource-adaptive Cognitive Processes" Saarland University, Saarbrücken, Germany. Email: lars@cognition.iig.uni-freiburg.de.

attempt to reduce the entire issue of parsing complexity to locality: Hawkins' (1994) *early immediate constituents* (EIC) approach, and Gibson's (1998) *syntactic prediction locality theory* (SPLT). Both accounts provide a detailed metric to calculate on-line parsing complexity, both attribute locality phenomena more or less directly to working memory limitations in the human sentence processing mechanism, and both predict that the processing of more distant dependent entities should be harder. These predictions are supported by a variety of observations, such as corpus counts on word-order preferences, ambiguity resolution preferences, and reading times on unambiguous, but complex, constructions. Although these data seem persuasive, locality-based accounts of on-line parsing complexity generally suggest that processing items toward the end of longer sentences should be harder, since they most often have to be integrated with items further away.

Locality-based predictions stand in sharp contrast to accounts that predict anticipation of subsequent items through integration and projection of previous items. For example, verb-complements can be anticipated through valency or thematic properties of a verb that has been encountered previously in the string (Mac Donald, Perlmutter, & Seidenberg 1994; Konieczny, Hemforth, Scheepers, & Strube 1997). However, if the verb is placed at the end, it might be anticipated through the number and types of arguments preceding it (cf. Konieczny, 1996; Konieczny & Hemforth, 1994). There is a growing body of evidence that prehead integration of arguments takes place incrementally, such that the head (e.g., the verb) does not have to collect previous items kept unattached in a buffer, and integrate them into the structure (Hemforth, Konieczny, & Strube 1993; Bader and Lasser, 1994; Kamide and Mitchell, 1999). On the other hand, thematic assignment is often assumed to take place at the semantic head (cf. Ferreira and Henderson, 1991), i.e., the verb in the given context, which might predict locality effects for head-final constructions. There is also evidence that when subjects read a string of several arguments followed by a verb, they can predict subsequent arguments on the basis of previous ones, probably by anticipating the class of the verb and its further arguments (Kamide, Attmann, & Haywood, 2000). If this is the case, a larger number of arguments preceding a verb might even facilitate its processing.

Locality-based accounts are also challenged by the common observation that reading times usually do not increase toward the end of clauses or when clauses get longer (in fact, reading often even speeds up toward the end of sentences). How can these mutually incompatible findings be explained? A possible answer might be that locality affects parsing complexity only in certain constructions, in which items further downstream cannot be anticipated easily. Such cases might include noncanonical argu-

ment orders, such as English object-extraction RCs, or items that are generally not predicted, such as adjuncts. Another explanation might be that locality determines production preferences, but leaves language perception widely unaffected (Wasow, 1997).

In this paper, two experiments are presented based on the same set of materials made up from simple German sentences without violations of canonical word-order preferences. Example (1) is a simple German verb-final sentence with an embedded RC modifying the direct object NP.

(1) a. Er hat *das Buch, das Lisa gestern gekauft hatte,* hingelegt.
       He has the book, that Lisa yesterday bought had, laid_down.
       "He has laid down the book that Lisa had bought yesterday."

    b. Er hat *das Buch* hingelegt, *das Lisa gestern gekauft hatte.*
       He has the book laid_down, that Lisa yesterday bought had.
       "He has laid down the book that Lisa had bought yesterday."

The RC can be placed directly after its host NP (1a), or it can be extraposed to the end of the matrix clause (1b). Both versions are unproblematic although intuitively, there might be a small preference for the extraposed version in (1b). However, the position of the matrix participle *hingelegt* varies strongly between the two versions: In (1a), it is the final word in the sentence and thus relatively far away from both the subject-NP (Er) and the object-NP (das Buch). If parsing complexity is affected by locality, processing the participle in (1a) should be noticeably harder.

EARLY IMMEDIATE CONSTITUENTS

Hawkins' (1994) metric is based on the idea that processing a phrasal unit is more expensive if it takes longer to identify its immediate constituents (ICs). The central concept is the "*constituent recognition domain*" (CRD) of a phrase: starting with the first IC, it includes all nodes that are built until the final IC can be identified. The last IC is identified at the first word that allows its construction, typically the first word of the phrase. According to principle of *Early Immediate Constituents* (EIC, Hawkins, 1994, p. 77) "the human parser prefers linear orders that maximize the IC-to-non-IC ratios of constituent recognition domains." To simplify matters, the IC-to-non-IC ratio can be estimated by the simpler IC-to-word ratio.

In cases of several possible word orders, the IC-to-word ratios of all phrases whose IC-to-word ratios are affected must be considered. The relative clause in example (1) is either adjacent to its host or extraposed to the end of the clause. Two phrases have to be considered: the NP host of the RC and the VP containing the NP host. Following Hawkins, the CRD of the NP

(NP→det N RC) starts with the determiner and ends with the relative pronoun, at which the RC can be constructed. The VP includes the direct object NP and the verb (VP→NP verb). The IC-to-word ratios for the extraposed version are $2/3 = 66.7\%$ for the VP and $3/4 = 75\%$ for the NP (provided that only the verb intervenes between the RC and its host and provided that the relative pronoun allows the construction of the RC phrase marker). The aggregate ratio for the extraposed version is $(2/3 + 3/4)/2 = 70.83\%$, serving as an estimate of the relative over-all quality of the sentence.

