Exposure-Based Models of Human Parsing: Evidence for the Use of Coarse-Grained (Nonlexical) Statistical Records

Don C. Mitchell,^{1,4} Fernando Cuetos,² Martin M. B. Corley,¹ and Marc Brysbaert³

Accepted August 11, 1995

Several current models of human parsing maintain that initial structural decisions are influenced (or tuned) by the listener's or reader's prior contact with language. The precise workings of these models depend upon the 'grain,' or level of detail, at which previous exposures to language are analyzed and used to influence parsing decisions. Some models are premised upon the use of fine-grained records (such as lexical cooccurrence statistics). Others use coarser measures. The present paper considers the viability of models based exclusively on the use of fine-grained lexical records. The results of several studies are reviewed and the evidence suggests that, if they are to account for the data, experience-based parsers must draw upon records or representations that capture statistical regularities beyond the lexical level. This poses problems for several parsing models in the literature.

INTRODUCTION

There is a substantial body of evidence showing that an individual's parsing decisions are influenced in some way by his or her prior contact with com-

Aspects of this work were supported by ESRC grant No. R0023 4062 to Don Mitchell, by a Spanish Government grant DGICVT No. PB-92-0656-C04-02 to Fernando Cuetos, and by the Belgian National Fonds voor Wetenschappelijk Onderzoek, of which Marc Brysbaert is a Research Associate. We are grateful to Chuck Clifton, Barbara Hemforth, Martin Pickering, Matt Traxler, and an anonymous reviewer—all of whom made helpful comments on an earlier draft of the paper.

¹ University of Exeter, Devon, United Kingdom.

- ² University of Oviedo, Oviedo, Spain.
- ³ University of Leuven, Leuven, Belgium.
- ⁴ Address all correspondence to Don Mitchell, Department of Psychology, University of Exeter, Exeter, Devon EX4 4QG, United Kingdom.

parable strings or structures. There is good evidence that the choice of one structural interpretation over another can be influenced by a very recent decision in the analysis of a sentence with comparable structure (e.g., Branigan, Pickering, Leversedge, Stewart, & Urbach, 1995; Carey, Mehler, & Bever, 1970; Frazier, Taft, Roeper, & Clifton, 1984; Mehler & Carey, 1967). In addition, influences over a period of 2 weeks have been demonstrated by Cuetos, Mitchell, and Corley (in press) (see Mitchell, 1994, p. 398, for a brief summary). These results suggest that parsing decisions may depend in part on the person's history of experience with the structures under scrutiny, in which case there may be a general bias in favor of analyses that occur most frequently in the language (a general proposal suggested, amongst others, by Bates & MacWhinney, 1987; Bever, 1970; Cuetos & Mitchell, 1988; Cuetos et al., in press; Ford, Bresnan, & Kaplan, 1982; MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell and Cuetos, 1991a; Mitchell, Cuetos, & Corley, 1992; Tanenhaus & Juliano, 1992 Trueswell, Tanenhaus, & Kello, 1993). One particular version of this proposal, termed the linguistic tuning hypothesis, is that structural ambiguities are initially resolved on the basis of stored records relating to the prevalence of the resolution of comparable ambiguities in the past (cf. Cuetos & et al., in press; Mitchell & Cuetos, 1991a).

The present paper is concerned with exposure-based phenomena of this kind.

THE "GRAIN" PROBLEM

Any algorithm that makes decisions on the basis of past experience depends in part on procedures for recording and storing relevant features of that experience. In addition, to implement the decisions, there must be a process for recovering "appropriate" records and using this information to execute the resulting action. The success of this process depends upon establishing a useful link between aspects of the current material and corresponding features of the established records. This is essentially a category selection or pattern-matching problem. In the case of an exposure-based parsing mechanism this problem is complicated by the fact that almost any kind of linguistic pattern could potentially be exploited in procedures of this kind. For example, consider sentences like (1a) and (1b):

- (1) a. Someone stabbed the wife of the football star who was outside the house.
 - b. Someone stabbed the estranged wife of the moviestar outside the house.

In selecting a host for the ambiguous relative clause who was ... in (1a), an actuarial decision process might refer to the accumulated records tabulating frequencies of attachment to the two alternative nouns in all structures of the form NP-PP-RC. Likewise resolutions of sentences like (1b) might be decided by referring to records for NP-PP-PP sequences. Arguably, however, the records could pool the statistics for both structures, keeping figures for all structures of the form NP-PP-(modifying constituent), and indeed the categories of classification could be even more coarse grained than this (e.g., with records being kept for NP-Modifier 1-Modifier 2 structures of the number of times Modifier 2 attaches to the NP rather than Modifier 1). Equally, it is easy to imagine a system in which the records are more detailed and more fine grained than those initially proposed above. For example, there could be separate bins for NP-PP-RC structures where the preposition in the PP is of [as, for examples, in (1a) and (1b)] or with, by, and so on. There could be different tabulations when the first NP has no adjectives [as in (1a)], when it has one adjective [as in (1b)], etc. There could be distinct records when the whole NP-PP-modifier complex appears before or after the main verb of the sentence, or when one or other of the host nouns is human, animate, or inanimate; and it is even possible that separate statistics are kept for modifier attachments to individual nouns (e.g., wife, moviestar, servant, actress, etc.).

