# PERSPECTIVES ON SENTENCE PROCESSING

Edited by
**Charles Clifton, Jr.**
**Lyn Frazier**
**Keith Rayner**
*University of Massachusetts at Amherst*

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1994 Hillsdale, New Jersey Hove, UK

# 8 A Corpus-Based Analysis of Psycholinguistic Constraints on Prepositional-Phrase Attachment

Edward Gibson
*Massachusetts Institute of Technology*

Neal J. Pearlmutter
*University of Illinois at Urbana-Champaign*

Most empirical psycholinguistic data are currently obtained by performing experiments on subjects (e.g., measuring reading times or eye movements associated with the processing of sentences). The advent of large on-line corpora, along with computational tools for processing these corpora, provides us with a new psycholinguistic testing ground. In particular, it is now possible for researchers studying psycholinguistic complexity to measure the frequencies of target-linguistic constructs in corpora relative to the frequencies of control-linguistic constructs. Our claim is that, other factors being equal, structures that are harder to understand (as measured, e.g., by reading times) should be less frequent than structures that are easier to understand, so that the relative frequency measurements of two or more constructions are inversely related to the relative comprehension-complexity ranking of the constructions.

Before we discuss reasons for believing this claim, note that the claim we are making here is *not* the following, more general claim, which is too strong: That the frequency of a structure varies inversely with its complexity alone. This stronger hypothesis is false because many factors other than complexity are involved in determining the frequency of a structure. One structure may be frequent, yet still relatively complex, whereas another might be easy to comprehend, yet rare for reasons other than complexity. Therefore, when attempting to make conclusions about relative syntactic complexities on the basis of frequency information, it is necessary to ensure that there are no differences between the target and control structures other than the complexity difference that is being studied. For example, if two constructions function in different ways, and thus relay different kinds of meaning, then no conclusions can be drawn about their relative syntactic complexities based on frequency comparisons because the difference in frequencies may be primarily due to the difference in function of the two constructions. Thus, it would not be appropriate to compare the number of declarative structures of a certain form to the number of interrogative structures of a similar form because, although declaratives and interrogatives may be closely linked syntactically, their functions are very different.

However, if factors such as function can be controlled for, and a sufficient quantity of target and control items can be observed, then our claim applies: The frequency of the target construction relative to the control construction varies inversely with the relative complexity of the two constructions. One instance where such a comparison seems reasonable is in the area of ambiguous attachment of a constituent (e.g., a prepositional phrase [PP]). Suppose that we can obtain a large number of instances of a structure *S* containing *n* prospective attachment sites, where each *S* token is followed by the same kind of constituent *C*, so that *C* attaches to one of the *n* prospective attachment sites in *S*. If the tokens of *S* and *C* are the same in all relevant aspects, then attachments to sites within *S* that *C* attaches to more frequently are less complex than attachments that occur less frequently.

There are two main reasons to believe the hypothesis relating the relative comprehension complexities of two constructions to the inverse of their relative frequencies. First, because most on-line corpora are collections of sentences that were produced naturally, it is clear that the relative frequencies of two or more constructions are inversely related to the relative complexities of these constructions with respect to the human sentence-production mechanism. This complexity measurement may also directly reflect the human sentence-processing mechanism (HSPM), if the comprehension and production systems are closely linked to one another and rely on similar definitions of complexity. Although the specific nature of the relationship between the two systems is yet unclear, the existence of some connection between the comprehension and production systems is uncontroversial. At the very least, the two systems must access some of the same lexical representations, and, presumably, the same grammatical knowledge must be applied in each system, regardless of how it is encoded. Currently, there is no evidence distinguishing the complexity measures of the two systems. Until such evidence appears, the null hypothesis is that the two systems share the same complexity measures, so that measurements of production frequencies should reflect comprehension processing difficulty.

An additional reason to expect that corpus frequencies should reflect human sentence-processing complexity is that most of the text currently available in on-line corpora has been edited at some stage to make it more understandable (e.g., newspaper stories, etc.). Thus, naturally generated constructions that are too complex for the human sentence processor will often be edited to make them easier. For example, some types of ambiguity, which often give rise to processing difficulty, might not be noticed in the generation process, but might be filtered out by an editing process.[1]



Therefore, we expect to find the same patterns of complexity in corpus analyses as in comprehension data. If such patterns are observed, we can infer the existence of some combination of an effect of editing and a link between production and processing systems, at least in terms of shared complexity measures. In this case, corpus-based analyses can at least supplement data from comprehension experiments, and, given enough support across a range of comparisons, corpus data could be taken as independent evidence for or against processing hypotheses.

However, it is possible that the production and comprehension systems do not share the same complexity measures, and/or that the effects of editing are relatively weak, resulting in two other possible general patterns of data from corpus-based analyses:

1. Frequency data from corpus analyses might be unrelated to complexity data from comprehension experiments, indicating that the production and processing systems are relatively isolated from each other, and that editing processes do not have a strong effect.
2. Frequency data might reflect comprehension data only in certain aspects, making direct conclusions about the processing system on the basis of corpus data along somewhat suspect. In this case, a characterization of how the two sources of data differ—in the form of a range of comparisons between corpus and comprehension data (and possibly other independent production data)—could help to determine the relationship between the processing and production systems.