The adjacent version is $3/3 = 100\%$ for the NP and $2/8 = 25\%$ for the VP (five words from the RC), resulting in an aggregate of $(3/3 + 2/8)/2 = 62.5\%$. The word-order preference is predicted by the difference of aggregates of both versions: $70.83\% - 62.5\% = 8.33\%$. For the given constellation of number of words in the RC and the potential distance of the RC of its host (one word), extraposing the RC should be advantageous. However, the advantage should disappear with longer distances and shorter RCs.

Hawkins' predictions were (roughly) confirmed in a corpus study on German sentences (Uszkoreit *et al.,* 1998). Extraposition distance strongly affected the proportion of extraposed RCs found in a German corpus: With only one word (potential) distance, about 96% of the RCs were extraposed, about 70% for two words and only few (about 22%) for three words. This result was modulated by the length of the RC, such that the proportion of extraposed RCs was the higher the longer they were, but the effect was less drastic than expected. Although these data can be seen as supporting locality-based predictions, it is still an unresolved issue whether or not such preferences are, in fact, *perceptual* in nature, or whether they reside in the production faculty, as suggested by Wasow (1997). An on-line comprehension advantage of extraposed or nonextraposed RCs, given a variation of lengths, still has to be established.

SYNTACTIC PREDICTION LOCALITY THEORY

More recently, Gibson (1998) has proposed a working memory-based metric of parsing complexity. There are two components: (1) storage for syntactic predictions and (2) integration cost for items being integrated into the current structure. While both components are supposed to be locality-based[2] and to consume energy from the same pool of working memory resources, only integration cost is of interest here, as it is supposed to be

---

[2] Toward the end of Gibson (1998), SPLT is revised such that only integration cost is assumed to be locality based.

reflected by reading time most directly. Integration cost is a monotonically increasing function of the distance of the current item to its previous dependents. Different from EIC, SPLT measures distance in terms of the number of new discourse referents (NDRs) that have been processed since a dependent item has been encountered. Applied to (1), repeated in (2) for convenience, there are two locations where integration cost is affected by the RC extraposition. First, the verb participle (*hingelegt*) must be integrated with the direct-object NP (*das Buch*), crossing three NDRs (Lisa, *gekauft hatte, hingelegt*[3]) if the RC is host-adjacent, but only one NDR (*hingelegt*) in the extraposed case.[4]

(2) a. Er hat *das Buch, das Lisa gestern gekauft hatte,* hingelegt.
$[_{IP}$He $[_{II}$has $[_{VP}[_{NP}$ the book that$_i$ Lisa yesterday t_1 bought had] laid_down]]].
...   I(0)   ...   I(3)

b. Er hat *das Buch* hingelegt*, das Lisa gestern gekauft hatte.*
$[[_{IP}$He $[_{II}$has $[_{VP}[_{NP}$ the book] laid_down]]] $[_{RC}$ that$_i$ Lisa yesterday t_1 bought had]].
...   I(1)   I(1)   ...

Second, the relative pronoun (*das*) is bound to the NP (*das Buch*), crossing either zero NDRs in the adjacent RC (2a) or one NDR in the extraposed RC (2b).

SPLT thus clearly predicts that reading times for the clause final verb should be elevated if it is preceded by a RC and that reading time on the relative pronoun should be higher if the RC is extraposed.

The main goal of this paper is to investigate whether locality affects sentence processing when sentences with canonical argument order are used. The corpus data (Hawkins, 1994; Uszkoreit et al., 1998) suggest that locality does influence word-order preferences in these constructions, but a perceptual effect has yet to be established. If it turns out that integration cost is not affected by locality in these constructions, or that the effect is even reversed, the question must be raised as to what distinguishes these from those cases that allegedly elicit locality effects.

---

[3] It is not quite clear whether past-participle verbs would count as referential entities. Note, however, that the predictions are not affected by this issue.

[4] It might be argued that the subject-NP (*he*) must be integrated with the participle as well, because the participle holds the subcategorization requirements. If so, extraposition should have an even more dramatic effect, since integrating the subject-NP also has to cross an adjacent RC. However, it can be argued that the subject-NP is integrated within the IP headed by the finite auxiliary *had,* leaving only the direct-object NP to be integrated with the participle within the VP. Note, however, that this is irrelevant for the discussion at hand, since we are interested in a locality effect *per se* and not in the size of it.

EXPERIMENT I

Materials and Design

Forty-two subjects read 63 sentences constructed according to a $2 \times 3 \times 3$ (+ 3) design including the factors *RC position, extraposition distance* (3), and *length of the RC* (4). For norming purposes, three sentence versions with no RC were added, keeping the *distance* variation (no vs. short vs. long preverbal PP; see 3a–c). Twenty-one lists were prepared with each list containing an equal number of each of the experimental conditions, but only one version of each sentence. Forty filler items of different sentence types were added to each list. Questionnaires were generated from these lists with items ordered randomly. In the remainder of the paper, the three sentence versions without a RC will be ignored, as they were not included in the statistical analysis.