In principle, then, it may be possible to keep records of the way ambiguities are resolved over an almost infinite range of grains. Given any arbitrary choice of record-keeping detail it would be possible to formulate the outline of an exposure-based account of ambiguity resolution as follows: Faced with a structure that needs to be disambiguated, first classify it in terms of the prestored categories assembled for this purpose; second, "read" off the structural interpretation that predominates in the accumulated records (perhaps in the form of a partial syntactic representation); third, copy this into the developing structure for the current sentence. Setting aside the obvious computational problems there would be in "merging" prestored frames with the new content of the material under analysis, an account of this kind could in principle make firm predictions about ambiguity resolution preferences. However, as a theory of parsing this remains unsatisfactory because it is clear that the precise predictions depend intimately on the exact grain of record-keeping underpinning the model. A system that maintains different records for different kinds of NPs or PPs will have a database which allows, indeed forces, it to offer different solutions to one which pools such records over time.

It follows that if an exposure-based model is to make any firm predictions, it must first make some kind of commitment on the grain of recordkeeping in its database. In the literature to date it is difficult to identify explicit commitments of this kind. There are advantages and disadvantages of both fine- and coarse-grained record keeping. The more detailed the records, the higher the chance that the stored information can be used to predict the correct analysis in any new sentence. However, highly detailed records would presumably be costly to maintain in computational terms, and at the limits it would be possible to make the grain so fine that many of the bins contained no information at all—because the reader or listener has never before encountered that precise structure. Coarse records would be easier to maintain, but the pooling of information would inevitably lead to loss of precision.

In the discussion that follows we shall use a vocabulary which presupposes that statistical information is represented in one particular (oversimplified) form. This is not intended to convey a theoretical commitment to this specific form of representation. It is merely used as a presentational device to explore the issues raised by systems of this kind. We shall treat the statistical information as being represented in a "record" or "table" made up of two parts: (1) a subtable containing values which stand for the frequencies with which the actuarial parsing device has encountered each competing interpretation of the structural ambiguity in question and (2) a contingency statement indicating any linguistic constraint under which the counts are enumerated. Table I gives an indication of the kind of detail that might be included in records of this type. In principle we do not envisage any limit on the kinds of linguistic/logical statement that might be entered into the contingency column.

Of course it is most unlikely that transparency of this kind would be evident in any biologically plausible system for accumulating experience. For example, in a connectionist system the "records" would consist of a set of activation weights distributed throughout the network. In models of

	Number of attachment to		
	First host	Second host	Contingent upon
Record 1	350ª	600	Zero adjectives in first NP [as in (1a)]
Record 2	180	315	One adjective in first NP [as in (1b)]
Record 3	3	5	First $N = wife$ [as in (1a)]
Record 4	30	80	First NP has gender = feminine; number = singular [as in both (1a) and (1b)]
Record 5	210	400	Preposition in $PP = of [as in both (1a) and (1b)]$

 Table I. Illustrations of Various Types of Contingent Statistical Records for NP-PP-Modifier Structures

^a All counts are entirely hypothetical and are merely entered to indicate possible patterns of data.

this kind contingencies and regularities may be just as important despite the distributed nature of the information stored.

LEVEL OF GRAIN IN CURRENT PARSING MODELS

A variety of different grains of record keeping are presupposed in current exposure-based models of parsing. At one extreme it is possible to identify a small number of models which propose that parsing proceeds by referring only to the very finest grain of statistical record. In the middle ground there are models which draw on both coarse-grained and fine-grained (e.g., lexical) statistics, basing their initial parsing decisions on varying balances between the two kinds of source. At the other extreme, there are accounts that focus almost exclusively on coarse-grained regularities, disregarding structural contingencies that may be associated with individual words in the sentence. Examples of each kind of account are outlined below.

(I) Models Using Fine-Grained Information Alone

An example of a purely fine-grained approach to sentence processing can be found in a recent paper by Spivey-Knowlton and Sedivy (1995). These authors argued that once lexical and discourse effects have been partialed out there does not seem to be any role at all for higher-level structural regularities. Consistent with this position, Spivey-Knowlton (1994) has described a computational model of ambiguity resolution which carries out its calculations without referring to any kind of statistic that is pooled over lexical items. Briefly, the activation levels for competing linguistic structures (i.e., alternative interpretations of ambiguous material) are calculated as the linear weighted sum of item-specific values for the stimulus string in question. The measures for the competing structures are then fed into an algorithm designed to resolve the competition (and thereby predict processing difficulties in on-line studies). Since the current (partial) implementation of this model is based on isolated calculations using context-specific or lexically specific records, it is not subject to the generalization effects that characterize connectionist systems with weights trained by repeated exposure to input materials. This genuinely seems to be a model which relies exclusively on fine-grained records for resolving structural ambiguities (see below for further discussion of connectionist generalization).

(II) Mixed-Grain Models

There are several models which make use of records over a range of different grains. Closest in spirit to the models of Type I above are the

numerous statements that have been made by researchers associated with Rochester University (e.g., Boland & Tanenhaus, 1991; Boland, Tanenhaus, & Garnsey, 1990; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994; Tanenhaus, Carlson, & Trueswell, 1989; Tanenhaus, Garnsey, & Boland, 1990: Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell et al., 1993). These proposals are often set up in opposition to models which highlight coarse-grained regularities [such as Frazier's (1987) garden-path model] and they place overwhelming emphasis on the importance of fine-grained (lexical) statistics, as can be illustrated by terms used in the titles of papers (e.g., use of verb control information-Boland et al., 1990; verb-specific constraints-Trueswell et al., 1993; combinatory lexical information-Tanenhaus, Boland, Mauner, & Carlson, 1993; Tanenhaus et al., 1990; lexical projection-Boland, Tanenhaus, Carlson, & Garnsey, 1989, etc.). The emphasis on fine-grained records is further illustrated by the frequent appeal to material ratings and corpus statistics categorized and enumerated at the lexical level (e.g., Tabossi et al., 1994; Tanenhaus & Juliano, 1992; Trueswell et al., 1994; Trueswell et al., 1993).