As an initial test of these claims, this chapter compares the results of a corpus analysis with the experimental results reported by Gibson, Pearlmutter, Canseco-Gonzalez, and Hickok (1993). Gibson et al. performed off- and on-line reading experiments on ambiguous noun phrases (NPs) of the following form, where the relative clause (RC) attached to one of $NP_1$, $NP_2$, or $NP_3$ (Prep = preposition, N = noun):

(1) $[_{NP_1} \ldots N_1 [_{PP_1} Prep_1 [_{NP_2} \ldots N_2 [_{PP_2} Prep_2 [_{NP_3} \ldots N_3 ]]]]]$ RC

[1]Note that this observation means that edited corpora may not be as directly relevant to the human sentence-production mechanism as unedited corpora. Thus, a better source of measurements relevant to the sentence-production mechanism would be naturally occurring speech, or naturally occurring dialogues where editing is somehow minimized (e.g., electronic mail interactions), as in studies of naturally occurring speech errors, where editing has clearly not occurred (see, e.g., Dell, 1986; Garrett, 1975).

The constructions analyzed had the additional constraint that $PP_2$ attached to $NP_2$, so that the RC could syntactically attach to any one of $NP_1$, $NP_2$, or $NP_3$.[2] In addition, the head nouns and RC were constructed so that the RC agreed in number with only one of the three possible NP attachment sites, as in the following NPs:

(2) a. the computer near the models of the buildings that was destroyed in the fire
b. the computers near the model of the buildings that was destroyed in the fire
c. the computers near the models of the building that was destroyed in the fire

The intuitions are fairly clear: (2c), in which the RC attaches to $NP_3$, is easy to process and seems perfectly acceptable. (2a), in which the RC attaches to $NP_1$, is somewhat harder to process, but is still relatively easy to understand. However, (2b), in which the RC attaches to the intermediate site $NP_2$, is very difficult to process and seems ungrammatical, giving rise to a noticeable garden-path effect.

Accordingly, the results from the off-line grammaticality judgment experiment reported by Gibson et al. indicated that NPs like (2a) and (2c) are usually perceived as grammatical (66% and 69% of the time, respectively), whereas NPs like (2b) are perceived as grammatical much less often (only 29% of the time). An on-line reading-time experiment also found that attachment to $NP_2$ was the hardest to comprehend. Furthermore, in the on-line task, attachments to $NP_3$ were significantly easier to process than attachments to $NP_1$, resulting in a difficulty ordering of the three sites as follows (easiest to hardest): $NP_3$, $NP_1$, $NP_2$.

To explain these effects, Gibson et al. proposed that the human parser contains the following independent preference factors:

1. Recency preference (cf. right association; Kimball, 1973, and late closure; Frazier, 1978, 1987; Frazier & Fodor, 1978), which states that the human parser prefers attachments to more recent words in the input stream over attachments to less recent words; and
2. Predicate proximity (cf. Frazier, 1990; Gibson, 1991; Gilboy et al., 1993), which states that the human parser prefers attachments to be as close as possible to the head of a predicate.

[2]If $PP_2$ attached to $NP_1$, then attachment of the RC to $NP_2$ would be syntactically blocked by a prohibition against crossed branches.



Given a RC following three potential NP attachment sites, as in (1), recency preference results in a preference for attachment to the most recent, or lowest, NP—$NP_3$. On the other hand, predicate proximity results in a preference for attachment to $NP_1$, the highest site, because $NP_1$ is structurally closer to the head of a predicate to follow (as its potential subject) than either $NP_2$ or $NP_3$.[3] Recency wins out in this case, partially because the attachment to $NP_1$ is so distant, with an intermediate attachment site $NP_2$ also available. The ranking of the intermediate (middle) site, $NP_2$, as least preferred derives from the fact that this site has neither recency preference nor predicate proximity in its favor. (See Gibson et al. (1993) for details.[4]) Other possible accounts of the observed complexity ordering include the tuning theory of Cuetos and Mitchell (1988), Mitchell (1993) and the references there, and the automaton theory of Langendoen and Langsam (1987), among other possibilities.[5]

The current chapter explores whether this on-line psycholinguistic result is replicated in corpus frequencies. The prediction is that, when a large number of instances of constructions of the form in (1) are considered, the frequency of low attachment (to $NP_3$) should be highest, followed by the frequency of high attachment (to $NP_1$), with middle attachments (to $NP_2$) least frequent of all. To evaluate this hypothesis, we examined most (as many as we could identify) of the NPs from the one million-word Brown corpus (Kučera & Francis, 1967) having a form similar to (1). In this analysis, any of the three attachments of the final modifying phrase (the RC in (1)) can be viewed as the target, relative to the two other attachment cases as controls. As predicted, we found that attachments to intermediate sites are indeed the most rare, and attachments to the most recent site ($NP_3$) are most common, with attachments to $NP_1$ falling in between. As a result, we conclude that corpus analysis can provide a useful additional method to test psycholinguistic hypotheses, either as a first test or as a validating measurement.