(3) Factors *RC position* (*adjacent:* RC_1 vs. *extraposed:* RC_2) and *distance:* one word (a) vs. three–four words (b) vs. five–six words c)

    a. Er hat die Rose {RC_1} *hingelegt* {RC_2}, und . . .
       He has the rose {RC_1} laid_down {Rc_2}, and . . .

    b. Er hat die Rose {RC_1} *auf den Tisch gelegt* {RC_2}, und . . .
       He has the rose {RC_1} on the table laid {RC_2}, and . . .

    c. Er hat die Rose {RC_1} *auf den kleinen runden Tisch gelegt* {RC_2}, und . . .
       He has the rose {RC_1} on the small round table laid {RC_2}, and . . .

(4) Factor *RC length:* three–five words (a) vs. six–eight words (b) and nine–eleven words (c)

    a. . . . die Rose . . ., *die wunderschön war,* . . .
       . . . the rose . . ., that was beautiful, . . .

    b. . . . die Rose . . ., *die auffällig schön und farbenprächtig war,* . . .
       . . . the rose . . ., that was remarkably beautiful and colorful, . . .

    c. . . . die Rose . . ., *die auffällig schön gewachsen und ganz besonders farbenprähtig war,* . . .
       . . . the rose . . ., that was remarkably beautifully grown and especially colorful, . . .

Procedure

Participants had to judge the acceptability of sentences presented in a printed questionnaire. The rating technique used was *magnitude estimation* (ME; see Bard, Robertson, & Sorace, 1996). Participants were instructed to provide a numeric score that indicates how much better (or worse) the cur-

rent sentence was compared to a given reference sentence (example: if the reference sentence was given the reference score of 100, judging a target sentence five times better would result in 500, judging it five times worse in 20). Judging the acceptability ratio of a sentence in this way results in a scale that is open-ended on both sides. It had been demonstrated that ME is therefore more sensitive than fixed rating scales, especially for scores that would approach the ends of such scales (Bard et al., 1996).

Hypotheses

EIC

There are two ways to derive predictions for acceptability judgments from EIC (Hawkins, 1994). First, Hawkins (1994) seems to suggest that acceptability corresponds to the *difference* of aggregated IC-to-word ratios for the adjacent and extraposed versions of the same sentence. Judging the difference of two-word order variants implies that the alternative word order must be present somehow to be compared with the given one. This version of EIC predictions will be dubbed the *difference hypothesis* in the remainder of the paper. Second, acceptability might as well directly correspond to the aggregate IC-to-word ratios of the relevant phrases in only the word order given by the sentence. This will be called the *judge-as-is hypothesis.*

Table I shows the aggregated IC-to-word ratios of each design condition, calculated as follows: for RC extraposed cross a short distance, the VP (NP verb) can be identified after three words (det n v), and the NP (det n RC) only after four words (det n v relPro). This is $2/3 = 66.6\%$ for the VP and $3/4 = 75\%$ for the NP, and 70.8%, on average. If a short PP (prep det N) is added to the distance, the ratio for the VP (now: NP PP verb) is now $3/6 = 50\%$; the ratio for the NP is $3/7 = 42.9\%$ yielding an aggregate of 46.45% (note that in the materials, short PPs have either three or four words. The 50.71% for middle distances in Table I is the average aggregate of three and four-word PPs).

Table I. Aggregate IC-to-Word Ratios[a]

	Extraposed RC			Adjacent RC		
	Distance short (%)	Distance middle (%)	Distance long (%)	Distance short (%)	Distance middle (%)	Distance long (%)
RC short	70.83	50.71	37.8	64.29	65.15	63.07
RC middle	70.83	50.71	37.8	60.0	61.61	60.36
RC long	70.83	50.71	37.8	57.69	59.41	58.58

[a] Hawkins, 1994.

Adding two to three words to the distance thus causes a dramatic decrease in aggregated IC-to-word ratios. This is due to the fact that both the VP and the NP are affected by the distance in extraposed cases. The RC-length, on the other hand, plays no role for extraposed RCs.

Turning to adjacent RCs, Table I demonstrates that changes are comparably mild and that only RC length seems to affect IC-to-word ratios, such that longer RCs lead to worse ratios. Distance plays a minor role, since it only affects VP recognition, which suffers more from the intervening RC in this construction. Figures 1 and 2 illustrate the predicted effects of distance and RC length for extraposed and adjacent RCs.
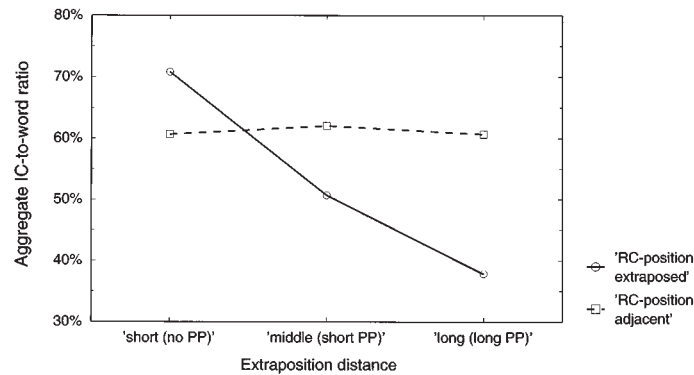
Fig. 1. Mean IC-to-word ratios by levels of RC position (extraposed vs. adjacent) and extraposition distance.
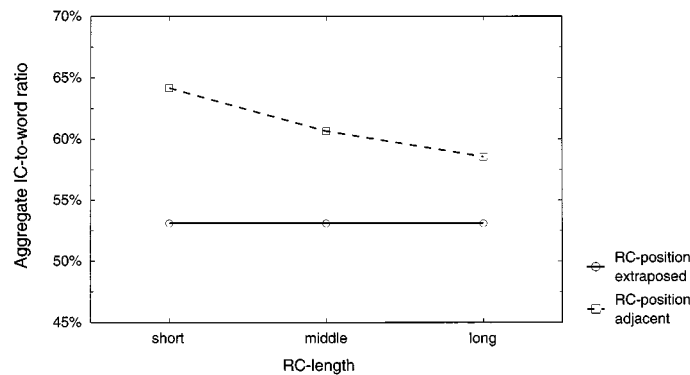
Fig. 2. Mean IC-to-word ratios by levels of RC position (extraposed vs. adjacent) and RC length.