While there is no question that these proposals place considerable emphasis on fine-grained (lexical) influences in parsing, their implicit (and on occasions explicit) connectionist structure would, by the very nature of interactive connectionist approaches, seem to necessitate some kind of role for coarse grained records.⁵ This is neatly illustrated by Juliano and Tanenhaus's (1994) recent connectionist implementation of the "constraint-based lexicalist" model-which provides a clear facility for encompassing coarsegrained effects. In this model an input layer of 208 lexical items (plus an eight-unit Jordan net) is fed through eight hidden units to a small bank of units representing syntactic categories. In this network the only place where item-specific (lexical) information can be stored is in the set of weights between the input and hidden layers. In the network's training phase all other weights are subject to back-propagated adjustment following each and every input. As a result of the algebraic properties of the back-propagation algorithm, these weights in the second layer must inevitably integrate information over broader categories than those entailed in the first (mainly lexical) layer of connections. It follows that when the system is in its final (fully trained) state its treatment of a particular combination of words will be influenced not only by the record of weight adjustments associated with prior exposure to those identical strings, but also by revisions brought about by integrations over numerous other inputs. In short, the mere introduction of a (small) hidden layer in the model (together with the use of a back-

⁵ Such a role for supralexical regularities is explicitly acknowledged by Trueswell et al. (1993) (p. 548).

propagation learning algorithm) is enough to ensure that the output will be influenced by statistical input/output relations beyond the lexical level.⁶ To the extent that this happens, the model can be viewed as stating that initial parsing decisions are based on regularities other than those associated with particular lexical items (and this despite the fact that this is a view often contested by Tanenhaus and his colleagues, though promoted by others: e.g., Frazier, 1987; Mitchell, 1987, 1989, 1994). To the extent that this particular implementation is representative of the Rochester position, it would appear that, however strong the emphasis on lexical statistics, the underlying model is, in essence, a mixed-grain model.

Another somewhat earlier example of this kind of parsing theory is one put forward by Ford et al. (1982). In this model performance is assumed to be affected (in different circumstances) both by principles of "lexical preference" and by coarser-grained principles of "syntactic preference." Of these, the first was held to predominate, with the model giving priority to parsing decisions governed by specific properties of individual verbs. Thus, the authors suggested that "alternative lexical forms differ in their strength, perhaps because of variations in frequency of usage" (Ford et al., 1982, Sec. 3) and that the relative "strengths" of the competing verb forms are used to guide or control structural decisions. While Ford and her colleagues did not attempt to spell out the exact form in which verb-strength information was encoded in the system, it is quite clear that for this primary aspect of processing the mechanism they had in mind was one that depends crucially on reference to fine-grained lexical records.

In contrast with the dominant profile of lexical statistics, the model assigned a relatively minor role to coarse-grained "syntactic preference" information. The parsing theory is quite explicit in stipulating that such information is only used as a last resort (or "default") in circumstances where detailed lexical information fails to provide helpful guidance. With these specifications the model qualifies as one that makes use of a variety of different grains while placing major reliance on fine-grained (lexical) statistics.

In a more recent model of this kind MacDonald et al. (1994) outlined a number of lexically based network mechanisms for resolving a range of different kinds of structural ambiguity. As with any connectionist system of

⁶ Similar observations apply to the connectionist model of Pearlmutter, Daugherty, Mac-Donald, and Seidenberg (1994), notwithstanding the fact that MacDonald et al. (1994) cited this as an example of the kind of model that implements "exactly the kinds of lexical and probabilistic information" which they endorse (p. 699). The mere use of hidden units and back-propagation learning ensures that the output of the model is influenced by nonlexical regularities—contrary to the entire spirit of the MacDonald et al. (1994) paper.

this kind, the activation weights in an implemented version of this model would undoubtedly be influenced by the model's training set, capturing the fine-grained statistical properties of the words it is exposed to during "learning." To the extent that the local lexical networks capture this information, the model can be regarded as drawing upon fine-grained lexical information. Furthermore, in a discussion of the relative reliance of different kinds of information (see MacDonald et al., 1994, pp. 697-698), the authors pointedly offered fine-grained interpretations of each of the ostensibly nonlexical phenomena they discussed—leaving the strong impression that an implicit purpose of their theory is to avoid postulating mechanisms that exploit regularities beyond the lexical level. However, on p. 697 the authors explicitly refrained from ruling out the possibility that higher-level syntactic regularities may be represented independently of the lexicon-and hence that the parser might also make use of coarse-grained measures in making at least some of its decisions. So, as an explicit statement, the model has to be categorized as one that allows for the use of a range of grains, while placing overwhelming emphasis on fine-grained lexical measures.

The mixed-grain models considered in this section have all emphasised fine-grained (lexical) statistics over coarser-grained statistics. However there is nothing to prevent a mixed-grain model from emphasizing any level of statistics; ultimately in interactionist accounts it becomes a question of the relative weights assigned to different types of statistical information.