[3]Note that, at the point of attaching the RC, the predicate has not yet been processed. Thus, we are implicitly assuming a predictive-parsing algorithm—one that expects $NP_1$ to be subject of a matrix verb to follow. See Kimball (1973, 1975), Frazier and Fodor (1978), and Gibson (1991) for evidence that the human parser uses a predictive-parsing algorithm. See Gibson (1991) for an algorithm that is appropriate for use with predicate proximity.

[4]It should be noted that the presented theory also accounts for numerous other English parsing preferences. Furthermore, it is parameterized to account for the Spanish attachment results reported in Cuetos and Mitchell (1988), Mitchell, Cuetos, and Zagar (1990), and Mitchell and Cuetos (1991).

[5]An explanation for these findings on the basis of the serial-position effect from the memory and free-recall literature (e.g., Murdock, 1962; Tulving, 1968; and the references in each) might also seem possible, with recency effects and primacy effects corresponding to preference for low and high attachments, respectively. However, such an approach is empirically inadequate to explain a range of contrasts in attachment preferences beyond those described here, as discussed in Gibson et al. (1993).

## ANALYSIS OF PP MODIFIERS OF NPS IN THE BROWN CORPUS

To test the psycholinguistic attachment-complexity hypotheses put forward by Gibson et al. (1993), we examined most of the NPs from the Brown corpus having the form in (3), which is the same as (1), but including a third PP modifier in the place of the RC:[6]

(3) $[_{NP_1} \ldots N_1 [_{PP_1} Prep_1 [_{NP_2} \ldots N_2 [_{PP_2} Prep_2 [_{NP_3} \ldots N_3 ]]]]]$
$[_{PP_3} Prep_3 NP_4]$

Using text-extraction tools (from the CLARIT project; Evans, Handerson, Lefferts, & Monarch, 1991) that are based on the work of de Marcken (1990) and Grefenstette (in press), every complex NP containing three or more prepositions was extracted from the Brown corpus, resulting in over 4,000 such phrases. However, many of these extracted phrases did not match the desired pattern, for a number of reasons:

1. Failure to conform to (3) because of poor category taggings.
2. One of the PPs attached to a site other than $NP_1$, $NP_2$, or $NP_3$. For example, in (4), "his supervision with a reduction in the number of forest fires in the state" was identified as a possible candidate NP. This particular sequence of words was ruled out because the PP "with a reduction" attaches to a preceding verb "credits," rather than to the NP "his supervision"[7]:

   (4) A citation from Conservation Commissioner Salvatore A. Bontempo credits his supervision with a reduction in the number of forest fires in the state.
3. An intermediate modifying phrase (other than a PP, as in a conjunction or a verb phrase) was attached to $NP_1$ or $NP_2$, making attachment to $NP_2$ and/or $NP_3$ impossible.

Thus, all of these kinds of examples were eliminated from consideration. Furthermore, we considered only the first three PPs that satisfied our requirements in a given example. Finally, we ruled out examples in which one of the prospective NP attachment sites was not a possible site for attachment. In particular, we eliminated examples in which one of the identified PPs was part of a

[6]Patterns of the form in Sentence (3) are much more common than are patterns of the form in Sentence (1), so that better statistics can be computed on the former.

[7]Although the pattern initiated by "his supervision" is not an appropriate item for our analysis, the NP that immediately follows the preposition "with" in Sentence (4)—"a reduction in the number of forest fires in the state"—satisfies the target pattern, and thus was included.





complex proper noun, such as "the United States of America," "Massachusetts Institute of Technology," or "Doctor of Philosophy." Such examples were eliminated because they contain only one possible attachment site for following modifiers, rather than two possible sites as in other NP–Prep–NP cases. Similarly, we ruled out examples that contained pronouns as one of the NP sites because pronouns in general do not allow modifier attachment.

Examples were also ruled out in which one of the identified NPs in the target template was a quantifier, such as "one of," "some of," "all of," or "many of." These examples were eliminated because attachment of a following modifier to the quantifier is extremely marked, and perhaps ungrammatical (except in appositive constructions, which are rare). Of the 69 examples we observed that contained such quantifiers, few contained a following modifier that could attach to the quantifier rather than the quantified NP.