Under the judge-as-is hypothesis, the aggregate IC-to-word ratios should be direct estimates of acceptability judgments. Under the difference hypothesis, however, judgments should be a function of the *difference* of the ratios for extraposed and adjacent RC versions. For example, the acceptability of a sentence with a short RC extraposed over only a short distance might be based on the ratio difference $70.83\% - 64.29\% = +6.54$. The adjacent RC version of the same sentence, on the other hand, would be based on the same difference multiplied by -1, yielding -6.54.

Acceptability judgments are now affected by the quality of the sentence itself and by the quality of its word-ordering alternative. This amounts to the prediction that both the extraposed version and the adjacent version should be affected by varying the length of the RC and the extraposition distance. Therefore, sentences should be rated higher when extraposed RCs are long rather than short and adjacent RCs are short rather than long. Sentences with extraposed RCs should also be judged better if the extraposition distance is short rather than long, whereas sentences with adjacent RCs should be the better the longer the distance would be if the RC had been extraposed.

Under both the difference hypothesis and the judge-as-is hypothesis, EIC predicts reliable position by distance and position by RC-length interactions. Only the difference hypothesis, however, predicts a reliable effect of RC length for extraposed RCs and a reliable effect of distance for adjacent RCs.

Given that only short distances seem to predict an advantage for extraposed RCs in the given materials, a main effect of RC position can also be hypothesized, showing an overall advantage for adjacent RCs.

SPLT

Predictions derived from SPLT (Gibson, 1998) are less straightforward. Integration cost might affect acceptability judgments via processing difficulties at places of long-distance integrations.[5] The regions affected by the sentence variation are (1) the matrix verb participle, whose integration with its argument NP has to cross an adjacent RC, but not an extraposed RC, and (2) the RC pronoun, whose integration with its host NP has to cross the "distance" region (PP and verb participle) if the RC is extraposed, but not otherwise. Verb participle integration should thus be affected by the factor's

---

[5] Memory cost might also influence acceptability, but it is not clear whether memory cost should be considered locality based. Furthermore, it is not clear whether there is memory cost for storage of the matrix verb *participle* at the end of the clause (matrix verbs do not produce memory cost). Note, however, that whether or not storage should be considered does not change the hypotheses.

RC position (adjacent harder than extraposed), length of the RC, if the RC is adjacent (i.e, a two-way interaction is predicted: processing should be harder for longer RCs, if they are adjacent, but not for extraposed RCs), and the distance (verbs following a PP should be harder regardless of RC position). RC-pronoun integration should only be affected by extraposition and distance, in case of extraposition (i.e., a two-way interaction RC position by distance is predicted.).

Note that in SPLT (Gibson, 1998), integration cost increments are caused by crossing new discourse referents rather than words *per se* (although it is not ruled out, in principle, that words count as well). After reviewing the materials, it turns out that lengthening the RC was often, but not always, accomplished by adding NDRs. RCs included at least one NDR (the verb) and up to four NDRs in the long version. For extraposition distance, adding a PP also adds a NDR, but lengthening the PP is mostly done without adding another NDR. Distance effects are, therefore, mainly predicted for short (no PP) vs. middle and long distances (one PP). Table II shows integration cost predictions for the verb participle and the RC pronoun for adjacent and extraposed RCs and short vs. long distances. Note that predictions are based on the simplified assumption that crossing an RC averages to two NDRs.

The relationship of on-line integration cost to overall acceptability of a sentence is somewhat vague. A potential solution could be to sum up (or calculate the means of) integration costs of critical regions, as shown in the two right-most columns in Table II. The sums indicate that short distances should be better than long distances and possibly an interaction of RC position and distance, such that extraposed RCs profit more from short distances than adjacent RCs.

The same kind of estimate can be applied to RC length and PC position (Table III), assuming that a short RC contains, on average, 1.5 NDRs and long 2.5 NDRs and there are, on average, 0.5 NDRs (PPs) preceding the verb.

Only the verb is affected by longer RCs and only for adjacent RCs. Similar to EIC, SPLT thus predicts that sentences with adjacent RCs get worse with increasing RC length.[6,7]

---

[6] This claim is based on the simplified assumption of a linear integration cost function. Using a more realistic sigmoid function might predict smaller additional effects if the verb follows an RC with several NDRs. We ignore this matter for expository reasons.