(III) Coarse-Grained Models

Models of this kind are based on the premise that fine-grained information is specifically excluded from all records relating to the prevalence of different readings of ambiguous structures. In one example of this kind, Cuetos et al. (in press) put forward a variant of the linguistic tuning hypothesis to account for cross-linguistic parsing preferences in sentences like (1a) and (1b) (see also Brysbaert & Mitchell, in press; Corley & Corley, 1995; Mitchell & Cuetos, 1991a). Put simply, the proposal is that in sentences of this type readers and listeners initially attach the relative clause to whichever of the competing host positions has been modified most frequently by relative clauses in the past. Because previous exposures to language determine current behavior, tuning accounts are sometimes referred to as "exposure-based." By hypothesis, the statistical records that are consulted in this process are held to exclude all details of the nouns (or NPs) that occupied these positions in past encounters with the sentence frame. The initial parsing decisions are based on tallies that are integrated over details of this kind. While there is nothing in the current general formulation of the tuning hypothesis to rule out the retention and use of more fine-

grained statistics as well as (or instead of) such coarse-grained measures, at least one of the present authors (D.C.M.) favors variants of the model which ignore such information in initial decision-making (cf. Mitchell, 1994). On such an account all lexical influences in parsing are interpreted not as first-pass effects, but as the results of later filtering phenomena (for details see Mitchell, 1987, 1989; for related proposals see Frazier, 1987, 1989).

Given the range of models and proposals outlined above, the following questions arise: Is it viable to propose that record keeping in statistically driven parsing is based solely on aggregations of detailed (lexical) tabulations in a statistically driven model of parsing? Or, alternatively, is it necessary to assume that at least part of the statistical information used to resolve ambiguities is based on counts using categories that are higher than the lexical level? For the remainder of the paper we shall examine empirical evidence which suggests that higher-level categories play a crucial role in parsing processes.

EVIDENCE AGAINST (EXCLUSIVELY) FINE-GRAINED RECORD KEEPING

The evidence presented here will be based predominantly from data on ambiguity-resolving preferences in sentences like (1a)—repeated here for the reader's convenience:

(1) a. Someone stabbed the wife of the football star who was outside the house.

The reason for concentrating on this particular ambiguity is that it provides a situation in which lexical constraints are likely to be weak. The use of adjuncts (like relative clauses) in a sentence is usually entirely optional and therefore not so intimately governed by lexical considerations as might be the case with the verb-argument attaching ambiguities which dominate much of the current parsing literature. With sentences like (1a) lexical influences are unlikely to swamp and obscure all others. It follows that these materials would seem likely to provide a good opportunity to demonstrate any higher-level effects that might exist.

Basically, the parsing problem in (1a) is that the relative clause (*who was outside the house*) can reasonably be interpreted as modifying two different host nouns in the sentence (i.e., *wife* and *football star*). The parser is therefore faced with the problem of deciding which of these to attach the final clause to in the first place. The evidence suggests that in most languages with complex noun phrases including genitive PPs the initially preferred attachment is to the site within the higher NP rather than the noun within

the PP⁷ (the major exceptions being English, and perhaps Italian, German, and Swedish in certain circumstances: see Cuetos et al., in press, for an extensive review of the evidence on this topic). Given that there is a clear preference in most languages, the problem for any exposure-based parser is to identify the types of statistical records which would account for the biases. (This is not to deny that there are viable alternatives to exposure accounts as detailed in the Cuetos et al. (in press) review).

For a fine-grained model which fails to record syntactic information such biases may be difficult to explain. This type of model would not have any records to help it distinguish between attachment points in their different syntactic environments. To make a decision it would have to fall back on lexically related biases—perhaps selecting *football star* rather than *wife* as host because its tabulations show that the former is more likely to take a modifier of a particular kind. This is, in fact, exactly the proposal MacDonald et al. (1994, pp. 697–698) put forward to account for ambiguity resolution in sentences of this kind.

On the face of it this proposal is implausible and unlikely to provide a satisfactory account of the data. On-line studies have frequently produced a pattern of results in which there is no detectable attachment to one site (N2) at the same time as there is reliable evidence of links to the other (N1) (see Cuetos et al., in press; Mitchell & Cuetos, 1991a, 1991b). Within the framework of a model based exclusively on the use of lexical records, this would imply that the materials were so badly designed that almost every noun in the first position was more readily modified (on the evidence of statistical records) than each noun in the second position. Moreover, to account for the data in numerous replications by other authors in other labs, the proponents would have to argue that the same bias occurred in each new experiment carried out to examine the phenomenon (for details of replications see Cuetos et al., in press).

In addition to the problems just outlined, there is more direct evidence that this account of the data is unsatisfactory. In a recent study, Corley and Corley (1995) examined clause-attachment biases in materials which enabled

⁷ Gilboy et al. (1995) have presented off-line evidence that attachment to the higher site (N1) in Spanish may not be equally marked for all classes of complex noun phrases incorporating genitive PPs. Nevertheless their data confirmed that for NPs including the word *de* (*of*) their subjects' overall preference remained the same as that in earlier investigations of this structure in Spanish [i.e., high (N1) attachment]. At the time of writing these variations across different types of *de* phrase have not been corroborated in on-line studies and so it is not known whether they are characteristics of initial attachment or of later operations such as revision and reanalysis. For these reasons it is not clear whether these findings have any direct bearing on the present discussion of initial attachment preferences.

the noun position to be counterbalanced over the experiment. Examples are shown in (2a) and (2b):

- (2) a. The satirist ridiculed the lawyer of the firm wh . . . (human noun first; nonhuman noun second)
 - b. The satirist ridiculed the firm of the lawyer wh . . . (nonhuman noun first; human noun second)

Subjects were instructed to complete each sentence with a clause starting with either who or which, and these continuations were used to assess whether the partially seen relative clauses at the ends of the sentence segments had been interpreted as being attached to one host rather than the other. Three weeks after the initial test 24 of the subjects were given a parallel version of the materials, in which all noun pairs were reversed for each subject. If MacDonald et al. (1994) had been correct in interpreting these biases in lexical terms alone there would have been no bias in relation to attaching clauses to nouns in either of the two positions as the two nouns appeared equally often in each position over the entire study. In fact there was a reliable bias in favor of attaching the relative clause to the second noun in both test phases and a reliable (p < .001) positive correlation between the two sessions (r = .794),⁸ indicating that subjects who opted to modify one noun position in Session 1 tended to attach the clause to the same position 3 weeks later. The fact that they did this rather than reverse the attachment in order to modify the same noun as before indicates that it is something about the structural *position* of the noun that is attracting a modifier and not (or at least, not only) some lexical/statistical property of the noun itself.