The remaining 543 tokens were independently hand parsed by both authors, each resulting in a parse in which each of $PP_1$, $PP_2$, and $PP_3$ is attached to one of $NP_1$, $NP_2$, or $NP_3$. Because of the way the NPs were selected, $PP_1$ unambiguously attaches to $NP_1$ in all of the examples. The first ambiguity occurs at $PP_2$, which can attach to either $NP_1$ or $NP_2$. Of the 543 tokens, 202 are clear examples of $NP_1$ attachment, whereas 319 clearly attach to $NP_2$. Choosing a preferred attachment site in the remaining 22 cases is difficult because the meaning of the phrase is very similar with either attachment (cf. similar observations by Hindle & Rooth, 1993; Hobbs & Bear, 1990). Many of these ambiguous tokens involved the attachment of a locative PP; some of these are shown in (5), with the phrase to be attached in brackets:

(5) a. the topic for a round-table discussion [at the Bayerische Rundfunk in Munich]
b. The inventories of unsold houses [in some areas of the country]

For these examples, a (slightly) preferred attachment of $NP_1$ was agreed on in 12 cases (as in (5a)), with the other 10 favoring $NP_2$ in our opinion (as in (5b)). Thus, there were 214 tokens of $NP_1$ attachment of $PP_2$, along with 329 tokens of $NP_2$ attachment (see Table 8.1). Therefore, we find a general preference to attach $PP_2$ to the most recent attachment site, although there are still many examples that violate this preference. This result replicates similar findings reported by

TABLE 8.1
Preferred Attachment Sites of $PP_2$

| *Attachment Site* | *Number of Tokens (%)* |
|---|---|
| $NP_1$ | 214 (39) |
| $NP_2$ | 329 (61) |

TABLE 8.2
Preferred Attachment Sites of $PP_3$

| *Attachment Site* | *Number of Tokens (%)* |
|---|---|
| $NP_1$ | 62 (19) |
| $NP_2$ | 63 (19) |
| $NP_3$ | 204 (62) |

Whittemore, Ferrara, and Brunner (1990), Hobbs and Bear (1990) and Hindle and Rooth (1993).

Examples in which $PP_2$ attached to $NP_1$ were then removed from consideration because such examples do not allow attachment of $PP_3$ to $NP_2$ (crossed branches would result). Of the 329 remaining examples, 62 prefer $PP_3$ attachment to $NP_1$, and 63 prefer $PP_3$ attachment to $NP_2$, with the remaining 204 preferring $PP_3$ to attach to $NP_3$. These results are presented in Table 8.2.

The results for attachment of $PP_3$ (Table 8.2) clearly indicate that the attachment pattern is not random [$\chi^2(2) = 121.72, P < .001$], with attachments to $NP_3$ occurring far more often than would be expected [vs. $NP_1$: $\chi^2(1) = 74.74, p < .001$; vs. $NP_2$: $\chi^2(1) = 73.41, p < .001$], and the frequency of attachments to $NP_1$ and $NP_2$ obviously not differing.[8,9]

The preference for attachments to $NP_3$ is as expected; this attachment is preferred in Gibson et al. (1993) as well. The lack of a difference between the number of attachments to $NP_1$ and $NP_2$ does not follow Gibson et al., but this is also not surprising because the set of 329 $PP_3$ attachment tokens includes: (a) tokens in which the three prospective attachment sites all have the same adjunct–argument status (i.e., either all adjunct sites or all argument sites), and (b) tokens in which one of the prospective sites has a different adjunct–argument status than the others (i.e., one or two of the sites are adjunct sites, and one or two sites are argument sites). It is well known that argument attachments are generally preferred over adjunct attachments, often even in cases in which recency favors the nonargument attachment. For example, Frazier's (1978, 1987) minimal-attachment principle favors argument attachments over adjunct attachments because argument attachments involve fewer phrase-structure nodes. Similarly, Pritchett's (1988, 1992) and Gibson's (1991) proposals involving the local satisfaction of the theta criterion (Chomsky, 1981) generally cause preferences for argument attachment over adjunct attachments (see also Hobbs & Bear, 1990; Whittemore et al., 1990, for corpus-based evaluations of parsing heuristics based partially on preferences for argument attachments). Thus, we would not expect the cases involving a heterogeneous set of adjunct–argument attachment sites to pattern in the same way as those involving homogeneous sites because, in the heterogeneous cases, attachment to the argument sites will tend to be preferred for independent reasons. Hence, the examples that we are interested in are those with homogenous attachment sites: either all prospective adjunct sites or all prospective argument sites.

[8]In order to use a chi-square test, we need to assume that each observation is independent, which, as pointed out by Don Mitchell (personal communication August 12, 1993), need not necessarily be the case because it is possible that some of the tokens come from the same passage in the corpus. However, independence seems to be a reasonable initial assumption.

[9]All chi-square tests with exactly one degree of freedom in this chapter have Yates' correction for continuity applied (Hays, 1988).



As a result, we determined by hand for all 329 complex NPs whether the prospective attachment of $PP_3$ to each of $NP_1$, $NP_2$, and $NP_3$ was an argument attachment, satisfying a thematic role assigned by one of the NPs, or whether the attachment was an adjunct attachment (cf. Hindle & Rooth, 1993, who attempted to obtain related lexical information automatically). Standard syntactic tests for argumenthood (see, e.g., Pollard & Sag, 1987) were used to make this judgment. Most decisions were uncontroversial, although there were a few difficult cases decided by discussion between the authors. Representative examples of argument attachments to each of $NP_1$, $NP_2$, and $NP_3$ are given in (6a), (6b), and (6c), respectively. Representative examples of adjunct attachments to $NP_1$, $NP_2$, and $NP_3$ are given in (7a), (7b), and (7c), respectively. The attaching phrase is in brackets in each example.