[7] As pointed out by an anonymous reviewer, there might be alternative ways of calculating global acceptability predictions from SPLT. For instance, acceptability might be estimated more correctly using the aggregate of all local integration cost scores in the sentence, or using the maximum of those. Predictions are likely to vary with these calculations, although both of them would probably correlate with the simplified calculation presented here. Unfortunately, demonstrating this is beyond the scope of this paper.

Table II. Predicted Integration Costs for the Verb Participle and the Relative Pronoun with Short Extraposition Distances (no PP) and Long Distances (PP)

RC-pos	Verb		relPro		Verb + relPro	
	Short	Long	Short	Long	Short	Long
Extrap.	I(1)	I(2)	I(1)	I(2)	2*I(1)	2*I(2)
Adjacent	I(3)	I(4)	I(0)	I(0)	I(3) + I(0)	I(4) + I(0)

Table III. Predicted Integration Costs for the Verb Participle and the Relative Pronoun with Short RCs (1.5 NDRs) and Long RCs (2.5 NDRs)

| RC-pos	Verb		relPro		Verb + relPro	
	Short RC	Long RC	Short RC	Long RC	Short RC	Long RC
Extrap.	I(1.5)	I(1.5)	I(1.5)	I(1.5)	2*I(1.5)	2*I(1.5)
Adjacent	I(3)	I(4)	I(0)	I(0)	*I(3)* + I(0)	*I(4)* + I(0)

Results

The standardized logarithms of the raw scores (z-standardized per participant and multiplied by 20 for expository reasons) were submitted to a $2 \times 3 \times 3$ GLM (MANOVA) for repeated measures. Extremes were eliminated from the data before aggregation for participants or items, resulting in 5.1% missing data before aggregation.[8]

Table IV shows the means of acceptability judgments for each condition. Statistical analysis revealed a strong main effect of EC position: Sentences with adjacent RC were judged much better than extraposed RCs [adjacent: 5.96; extraposed; −7.14; $F_1(1,39) = 65.4$, $p < .001$; $F_2(1,54) = 307.12$, $p < .001$]. There was also a main effect of extraposition distance [$F_1(2,78) = 7.86$, $p < .001$; $F_2(2,108) = 12.34$, $p < .001$], which was due to the fact that sentences with long preverbal PPs were judged worse (no PP: 0.83; short PP: 0.40; long PP: −3.01). No main effect of RC length was found. The relevant predictions referred to two-way interactions of RC position and the length factors. Both turned out to be highly reliable. RC position interacted with both extraposition distance [$F_1(2,78) = 28.4$, $p < 001$; $F_2(2,108) = 58.28$, $p < 001$] and RC length [$F_1(2,78) = 13.23$, $p < .001$; $F_2(2,108) = 6.13$, $p < .01$]. No other interaction attained significance.

---

[8] Extremes were identified per condition by the EXAMINE (PLOT STEMLEAF) procedure in SPSS 9.

Table IV. Mean Acceptability Scores (Standardized Logarithms of Raw ME Scores)

	Extraposed RC			Adjacent RC		
	Distance short	Distance middle	Distance long	Distance short	Distance middle	Distance long
RCshort	−2.28	−10.11	−17.62	5.23	8.7	8.9
RC middle	0.75	−5.14	−11.73	3.04	8.59	7.97
RC long	−1.63	−4.45	−12.13	−0.14	4.74	6.55
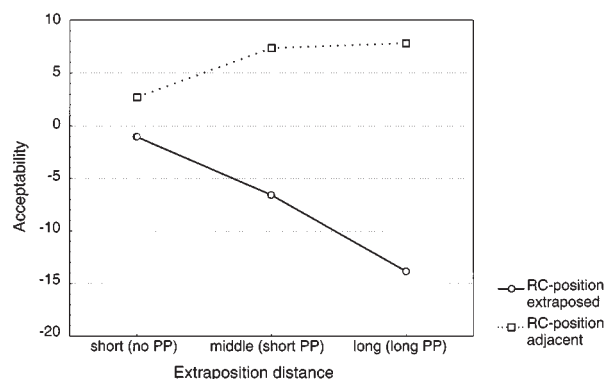
Fig. 3. Mean acceptability scores by levels of RC position (extraposed vs. adjacent) and extraposition distance.

As can be seen in Figure 3, sentences received lower ratings with increasing distances for extraposed RCs [$F_1(2,78) = 29.41$, $p < .001$; $F_2(2,112) = 51.24$, $p < .001$], while they received higher ratings with increasing distances for adjacent RCs [$F_1(2,82) = 7.32$, $p < 01$; $F_2(2,110) = 14.18$, p < 001]. The opposite picture emerged for RC lengths (see Fig. 4): sentences received higher ratings with increasing RC lengths for extraposed RCs [$F_1(2,78) = 8.0$, $p < .001$; $F_2(2,112) = 6.1$, $p < .001$], while they received lower ratings with increasing RC lengths for adjacent RCs, although only so for analysis by participants [$F_1(2,82) = 4.2$, $p < .05$; $F_2(2,110) = 1.72$, ns].