As a fallback position MacDonald et al. (1994) might have argued that something is added to the statistical records on noun modification. For example, the records indicating how much each noun attracts a modifier may be refined so that they are contingent upon the linear position of the noun within the preceding string. On this extension of the proposal, any noun late in the string would have a stronger (or weaker) "modifier attracting tendency" than nouns in earlier positions. While something of this kind could presumably be achieved (albeit at some cost in record proliferation), this kind of proposal in turn fails to provide an explanation of certain other data. Mitchell and Cuetos (1991b) showed that the relative attractiveness of a

⁸ The correlation statistic cited here is the figure for sentences of the form (2a) and (2b) for subjects who participated in the two off-line sessions as well as a series of on-line tests. The statistics reported by Corley and Corley (1995) included data from a further sentence type not considered here and, in some cases, the data from a larger sample of subjects (N = 52) including 28 who did not complete the on-line tests and whose data are not included here.

noun as a host depends not only on its linear position within the string but on the type of syntactic structure it is embedded within (see also Cuetos et al., in press). Specifically, the second noun (e.g., *chica*, or *girl*) is strongly *dispreferred* when it appears within a prepositional phrase [as in (3a) below]. However, it is strongly *preferred* when it appears within a relative clause conveying essentially the same semantic content [as in (3b)]:

- (3) a. Pedro miraba los libros de la chica que ...(Peter was looking at the books of the girl who/which ...)
 - b. Pedro miraba los libros que pertencian a la chica que ...
 (Peter was looking at the books which belonged to the girl who/which ...)

Any statistical model seeking to account for this finding will have to go beyond postulating mechanisms which compile statistics exclusively on the basis of lexical distinctions and on the basis of the linear order of words. (For further arguments along the same lines see Hemforth, Konieczny, & Scheepers, 1994, who argued that attachment biases in German vary depending upon whether the modifying structure is a relative clause or a prepositional phrase.)

EVIDENCE AGAINST SUBDIVISION OF NP RECORDS

If record keeping at the lexical level is not sufficient to explain the data, one possibility is that an individual's statistical database contains a single table summarizing the accumulated statistics for all encounters with structures of the form NP-PP-RC. In this case all ambiguities of this form would have to be resolved by referring to the same central record.

However, this is not the only possibility. Instead, the general tabulations for the NP-PP-RC structure might be broken down into a number of subclassifications. For example, the attachment statistics for NPs starting with a definite article might be recorded separately from those for indefinite NPs, or NPs without any article at all. Proper nouns (which are presumably modified relatively rarely) might be entered separately from other categories of noun. Distinctions might be drawn between animate and inanimate hosts, human and nonhuman nouns, and so on.

While there is currently little empirical evidence which can be used to examine these possibilities, there is some preliminary evidence that certain types of subclassification may be ignored for the purposes of ambiguity resolution.

French Evidence

Using corpus data derived from the newspaper Le Monde, Charolles, Kister, and Saltazart (Michel Charolles, personal communication) have

shown that with NP-PP-RC structures there are different relative-clause attachment patterns for different types of noun phrases. In particular, when the complex NP is of the form (definite article-Noun 1-de-definite article-Noun 2)-as in *le sable de la plage* (*"the sand of the beach"*), then in corpus cases where this complex was followed by a relative clause, in 135 of the 213 resolvable cases (53.4%) the attachment was to the second of the two nouns. However, this pattern changed with different kinds of NPs (e.g., NPs with indefinite articles or with no articles at all). Overall statistics showed that when all types of NPs were pooled the resulting bias was reversed [with only 250 out of 719 (34.9%) clauses being interpreted as being attached to the second noun].

Let us suppose that these corpus statistics provide a reasonably accurate reflection of the general pattern of biases that individual French listeners and readers might encounter in their language. If there were distinct records for different subclasses of noun phrases one might expect that an exposurebased parsing mechanism might opt for Noun 2 attachment in the case of definite NPs and Noun 1 attachment when the NP-PP complex contains certain other kinds of noun phrases. However, this is not what happens (at least in the case of the first prediction). Using both eve-tracking procedures and subject-paced reading, Zagar and his colleagues have consistently found that, contrary to the corpus statistics, relative clauses are attached to the first of the two alternative sites in complex NPs of the form le sable de la plage (Zagar & Pynte, 1992; Zagar, Pynte, & Rativeau, 1995; see also Mitchell, Cuetos, & Zagar, 1990). This pattern of results is clearly not what would have been expected if the parser were a statistical device capable of using fine-grained records. In contrast, while the empirical evidence is too sparse to provide strong support for coarse-grained records, the data are perfectly compatible with the proposal that attachment data for all occurrences of NP-PP-RC structures are pooled or integrated into a single record. Such an account would be consistent with the proposal that record keeping may generally be undertaken at the level of phrasal categories.