(6) Argument attachment:
a. $NP_1$: the relation of the figure of the dancer [to light and color]
b. $NP_2$: the host of novel applications of electronics [to medical problems]
c. $NP_3$: the lack of scientific unanimity on the effects [of radiation]

(7) Adjunct attachment:
a. $NP_1$: periodic surveillance of the pricing practices of the concessionaires [for the purpose of keeping the prices down]
b. $NP_2$: headmaster of a private school for boys [in Louisiana]
c. $NP_3$: an area about the size of Rockefeller Center [in New York]

In (6a), the noun "relation" subcategorizes for a PP headed by "to". Similarly, "application" subcategorizes for a PP headed by "to" in (6b), and "effect" subcategorizes for a PP initiated by "of" in (6c). In contrast, there is no subcategorization relationship involved in the attachment of $PP_3$ in the adjuncts attachment examples: In (7a), "surveillance" does not subcategorize for a PP headed by "for"; in (7b), "school" does not subcategorize for a PP headed by "in"; and in (7c), "Rockefeller Center" does not subcategorize for a PP headed by "in". These are all adjunct attachments.

Of the 329 cases, 155 are examples in which all of the prospective attachments are adjunct attachments, whereas 174 involve at least one argument site.

TABLE 8.3
Adjunct Attachments of $PP_3$

| Attachment Site | Number of Attachments (%) |
|---|---|
| $NP_1$ | 39 (25) |
| $NP_2$ | 23 (15) |
| $NP_3$ | 93 (60) |
| Totals | 155 (100) |

Of these 174 cases, almost all involve exactly one prospective argument-attachment site and two prospective adjunct-attachment sites. Because of the extremely low number of instances in which all three potential sites are argument sites, we restrict our analysis to the homogeneous adjunct-attachment cases.[10] The breakdown of adjunct-attachment location of $PP_3$ is presented in Table 8.3.

A chi-square analysis including all three attachment types reveals a significant effect [$\chi^2(2) = 52.08, p < .001$]. As in the analysis including arguments, the number of adjunct $NP_3$ attachments is far higher than would be expected on a random distribution of attachments [vs. $NP_1$: $\chi^2(1) = 21.28, p < .001$; vs. $NP_2$. $\chi^2(1) = 41.04, p < .001$]. Furthermore, the higher frequency of $NP_1$ versus $NP_2$ adjunct attachments is now apparent, although the effect is not quite significant [$\chi^2(1) = 3.63, p < .06$].[11]

As discussed earlier, however, it is difficult to unambiguously determine the attachment site for some modifiers because attachment to more than one site may result in identical or nearly identical interpretations. The counts in Tables 8.1, 8.2, and 8.3 reflect our intuitions about preferred attachments. In a number of cases, an alternative attachment still results in the same meaning. The existence of the possibility of alternate parses complicates the argument that the attachment site ordering presented in Table 8.3 reflects a complexity ordering. In particular, it could be that many sites that we have marked as high or low attachments also allow middle-site attachment to yield the same interpretation. Alternatively, it could be that many sites that we have tagged as middle attachments also allow high or low attachment. Thus, we examined all the adjunct $PP_3$ attachments and

[10] Another reason to restrict our analysis to the adjunct-attachment cases is that the experiments reported by Gibson et al. (1993) were also restricted to adjunct-attachment cases. Thus, although the Gibson et al. theory predicts that the pattern in homogeneous argument-attachment cases should be the same as in homogeneous adjunct cases, we do not yet have evidence that this is the case.

[11] In the 174 cases involving an argument attachment site, frequencies of attachment were 111 to $NP_3$, 40 to $NP_2$, and 23 to $NP_1$ (all three differ, *ps* <.05). This pattern can be partially explained by noting that $PP_1$ and $PP_2$ often attach as arguments, so that $PP_3$ is often a second argument for $NP_1$ or $NP_2$, but only a first argument for $NP_3$. Assuming that more NPs take one argument than take two or more, second arguments will be rarer than first arguments, and thus $NP_3$ attachments will be more common than higher attachments.



TABLE 8.4
Unambiguous $PP_3$ Attachments*

| Attachment Site | Number of Tokens (%) |
|---|---|
| $NP_1$ | 27 (26) |
| $NP_2$ | 10 (10) |
| $NP_3$ | 68 (65) |
| Totals | 105 (100) |

*Note.* Adjuncts only.

determined which of these contained $PP_3$-attachment ambiguity. The unambiguous examples were then tabulated. The results are shown in Table 8.4.

These results provide further evidence for the complexity ordering ($NP_3$, $NP_1$, $NP_2$), from least to most complex. As with the preceding analysis, a chi-square test, including all three attachment categories, reveals a significant effect [$\chi^2(2) = 50.80, p < .001$], and all three categories differ significantly, with $NP_3$ attachments most frequent [vs. $NP_1$: $\chi^2(1) = 16.84, p < .001$; vs. $NP_2$. $\chi^2(1) = 41.65, p < .001$] and $NP_1$ attachments more frequent than $NP_2$ attachments [$\chi^2(1) = 6.92, p < .01$]. Therefore, the corpus analysis corroborates the complexity ordering given in Gibson et al. (1993).