Posthoc Scheffé tests revealed that for adjacent RCs, the effect of potential distance was due to short distances (no PP) being judged worse (2.84) than both middle (short PP: 7.41; $p < 01$ by participants and by items) and long distances (long PP: 7.84, $p < .01$ by participants and by items). Also, for adjacent RCs, the effect of RC length was mainly due to short RCs being judged better (7.47) than long RCs (4.06, $p < .05$ by participants, by items). For extra-
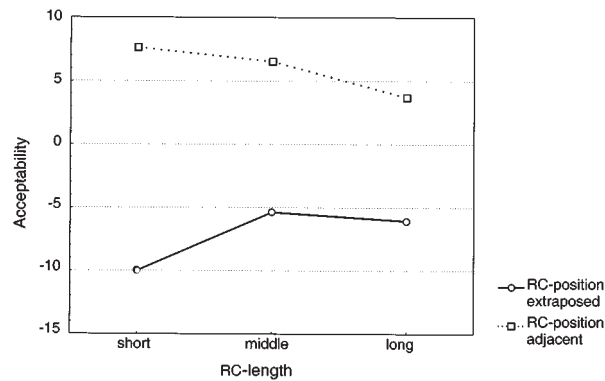
Fig. 4. Mean acceptability scores by levels of RC position (extraposed vs. adjacent) and RC length.

posed RCs, each increase in distance resulted in reliably worse acceptability judgments (no PP: −1.05; short PP: −6.57; long PP: −13.83; all ps < .01 for analyses both by participants and by items), while short RCs were reliably worse (−10.0) than both middle (−5.37, $p < .01$ by participants, and $p < .05$ by items) and long RCs (−6.07, $p < .01$ by participants and $p < .05$ by items).

Discussion

The somewhat unexpected result is the strong overall preference for adjacent RCs. Although EIC predicts a preference of adjacent RCs, at least those RCs extraposed across only one word should have been judged more acceptable than their adjacent counterparts. This has not been the case. This finding is even less compatible with SPLT's prediction of a general preference for extraposed RCs.

The interactions of RC position with both distance and RC length fit EIC's predictions almost perfectly. Sentences with adjacent RCs were judged worse with increasing RC lengths and decreasing potential distances. Conversely, sentences with extraposed RCs were judged worse if the RCs were shorter, or if distances were longer. This is the pattern of results predicted by EIC under the *difference hypothesis*. Although the pattern of interactions was also predicted by SPLT and EIC under the judge-as-is-hypothesis, they did not predict extraposed sentences to be negatively affected by decreasing RC length.

Acceptability scores are usually understood as language processing data rather than production data. However, since judgments have been made off-line, it remains open as to what extent they reflect on-line processing

only, or to which extent they also mirror stylistic after-thoughts, which potentially involve silent rereading, and possibly even the production of alternative word orders. The answer can only be given by an on-line study. Since the materials have proved to capture length effects, they seem well suited for this purpose.

EXPERIMENT II

A self-paced reading study was conducted to establish on-line processing complexity at relevant positions: the matrix clause final verb participle and the relative pronoun. Experiment II was based on the same $2 \times 3 \times 3$ design as Experiment I, with factors *RC position* and *length* and *extraposition distance* (3–4). Fifty-four participants read the 54 sentences taken from the materials of Experiment I, distributed in eighteen lists with each list containing an equal number of each of the eight experimental condition, but only one version of each sentence. The experimental items were embedded in a list of 58 fillers of various syntactic structures.

Procedure

Participants were instructed to read sentences on a computer display word-by-word by pressing a button for the next word in the sentence. The time between two consecutive button presses were taken as an indicator of the reading time for that word. Words were presented noncumulatively in a stationary window in the center of the screen to avoid line breaks, which would have been inevitable for longer versions of the sentences. After each sentence, a YES–NO question was presented on the screen, which subjects had to answer by pressing one of two buttons. Questions were answered highly accurately independent of conditions (between 92% and 97% correct). The experiment was set up in MEL 2.1, running on a Pentium PC with MS-DOS 6.

Hypotheses

EIC (Hawkins, 1994) predicts processing to be harder if constituent recognition domains (CRDs) grow long. However, EIC only predicts the overall quality of a sentence. It is not very clear which predictions for online reading time can be derived from EIC. If EIC should be conceived of as a principle of parsing, the most straightforward prediction would be similar to SPLT's, but for different reasons. Assuming that the last word in a CRD is the region where the constituent can finally be identified (i.e, constructed), most work should be expected here. For the VP, the verb-partici-

ple allows the construction of the VP. The NP, on the other hand, can only be identified at the relative pronoun.

SPLT's predictions are similar: the verb and the pronoun are the regions where integration costs of both word-order variants should differ from each other. Tables II and III provide the predictions of on-line integration costs for the verb and the relative pronoun. The verb should take longer to read (1) if it follows the RC (main effect of RC-position), (2) the longer the preceding RC (*RC-position* by *RC-length* interaction), (3) the longer the preceding PP (main effect of *distance*). The relative pronoun should be harder to read, (1) if the RC is extraposed and (2) for longer extraposition distances (RC *position* by *distance* interaction).