Dutch Evidence

Further evidence that attachment decisions may not be contingent on details relating to the internal features of noun phrases comes from recent work on relative clause attachment in Dutch (see Brysbaert & Mitchell, in press). Like readers of Spanish, French, and several other languages, Dutch readers show a tendency to resolve the attachment ambiguity by interpreting the relative clause as modifying the earlier of the two potential hosts. For example, Brysbaert and Mitchell used subject-paced reading and eye-tracking tasks to show that subjects took longer to process the disambiguation region (in italics below) when it forced attachment to the more recent host [as in (4b)] than when it forced attachment to the earlier head [as in (4a)].

- (4) a. De gangsters schoten op DE zoon van DE actrice DIE op het balkon zat met zijn arm in het gips.
 [The terrorists shot THE (nonneuter) son of THE (nonneuter) actress WHO (nonneuter) was on the balcony with his arm in a cast.]
 - b. De gangsters schoten op DE zoon van DE actrice DIE op het balkon zat met *haar arm in het gips*.
 [The terrorists shot THE (nonneuter) son of THE (nonneuter) actress WHO (nonneuter) was on the balcony with *her arm in a cast*.]

For present purposes, however, a more interesting finding was one that occurred when one of the competing nonneuter noun phrases was replaced by a singular neuter noun phrase as in (5a) and (5b):

- (5) a. De gangsters schoten op HET zoontje van DE actrice DAT op het balkon zat met zijn arm in het gips.
 [The terrorists shot THE (neuter) little son of THE (nonneuter) actress WHO (neuter) was on the balcony with his arm in a cast.]
 - b. De gangsters schoten op HET zoontje van DE actrice DIE op het balkon zat met *haar arm in het gips*.
 [The terrorists shot THE (neuter) little son of THE (nonneuter) actress WHO (nonneuter) was on the balcony with *her arm in a cast*.]

In these examples *de zoon (the son)* is replaced by a diminutive form (*het zoontje*). A crucial effect of this change is that it eliminates the clause attachment ambiguity. This is because RC attachment is subject to gender agreement between the host NP and the relative pronoun. Thus, clauses starting with the singular neuter relative pronoun (i.e., *dat*) can only be attached to the neuter head (*het zoontje*), whereas those headed by the pronoun *die* [like (5b)] are *precluded* from being attached to hosts of this kind. Given these observations, it would seem reasonable to suppose that faced with unambiguous sentences like (5a) and (5b) the parser would have been able to avoid the problems precipitated by opting for the wrong analysis [as it appears to do in the ambiguous cases in (4a) and (4b) above]. By using the gender information effectively, it might have been expected to opt immediately for the correct analysis—eliminating the need for any revisions in the last few words of the sentence. In fact, both eye-tracking and subject-paced reading measures showed that the difference between the final por-

tions of (5a) and (5b) was statistically indistinguishable from that between (4a) and (4b). In other words, the evidence suggests that, on the first pass at least, the parser pays absolutely no attention to the gender information initially opting for exactly the same analysis as it would have pursued in the absence of this information.

Like the results summarized in earlier sections, this surprising finding is readily explained if one assumes that initial parsing decisions in cases like this are based largely on coarse-grained records. So the suggestion is that the parser encounters a sequence of structures of the form NP-PP-RC and resolves the (apparent) ambiguity by referring to recorded tallies of actual attachments in the past. By hypothesis, this favors attachment to the first NP and so the parser adopts this analysis *despite the fact that it is not compatible with the fine-grained (gender) information*. As with the earlier studies, then, the Dutch data provide evidence for relatively coarse-grained mechanisms in initial parsing decisions.

Evidence from Other Structures

In the sections above our discussion has focused on the procedures readers use to resolve relative clause attachment ambiguities. This might give the impression that the use of coarse-grained statistics may be restricted to parsing problems associated with this kind of ambiguity. However, the evidence against exclusively fine-grained mechanisms extends to other structures as well. For example, in a recent study Traxler and Pickering (1995) used eye-tracking procedures to examine parsing preferences in reduced complement constructions like those in (6a) and (6b):

- (6) a. The athlete realized her goals would be far out of reach.
 - b. The athlete realized her shoes would be far out of reach.

Careful pretesting procedures were used to identify a sample of verbs that show a strong bias in favor of reduced complement rather than direct object readings. For example, with the verb *realized* production studies revealed that subjects were much more likely to interpret a noun phrase following the verb as the subject of a reduced complement [as in (6a) and (6b)] than to attach it to the verb directly as a simple object, as in the sentence: *The athlete realised her goals*.

Given verbs of this kind, a fine-grained (i.e., lexical level) exposurebased mechanism might have been expected to draw upon stored information about the structural preferences associated with the particular verb under scrutiny. If this had happened, it would presumably have led to systematic choice of the reduced complement reading, given the intentional bias of the verbs included in the study. On these assumptions, there would have been no more difficulty in reading noun phrases at the beginning of complements like that in sentence (6a) compared with the noun phrase starting the complement in (6b). In fact, Traxler and Pickering's (1995) eye-tracking data showed that *her shoes* was more difficult to read than *her goals*, suggesting that the NP following the verb is initially interpreted as a direct object rather than the start of a complement structure. The fact that this happened in the face of the lexical biases of the individual verbs in the sentences suggests, at least within exposure-based frameworks, that in this case as in the examples above, coarse-grained rather than fine-grained records must have been consulted to make the structural decisions. The finding is easily explained if the tallies consulted are ones which pool information on usage of the different structural forms over *all* verbs which share the same ambiguity. A tabulation of this kind would show a strong bias in favor of the direct object interpretation, providing a ready explanation of the results of the study.