Of the remaining 10 middle ($NP_2$) attachments, most are still relatively easy to understand, despite the comprehension result demonstrating the difficulty of interpreting middle attachments. In two of these cases, the lack of difficulty can be explained by the presence of what seems to be a path constituent, which consists of two separate PPs: [$_{PP}$ from . . . ] followed by [$_{PP}$ to . . . ], as in (8):[12]

(8) the tendency of some psychologists [from Heraclitus] [to Pirandello]

Because the PPs "from Heraclitus" and "to Pirandello" form a unit, they attach to the same location, which happens to be a middle ($NP_2$) attachment site in this example.

In some of the other leftover middle attachments, $NP_3$ is a proper NP, whereas $NP_1$ and $NP_2$ are headed by common nouns. Although proper nouns do allow some PP modifications (usually nonrestrictive attachments, as in "John Doe of Cambridge"), proper nouns do not readily allow the attachment of many PPs (those that are not easily interpreted nonrestrictively). Hence, if $NP_3$ is a proper NP to which $PP_3$ cannot attach, then there are effectively only two possible attachment sites for $PP_3$ ($NP_1$ and $NP_2$), as in (9), in which the PP "as the only effective method" is not permitted as a modifier of the proper NP "Cuba":

(9) the imposition of a naval blockade of Cuba [as the only effective method]

[12] Thanks to an anonymous reviewer for providing this analysis.

In other middle attachments, the attachment of $PP_3$ to one of $NP_1$ or $NP_3$ results in a meaning that is closely related to the meaning that results from attachment to $NP_2$, although this meaning is potentially distinct. It is possible that people do not notice this difference in most contexts, so that these NPs are interpreted based on a low or high attachment instead. For example, in (7b), the attachment of the PP "in Louisiana" to $NP_2$ or $NP_3$ might yield the same meaning, given an appropriate context (e.g., (7b): headmaster of a private school for boys [in Louisiana]).

Thus, the remaining middle-attachment examples that are present in the corpus are not overly difficult to understand. Of course, the lack of observed difficult middle-attachment examples in the corpus does not mean that all middle attachments are easy to understand. The middle-attachment stimuli in the experiments performed by Gibson et al. (1993) are examples of middle attachments that are difficult to understand. Furthermore, we can construct examples from the corpus of $NP_1$- and $NP_3$-attachment items in which attachment to $NP_2$ is forced semantically/pragmatically. These examples should be difficult to understand. For example, the NPs in (10b)–(10d) were constructed from $NP_1$-attachment examples by rewriting $PP_3$ so that the only plausible attachment was to $NP_2$. As expected, the constructed examples with forced middle attachment ((10b) and (10d)) are much more difficult to understand than the corresponding high-attachment versions:

(10) a. a program of prepayment of health costs [with absolute freedom of choice guaranteed]
b. # a program of prepayment of health costs [with a credit card]
c. a pile of wire cages for mice [from his time as a geneticist]
d. # a pile of wire cages for mice [with small doors]

## GENERAL DISCUSSION

The corpus-analysis results described here corroborate the results of Gibson et al. (1993), replicating in corpus frequencies the pattern that Gibson et al. found in reading times and in grammaticality judgments: In ambiguous modifier attachments to NP sites, attachments to the lowest site are preferred over attachments to high and middle sites, and attachments to high sites are preferred over attachments to middle sites. However, as discussed previously, we might have found that the frequencies of different attachment types in our corpus were unrelated to the complexity ordering in Gibson et al. (1993), or that the frequency results replicated only part of the experimental pattern, so that, for example, low attachments were most frequent, but middle attachments were more frequent than high attachments. The fact that the current results agree with independent experimental findings casts doubt on the hypothesis that corpora and comprehension data are completely independent.



We do not yet have enough data to answer more detailed questions about the relationship between corpus frequencies and comprehension data, but we can speculate: Although editing processes probably play a role in determining corpus frequencies, we do not think that they can provide a full account of the apparently similar corpus and comprehension data patterns. Thus, we would expect to find the same similarity between the Gibson et al. results and frequency data derived from more spontaneously produced corpora. This prediction remains to be investigated.

Furthermore, if editing processes are not the sole cause of similarities between corpus-frequency data and comprehension data, we are faced with the question of what other factors are responsible. One possibility is that raised by the linguistic tuning theory (Cuetos & Mitchell, 1988; Mitchell, 1993), which, in its most general form, suggests that processing decisions in comprehension reflect construction frequencies in the environment, so that the more frequent interpretation of an ambiguity will be the preferred one when the ambiguity is encountered during the course of comprehension. Thus, linguistic tuning suggests that, at least in the case of ambiguity, comprehension difficulty and corpus frequency will be closely matched because the latter determines the former—no additional factor is present. The comprehension system simply keeps track of (at least certain types of) frequency information. Although many details of linguistic tuning remain to be specified (see Gibson et al., 1993, for some discussion), it is certainly compatible with the results presented here.