Results

Raw reading times were cleaned from extremes (per condition) before aggregation.[8] Reading times on the verb-participle and the relative pronoun were submitted to a $2 \times 3 \times 3$ GLM (MANOVA) for repeated measures with the factors RC position and length and extraposition distance. Since some of the verbs differed in length between the no-PP condition and the PP conditions because of a directional affix (gelegt vs. *hin*gelegt), verb-length (number of characters) was included as a covariate for verb-reading times to eliminate its influence from the data.

Table V shows mean reading times in each of the conditions. A reliable main effect of RC position could be established, such that verbs were read faster when they followed a RC [$F_1(1,46) = 18.53$, $p < .001$; $F_2(1,52) = 6.29$, $p < .05$]. As illustrated in Figure 5, there was also a main effect of distance, but only by items [$F_1(2,92) = .21$, ns; $F_2(2,104) = 4.13$, $p < .05$]. All other effects and interactions were far from being significant (all $Fs < 1.5$).

For relative pronouns, a main effect of RC position was found [$F_1(1.51) = 25.04$, $p < .001$; $F_2(1,53) = 44.46$, $p < .001$), indicating that pronouns in extraposed RCs took longer to read (459 ms) than in adjacent RCs (433 ms). All other effects and interactions were not reliable ($Fs < 1.8$).

Table V. Mean Reading Times (msec) at the Matrix Clause-Final Verb

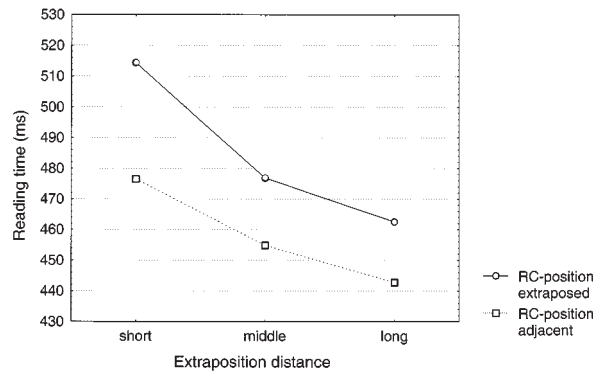|	Extraposed RC			Adjacent RC		
	Distance short	Distance middle	Distance long	Distance short	Distance middle	Distance long
RC short	526	488	460	498	448	432
RC middle	514	465	457	470	469	447
RC long	502	478	471	461	447	450

Fig. 5. Mean reading times at the clause final verb by levels of RC position
(extraposed vs. adjacent) and extraposition distance.

Discussion

On-line reading times did not support locality-based predictions. The
verb was read reliably faster when its integration with its arguments had to
cross an RC. This effect, however, did not seem to be modulated by the length
of the RC. Furthermore, distance did not elicit the predicted effect on the verb
(numerically, verbs even were read faster after longer distances, i.e., after PPs).

The RC pronoun, on the other hand, showed a mild sign of locality: it
was read longer when the RC was extraposed. Extraposition distance, how-
ever, did not modulate this effect. Contrary to the verb, the RC and its pro-
noun cannot be predicted syntactically (and not pragmatically either in the
given context). This leads us back to the question asked earlier on, namely,
whether locality-based integration cost might be restricted to nonpredictable
items. The present results support this hypothesis. On the other hand,
adding a PP to the distance did not strengthen the effect, indicating that fur-
ther surface distance does not matter. Why?

It has been hypothesized quite successfully that attaching a relative pro-
noun is, in fact, a dualistic process (Hemforth, Konieczny, & Scheepers,
2000a; Hemforth, Konieczny, Seelig, & Walter, 2000b; Konieczny and
Hemforth, 2000): It is both syntactic and anaphoric in nature. In cases of ambi-
guity, whichever process wins the race determines the attachment/binding pref-
erence. If the RC is extraposed, however, attachment is blocked and
interpretation is determined by anaphoric binding. Binding a pronoun to a dis-
course referent is not solely determined by surface distance, but also by dis-
course structure, including aspects of focus, etc. An intervening PP does not
take much of the prominence of a direct object, such that the object remains
accessible for the binding process. On the background of this model, the data

make sense. Extraposition prohibits fast attachment/binding, but adding a PP does not hurt that much.

GENERAL DISCUSSION

The on-line data stand in sharp contrast to the predictions made by locality-based models. Reading times on the clause final verb were shorter, not longer, when integration had to cross an RC. This result seems to support the anticipation hypothesis, which states that the verb can be anticipated through preceding arguments. However, anticipation should have gained from adding a PP argument, but it did not reliably so. Furthermore, adding an RC to one of its arguments does not (necessarily) constrain further the class of verbs yet to come. On the other hand, a RC might provide specific information about an argument that facilitates the anticipation of the verb on semantic grounds. Whether or not anticipation is operative here remains an open question.

The data are, however, compatible with *incremental partial commitment* approaches, such as suggested by Konieczny (1996, p. 207ff.; see also Scheepers, Hemforth, & Konieczny, 1999). When arguments of a clause final verb are being integrated into the structure, the *type* of the verb (i.e, its COMP feature) is further constrained to allow only certain arguments to come. When the verb arrives, everything is already in place and there is no need to retrieve an argument from any place far away. Constraining the verb type amounts to type- deduction, which might go beyond valency prediction, if enough time is provided. An RC preceding the verb might just provide enough time to finish type deduction as far as possible.