A similar appeal to the use of coarse-grained statistics may provide an explanation of failures to take account of detailed lexical information in resolving other forms of structural ambiguity—for example the apparent tendency to interpret a noun phrase following a verb as its direct object even when (i) the verb is classified as being intransitive and is highly unlikely to take a direct object (cf. Adams, Clifton, & Mitchell, 1991; Mitchell, 1987, 1989, 1994) or (ii) or when a direct object interpretation becomes implausible because an ergative verb follows an inanimate subject (as in Clifton, 1993). As before, results of this kind can be interpreted as evidence that parsing proceeds not by referring to individual lexical records but by tapping into pooled statistical values aggregated over wide classes of verbs (see Juliano & Tanenhaus, 1994, for an alternative gloss on the same general proposal and Konieczny & Strube, in press, for a somewhat different account of this phenomenon).

IMPLICATIONS

The evidence surveyed in this paper makes it clear that statistical or constraint-based parsing devices can be viable only if they are constructed in such a way that they are capable of using records beyond the lexical level. A system which works exclusively with fine-grained records would not be capable of accounting for numerous findings in the literature. To this extent the evidence is clearly in conflict with the fundamental assumptions of the more radical "lexicalist" models such as those listed in Category I, above.

The results are compatible with models classified as falling into the other two categories. However, given the relative lack of precision of many

current accounts, it is not a trivial matter to distinguish between these competing positions. Coarse-grained (Category III) models are consistent with all the evidence presented here and, as mentioned above, one of us (D.C.M.) has consistently argued that they provide a more parsimonious account of parsing data more generally (e.g., Mitchell, 1989, 1994, etc.). However, the case for models of this kind remains hotly contested (e.g., Trueswell et al., 1993) and it is likely that further work will need to be carried out before this issue is eventually resolved. Within the mixed-grain class of models authors frequently seek to draw further, more subtle distinctions concerning the balance of emphasis placed respectively on fine- and coarse-grained statistics. However, no one has suggested a metric that might capture and describe this balance. In the absence of such a measure it is not obvious how one could establish whether parsing processes are "predominantly" driven by reference to fine-grained or by coarse-grained exposure records. One consequence of this is that there is no overt rationale for dubbing any mixedgrain model "lexicalist" (notwithstanding the fact that this appears to have become common practice). Indeed, apparent indications in the past that one or other grain may dominate could well be associated with the fact that different studies concentrate on the analysis of different linguistic structures. It may be that parsing processes as a whole are not dominated by the use of one grain at the expense of another (on any putative balance measure), but rather that fine-grained measures play a relatively important role in resolving certain structures (e.g., verb argument ambiguities) while coarsegrained measures play an equally dominant role in others (e.g., adjunct ambiguities). On this analysis, the widespread current emphasis on finegrained (lexical) effects in the literature may simply be due to the overreliance on verb-argument ambiguities in much of the empirical work to date.

Overall, then, the studies reviewed here confirm that early parsing choices can be determined by high-level statistical regularities of the language. This, of course is one of the earliest insights in work on parsing, and has remained a dominant theme since the topic was first investigated in the 1960s. Human parsing is characterized by a tendency for preliminary choices to play down, and even sometimes to ignore, potentially informative (and often fine-grained) features of the text. In one of the earliest proposals, Bever (1970) suggested that certain structural analyses were tackled by the application of a heuristic procedure termed the *canonical order strategy*—which basically involved assigning one particular kind of syntactic structure on meeting any N-V-N sequence (regardless of the individual lexical items comprising the sequence). The recent work on exposure-based parsing reaffirms that such proposals capture important insights about the parsing process. Several tasks remain in developing models of this kind. Further empirical studies can be carried out to demonstrate that grains of one kind or another can exert an influence in early parsing decisions. However, a complete account of the decision process would need to do more than enumerate the factors that enter into the equation. It would have to establish a way of assigning appropriate weights to each of the terms representing different grains, and it would have to do this for a wide range of different linguistic structures. Only when this is achieved will we have a fully fledged model of exposure-based parsing. In the meantime there is little alternative to using much less detailed proposals (such as the tuning hypothesis) to provide a broad general framework for interpreting experience-related effects in parsing.

REFERENCES

- Adams, B. C., Clifton, C., Jr., & Mitchell, D. C. (1991). Lexical guidance in sentence parsing. Poster presented at the meeting of the Psychonomics Society, San Francisco.
- Bates, E., & MacWhinney, B. (1987). Competition variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Earlbuam.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), Cognition and the development of language (pp. 279-360). New York: Wiley.
- Boland, J. E., & Tanenhaus, M. K. (1991). The role of lexical representation in sentence processing. In G. B. Simpson (Ed.), Understanding word and sentence (pp. 331– 366). Amsterdam: Elsevier.
- Boland, J. E., Tanenhaus, M. K., Carlson, G., & Garnsey, S. M. (1989). Lexical projection and the interaction of syntax and semantics in parsing. *Journal of Psycholinguistic Research*, 18, 563–576.
- Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, 29, 413–432.
- Branigan, H., Pickering, M. J., Liversedge, S. P., Stewart, A. J. & Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24 (this issue).
- Brysbaert, M., & Mitchell, D. C. (in press). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology.*
- Carey, P. W., Mehler, J., & Bever, T. G. (1970). Judging the veracity of ambiguous sentences. Journal of Verbal Learning and Verbal Behavior, 9, 243-254.
- Clifton, C. E., Jr., (1993) Thematic roles in sentence parsing. Canadian Journal of Experimental Psychology, 47, 222-246.
- Corley, M. B., & Corley, S. (1995). Cross-linguistic and intra-linguistic evidence for the use of statistics in human sentence processing. Unpublished manuscript, University of Exeter.
- 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.