However, the theory involving recency preference and predicate proximity, described earlier and discussed in detail in Gibson et al. (1993), suggests an obvious alternative relationship between corpus and comprehension data, in which the two preference factors, already implicated in comprehension, apply in production as well. The details of how these preferences would operate in a sentence-production system remain to be worked out. One possibility would be to allow the production system to generate, in parallel, possible partial candidate syntactic structures for a given message to be expressed (perhaps along the lines of Garrett, 1975), with the two preference factors applying to select a single structure for use or to rank the alternatives. (See Gibson, 1991, for this conception of such preference factors in a comprehension system.) Alternatively, it is possible that the effects of the preference factors might follow from other properties of the production system: For example, recency might be a natural property of a production system that relies on the notion of spreading activation. Regardless of exactly how the preference factors end up being incorporated into the production system, on this approach the underlying explanation for the similarity between corpus-frequency and comprehension-difficulty data is that the systems responsible for generating the two patterns depend on the same underlying measure(s).

The data available so far are obviously much too limited to decide these issues. At the very least, the fact that the corpus-analysis results closely corroborate experimental results suggests that corpora can provide a useful supplement to experimental evidence in evaluating psycholinguistic processing theories. With further examinations of corpora and comparisons to related experimental data, it should be possible to develop a better understanding of how corpus statistics such as frequency are related to measurable properties of the comprehension process, and how such information can be used to evaluate psycholinguistic theories.

In addition to the psycholinguistic implications of these results, there are at least two relevant computational ramifications. First, the results reported here provide additional support for the conclusion that semantic-pragmatic information is necessary to achieve practically useful disambiguation (e.g., Crain & Steedman, 1985; Hirst, 1987; Hobbs & Bear, 1990; Schubert, 1984, 1986; Taraban & McClelland, 1988; Whittemore et al., 1990; Wilks, Huang, & Fass, 1985; among many others). In particular, although more than half of the $PP_2$ attachments in our corpus are to the most recent NP, 39% attach to the less recent NP. Even if the parser has access to argument- and/or lexical-preference information (cf. Ford, Bresnan, & Kaplan, 1982), it will get only 88 more of the 214 attachments to $NP_1$ correct (see Table 8.1), for a success rate of 79% on a single attachment decision.[13] Such a success rate is not practically useful, especially considering that a parse of a complete sentence may involve many such decisions. If only three decisions of this kind are necessary, the probability of a successful parse is only 49%. Thus, syntactic and lexical preferences are not sufficient for PP-attachment disambiguation in a practical natural-language processing (NLP) system.

[13]This success rate, using lexical preferences and recency, is somewhat low compared with several others in the recent computational literature: Whittemore et al. (1990) reported a success rate of 92% for a similar algorithm, and Hindle & Rooth (1993) achieved 85% success using automatically-acquired lexical preferences alone (no recency), when their algorithm's confidence in an attachment is 95% or higher. Furthermore, Hobbs & Bear (1990) reported a rate of 86% for a lexical preference/recency algorithm on cases not involving PPs headed by the preposition *of.* If we also exclude cases involving *of,* the success rate of the lexical preference/recency algorithm is only 61% in the current study.

A detailed consideration of the differences among these rates is beyond the scope of this chapter, but a major contributor to the lower success rate reported here is probably the difference in homogeneity of the attachment sites (in addition to differences in corpora). In particular, when potential attachment sites are homogeneous (of the same category), as in the current study (all NPs), the power of lexical preferences is likely to be degraded compared with attachment ambiguities involving heterogeneous sites. For example, certain prepositions strongly prefer to attach to nouns (e.g., *of*), whereas others (e.g., *during*) prefer to attach to events, which are canonically realized syntactically as verbs. When potential attachment sites are of heterogeneous categories, these preferences will often eliminate some sites from contention, whereas when all potential attachment sites are of the same category, strong category-based preferences will be irrelevant.



Second, when attempting to disambiguate an adjunct attachment in constructions of the form discussed here, an NLP system should adopt a search ordering of $NP_3$, $NP_1$, $NP_2$. More generally, when three or more NP sites are present, the most recent should be attempted first, followed by the least recent (followed by an as yet undetermined order of the other NP sites).

Previous NLP work in PP-attachment disambiguation has usually assumed a general recency preference, but no interacting high-attachment preference to lead the parser to rank $NP_1$ above $NP_2$. For example, Woods' (1973) selective modifier placement (SMP) heuristics ranked sites primarily by semantics, and then by recency if one site did not emerge as best. Schubert (1984, 1986) and Wilks et al. (1985) discussed similar semantics-based disambiguation methods.[14] These methods will fail on examples with three NP sites where (a) the most recent site is semantically blocked from attachment, and (b) there is no semantic difference between attaching to either of the two other sites. Examples of such constructions involving RC attachment are given in (11a) and (11b) (attaching clause in brackets):

(11) a. the sign above the memo to the committee [that was written in black ink]
b. the article on the movie about the murder [that was incorrect in many details]