Off-Line vs. On-Line Data

Perhaps the most striking result of this paper is the dissociation of the off- and the on-line data. Questionnaire acceptability judgments are usually understood to reflect preferences in language perception, since participants respond to what they *read* and there is a large body of evidence that this is, in fact, the case. However, this might not be the complete picture. In the present paper, the subjects' task was to judge the acceptability of rather weak stylistic variations of otherwise well-formed sentences. No garden-path sentences or any other strong linguistic violations were included in the materials. Subjects had all the time they wanted to perform that task. The data clearly show that they have been sensitive to word-order variations, and they responded in a way predicted by locality-based accounts. More specifically, the pattern of results seems to support the *difference hypothesis*, by which subjects compare a

given sentence with a possible word-order alternative. In order to do so, they have to generate this alternative first, which eventually involves *language production.* Remember that locality-based predictions were also supported by corpus counts of RC types (Uszkoreit *et al.,* 1998). Corpus data are primarily production data, although they might have been edited several times. The fact that the present off-line data pattern with corpus data, but are dissociated from on-line results, clearly indicate that locality-based processing cost is primarily a production phenomenon (cf. Wasow, 1997).

However, this might be a bit too strong a conclusion. First, there are elements in the data that could be interpreted as indicators for locality, namely, elevated reading times on the relative pronoun in an extraposed RC. Second, there is a substantial body of evidence supporting locality-based predictions in on-line studies. Most of them, however, compare canonical with marked word-order alternatives, such as object extractions in RCs (e.g., King and Just, 1991). However, marked word order might block the anticipatory mechanisms of incremental prehead processing, in which case it remains an open question as to whether integration cost is locality-based here. Future research will show.

CONCLUSION

The results of the off- and the on-line study suggest that locality in unambiguous sentences is primarily a factor which determines word-order preferences in sentence production rather than in perception. The off-line acceptability judgment study revealed that subjects take into account word-order alternatives of the present sentence and then judge on the basis of locality-based production preferences. When participants read the same materials on-line, reading times showed few signs of locality. It will be interesting to study other materials, including systematic deviations from canonical word-order to investigate anticipatory mechanisms in incremental prehead parsing.

REFERENCES

Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 225–242). Hillsdale, NJ: Erlbaum.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language, 72,* 32–68.

Ferreira, F., & Henderson, J. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30,* 725–745.

Fodor, J. D., & Frazier, L. (1980). Is the HSPM an ATN? *Cognition, 8,* 417–459.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *The psychology of reading. Attention and performance XII* (pp. 559–586). Hove, London: Erlbaum.

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

Hawkins, J. A. (1994). *A performance theory of order and constituency* (Cambridge studies in Linguistics). Cambridge: Cambridge University Press.

Hemforth, B., Konieczny, L., & Strube, G. (1993). Incremental syntax processing and parsing strategies. *In Proceedings of the 15th Annual Conference of the Cognitive Science Society,* July, 1993 (pp. 539–545). Hillsdale, NJ: Erlbaum.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000a). Modifier attachment: relative clauses and coordinations. In B. Hemforth, & L. Konieczny (Eds.) *German sentence processing,* (pp. 159–163), Dordrecht, NL: Kluwer.

Hemforth, B., Konieczny, L., Seelig, H., & Walter, M. (2000b). Case matching and relative clause attachment. *Journal of Psycholinguistic Research, 29/1*, 81–88.

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. *Language and Cognitive Processes, 14,* 631–662.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2000). Predictive eye-movements in incremental processing of head-final structures. *Thirteenth Annual CUNY Conference on Human Sentence Processing,* La Jolla, California.

Kimball, J. (1973). Seven principles in surface structure parsing in natural language. *Cognition, 2,* 15–47.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language, 30,* 580–602.

Konieczny, L. (1996). Human sentence processing: A semantics-oriented parsing approach (IIG-Bericht Nr. 3/96). Freiburg: Doctoral Thesis at the University of Freiburg, Institut für Informatik und Gesellschaft.

Konieczny, L., & Hemforth, B. (1994). Incremental parsing with lexicalized grammars. In G. Strube (Ed.), *Current research in cognitive science at the Center for Cognitive Science* (pp. 33–54). IIG-Bericht Nr. 1/94. Freiburg: University of Freiburg, Institut für Informatik und Gesellschaft.

Konieczny, L., & Hemforth, B. (2000). Modifier attachment in German: relative clauses and prepositional phrases. In A. Kennedy, R. Radack, D. Heller, & J. Pynte (Eds.). *Reading as a perceptual process.* (pp. 517–527.) New York: Elsevier.

Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G, (1997). The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes, 12,* 307–348.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Scheepers, C., Hemforth, B., & Konieczny, L. (1999). Incremental processing of German verb–final constructions: predicting the verb's minimum (!) valency. *The 2nd International Conference on Cognitive Science and 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference* (ICCS/JCSS99)

Stevenson, S. (1993). A constrained active attachment model for resolving syntactic ambiguities in natural language parsing. Doctoral dissertation, University of Maryland.

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik. Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft, 7,* 129–133.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexical grammar. *Cognition, 75,* 105–143.

Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research, 26,* 347–361.