- Cuetos, F., Mitchell, D. C., & Corley, M. M. B. (in press). Parsing in different languages. In M. Carreiras, J. Garcia-Albea, & N. Sabastian-Galles (Eds.), *Language processing* in Spanish. Hillsdale, NJ: Erlbaum.
- Ford, M., Bresnan, J. W., & Kaplan, R. M. (1982). A competence based theory of syntactic closure. In J. W. Bresman (Ed.), *The mental representation of grammatical relations* (pp. 727–796) Cambridge, MA: MIT Press.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), Attention and performance XII (pp. 601-681). Hillsdale, N.J.: Erlbaum.
- Frazier, L. (1989). Against lexical generation of syntax. In W. Marslen-Wilson (Ed.), Lexical representation and process (pp. 505-528). Cambridge, MA: MIT Press.
- Frazier, L., Taft, L., Roeper, T., & Clifton, C. (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory & Cognition*, 12, 421-430.
- Gilboy, E., Sopena, J. M., Clifton, C. Jr., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. Cognition, 54, 131-167.
- Hemforth, B., Konieczny, L., & Scheepers, C. (1994). Principle-based and probabilistic approaches to human parsing: How universal is the human language processor? In H. Trost (Ed.), *Tagungsband KONVENS '94* (pp. 161-170). Berlin: Springer.
- Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject-object attachment preference. Journal of Psycholinguistic Research, 23, 459-471.
- Konieczny, L., & Strube, G. (1995). SOUL: A cognitive parser. In Proceedings of the 17th Annual Conference of the Cognitive Science Society. (pp. 631–636). Mahwah, NJ: Erlbaum.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Mehler, J., & Carey, P. W. (1967). The role of serface and base structure in the perception of sentences. *Journal of Verbal Learning and Verbal Memory*, 6, 335–338.
- Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), Attention and performance XII (pp. 601-681). Hillsdale, NJ: Erlbaum.
- Mitchell, D. C. (1989). Verb guidance and lexical effects in ambiguity resolution. Language and Cognitive Processes, 4, 123-154.
- Mitchell, D. C. (1994). Sentence parsing. In M. Gernsbacher (Ed.), Handbook of psycholinguistics (pp. 375-409). New York: Academic Press.
- Mitchell, D. C., & Cuetos, F. (1991a). The origins of parsing strategies. In C. Smith (Ed.), *Current issues in natural language processing*. Austin: Center for Cognitive Science, University of Texas.
- Mitchell, D. C., & Cuetos, F. (1991b). Restrictions on late closure: The computational underpinnings of parsing strategies in Spanish and English. Unpublished paper, University of Exeter.
- Mitchell, D. C., Cuetos, F., & Corley, M. M. B. (1992, March). Statistical versus linguistic determinants of parsing bias: Cross-linguistic evidence. Paper presented at the 5th Annual CUNY Conference on Sentence Processing, New York.
- Mitchell, D. C., Cuetos, F., & Zagar, D. (1990). Reading in different languages: Is there a universal mechanism for parsing sentences? In D. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), Comprehension processes in reading (pp. 285-302). Hillsdale, NJ: Erlbaum.
- Pearlmutter, N. J., Daugherty, K. G., MacDonald, M. C., & Seidenberg, M. S. (1994). Modeling the use of frequency and contextual biases in sentence processing. In

Proceedings of the 16th Annual Conference of the Cognitive Science Society (pp. 699–704). Hillsdale, NJ: Erlbaum.

- Spivey-Knowlton, M. (1994). Quantitative predictions from a constraint-based theory of syntactic ambiguity resolution. In M. Mozer, P. Smolensky, D. Touretzky, J. Elman, & A. Weigland (Eds.), Proceedings of the 1993 Connectionist Models Summer School (pp. 130-137). Hillsdale, NJ: Erlbaum.
- Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. Cognition, 55, 227-267.
- Tabossi, P., Spivey-Knowlton, M., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta and M. Moscovitch (Eds.), Attention and performance XV (pp. 589-615). Hillsdale, NJ: Erlbaum.
- Tanenhaus, M. K., Boland, J. E., Mauner, G., & Carlson, G. N. (1993) More on combinatory lexical information: Thematic structure in parsing and interpretation. In G. T. M. Altmann and R. C. Shillcock (Eds.), Cognitive models of speech processing: The second Sperlonga meeting (pp. 297-319). Hillsdale, NJ: Erlbaum.
- Tanenhaus, M. K., Carlson, G., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, *4*, 211–234.
- Tanenhaus, M. K., Garnsey, S. M., & Boland, J. (1990). Combinatory lexical information and language comprehension. In G. T. M. Altmann (Ed.), Cognitive models of speech processing (pp. 383-408) Cambridge, MA: MIT Press.
- Tanenhaus, M. K., & Juliano, C. (1992, March). What to do about 'that': Use of cooccurrence information in parsing. Paper presented at the 5th Annual CUNY Conference on Sentence Processing, New York.
- Traxler, M. J., & Pickering, M. J. (1995, March). Evidence against statistical parsing. Poster presented at the 8th Annual CUNY Conference on Sentence Processing, Tucson.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 528– 553.
- Zagar, D., & Pynte, J. (1992, September). The role of semantic information and of attention in processing syntactic ambiguity: Eve-movement study. Paper presented at the 5th conference of the European Society for Cognitive Psychology, Paris.
- Zagar, D., Pynte, J., & Rativeau, S. (1995). Effects of context and strategy in human sentence parsing: The depth-tuning hypothesis. Manuscript submitted for publication.