Although people tend to prefer the attachment of the RC to the high-attachment site ("the sign" in (11a), "the article" in (11b)), the semantics/recency schemes will result in an attachment to the middle site. Naturally occurring examples of such high-attachment preferences are also present in the corpus discussed here. Two are given in (12a) and (12b):

(12) a. periodic surveillance of the pricing practices of the concessionaires [for the purpose of keeping the prices down]
b. a prohibition upon the withholding of patent rights [among A.L.A.M. members]

In Sentence (12a), the PP "for the purposes of keeping the prices down" cannot attach to $NP_3$ "the concessionaires" for semantic reasons. Attachment to either $NP_1$ "periodic surveillance" or $NP_2$ "pricing practices" is acceptable semantically, although the two readings have different meanings. Thus, the NLP systems of Woods (1973), Schubert (1984), and Wilks et al. (1985) would

[14]Some of these approaches rank attachments to a matrix verb phrase as better than a less recent NP attachment when semantics does not disambiguate, but none ranks a less recent NP attachment as better than a more recent one when semantics does not disambiguate.

probably choose $NP_2$. However, high attachment ($NP_1$) is preferred by people and turns out to be correct in context. Similarly, in Sentence (12b), attachment of the PP "among A.L.A.M. members" to $NP_3$ "patent rights" is semantically less preferred than attachment to either of the deverbal NPs ($NP_1$ or $NP_2$). Furthermore, there seems to be no semantic preference for either $NP_1$ or $NP_2$, so that $NP_2$ would be the choice of the semantics/recency heuristics. Here again, however, high attachment seems to be preferred and turns out to be correct in context. Thus, a system that relies on the ordering we have identified would have a better chance of rapidly finding the appropriate attachment site.

In further work, it should be possible to refine this ordering criterion by examining its performance on a wider range of constructions, both through corpus analysis and through comprehension experiments, considering differences across attachment-site type, attaching-site type, and, perhaps, languages. Results from a combination of corpus analyses and experimental methodologies should therefore be relevant for both computational and psycholinguistic perspectives.

## ACKNOWLEDGMENTS

We would like to thank the following people for helpful discussions about the work reported here: Bob Berwick, Enriqueta Canseco-Gonzalez, David Evans, Greg Hickok, Judith Klavans, Don Mitchell, Teddy Seidenfeld, Gregg Solomon, audiences at MIT and Carnegie Mellon, and an anonymous reviewer. Of course, none of the views expressed here necessarily reflects the views of these people, and all remaining errors are our own. The second author was supported by an NSF graduate fellowship administered with funds provided to MIT under NSF grant RCD 9054772. Parts of this chapter were completed with equipment provided by the McDonnell-Pew Center for Cognitive Neuroscience at MIT.

## REFERENCES

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language processing: Psychological, computational and theoretical perspectives* (pp. 320–358). Cambridge, England: Cambridge University Press.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30,* 73–105.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

de Marcken, C. G. (1990). Parsing the LOB corpus. *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics,* 243–251.

Evans, D. A., Handerson, S. K., Lefferts, R. G., & Monarch, I. A. (1991). *A summary of the CLARIT project* (Tech. Rep. No. CMU-LCL-91-2). Pittsburgh, PA: Carnegie Mellon University, Laboratory for Computational Linguistics.



Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 729–796). Cambridge, MA: MIT Press.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 303–330). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–177). New York, NY: Academic Press.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1993). *Cross-linguistic attachment preferences: Evidence from English and Spanish*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA.

Gilboy, E., Sopena, J. M., Clifton, C., Jr., & Frazier, F. (1993). *Argument structure and association preferences in Spanish and English complex NPs*. Unpublished manuscript, University of Barcelona, Barcelona, Spain.

Grefenstette, G. (in press). SEXTANT: Extracting semantics from raw text. *Heuristics: Journal of Knowledge Engineering*.

Hays, W. (1988). *Statistics* (4th ed.). Orlando, FL: Holt, Rinehart & Winston.

Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics, 19,* 103–120.

Hirst, G. (1987). *Semantic interpretation and the resolution of ambiguity*. Cambridge, England: Cambridge University Press.

Hobbs, J. R., & Bear, J. (1990). Two principles of parse preference. *Proceedings of the Thirteenth International Conference on Computational Linguistics, 3,* 162–167.

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

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 4, pp. 155–179). New York, NY: Academic Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Langendoen, T., & Langsam, Y. (1987). On the design of finite state transducers for parsing phrase-structure languages. In A. Manaster-Ramer (Ed.), *Mathematics of language* (pp. 191–235). Philadelphia, PA: John Benjamins.

Mitchell, D. C. (in press, 1994). Sentence parsing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375–409). New York, NY: Academic Press.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Conference proceedings: Current issues in natural language processing* (pp. 1–12). Austin, TX: University of Texas at Austin.

Mitchell, D. C., Cuetos, F., & Zagar, D. (1990). Reading in different languages: Is there a universal mechanism for parsing sentences? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 285–302). Hillsdale, NJ: Lawrence Erlbaum Associates.

Murdock, B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.