CHAPTER 19

Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences

J. L. McCLELLAND and A. H. KAWAMOTO

MULTIPLE CONSTRAINTS ON ROLE ASSIGNMENT

Like many natural cognitive processes, the process of sentence comprehension involves the simultaneous consideration of a large number of different sources of information. In this chapter, we consider one aspect of sentence comprehension: the assignment of the constituents of a sentence to the correct thematic case roles. Case role assignment is not, of course, all there is to comprehension, but it reflects one important aspect of the comprehension process, namely, the specification of who did what to whom.

Case role assignment is not at all a trivial matter either, as we can see by considering some sentences and the case roles we assign to their constituents. We begin with several sentences using the verb *break*:

(1) The boy broke the window.
(2) The rock broke the window.
(3) The window broke.
(4) The boy broke the window with the rock.
(5) The boy broke the window with the curtain.

We can see that the assignment of case roles here is quite complex. The first noun phrase (NP) of the sentence can be the Agent (Sentences 1, 4, and 5), the Instrument (Sentence 2), or the Patient

(Sentence 3). The NP in the prepositional phrase (PP) could be the Instrument (Sentence 4), or it could be a Modifier of the second NP, as it is in at least one reading of Sentence 5. Another example again brings out the ambiguity of the role assignment of with-NPs:

(6) The boy ate the pasta with the sauce.
(7) The boy ate the pasta with the fork.

In (6) the with-NP clearly does not specify an Instrument, but in (7) it clearly does.

Before we go much further, it should be said that there is no universally accepted set of case roles, nor universal agreement as to the correct assignment of constituents to roles. We have adopted conventions close to those originally introduced by Fillmore (1968) in "The Case for Case," but we do not think the details are of crucial importance to the behavior of our model. Later we will suggest ways in which an extension of our model might circumvent certain of the difficulties involved in specifying the correct assignment of cases.

These complications aside, it appears from the examples that the meaning of the words in these sentences influences the assignment of arguments to roles. However, the placement of NPs within the sentences is also very important. Consider these two cases:

(8) The vase broke the window.
(9) The window broke the vase.

Here we must rely on word-order constraints. That such constraints are very strong in English can be seen from sentences like:

(10) The pencil kicked the cow.

Even though semantic constraints clearly would indicate that the cow is a much more likely Agent and the pencil a much more likely Patient, Sentence 10 simply is not given this interpretation by adult readers who are native speakers of English.

Word-order constraints like those illustrated by (10) are very strong in English, but it is important to realize that such heavy reliance on such constraints is not universal. Bates and MacWhinney (in press; MacWhinney, Bates, & Kliegl, 1984) have shown that adult speakers of Italian will assign roles to sentences like (10) based predominantly on semantic constraints;[1] word order plays a very limited role and

[1] We use the phrase "semantic constraints" to refer to the constraints language users impose on the co-occurrence of constituents in particular roles in case-level representations. In the model, as we shall see, these constraints arise from the co-occurrences of constituents in the experiences the model is exposed to.

determines assignment only when semantics and case-marking inflections give no information.

As the work of Bates and MacWhinney amply demonstrates, case role assignment is influenced by at least three different kinds of factors: word order, semantic constraints, and (when available) inflectional morphology. Reliance on any one of these constraints is a matter of degree, and varies from language to language. In addition to these factors, there is one more that cannot be ignored, namely, the more global context in which the sentence is presented. Consider, for example, Sentence 11:

(11) The boy saw the girl with the binoculars.

We get one reading if prior context tells us "A boy was looking out the window, trying to see how much he could see with various optical instruments." We get quite a different one if it says "Two girls were trying to identify some birds when a boy came along. One girl had a pair of binoculars and the other did not." Crain and Steedman (1985) have experimentally demonstrated contextual influences on parsing decisions.

While the fact that word order and semantic constraints both influence role assignment has often been acknowledged (Bever, 1970; Fodor, Bever, & Garrett, 1974), there are few existing models that go very far toward proposing a mechanism to account for these effects. However, there are some researchers in language processing who have tried to find ways of bringing semantic considerations into syntactic processing in one way or another. One recent approach has been to rely on the lexicon to influence both syntactic processing and the construction of underlying functional representations (Ford, Bresnan, & Kaplan, 1982; Kaplan & Bresnan, 1982; MacWhinney & Sokolov, in press). Ford et al. (1982) considered cases like the following:

(12) The woman wanted the dress on the rack.
(13) The woman positioned the dress on the rack.

They noted that the preferred reading of the first of these had *on the rack* as a modifier of *the dress*, while the preferred reading of the second had *on the rack* as a locative argument of *positioned*. To account for this difference in role assignment, they proposed two principles: (a) *lexical preference* and (b) *final arguments*. Basically, lexical preference establishes an expected argument structure (e.g., Subject–Verb–Object in the case of *want*; Subject–Verb–Object–Prepositional Object in the case of *positioned*) by consulting an ordered list of possible argument structures associated with each verb. If a constituent is encountered that could fill a slot in the expected argument structure, the constituent is treated as an argument of the verb. However, if a constituent is encountered that appears to satisfy the conditions on the final argument of the expected argument structure, its attachment is delayed to allow for the incorporation into the constituent of subsequent constituents. Thus, with *want*, the NP *the dress* is a candidate for final argument and is not attached directly as a constituent of the VP; rather, a superordinate NP structure containing *the dress on the rack* is ultimately attached to the VP. With *position*, however, *the dress* would not be the final argument, and so is attached directly to the VP and closed. *On the rack* is then available for attachment as the final argument to the VP.

While this scheme certainly does some of the work that needs to be done in allowing the constraints imposed by the words in a sentence to influence role assignment, we do not think it goes nearly far enough. For as we saw in Sentences 4–7, the NPs of a sentence also influence syntactic decisions. Oden (1978) has verified that all three NPs in sentences like these influence subjects' role-assignment decisions.

In the literature on sentence processing, no one disputes that various factors influence the final reading that is assigned to a sentence. However, there are various views of the way in which these factors are taken into account on-line. Kurtzman (1985) argues that the parsing process is directly guided by an ongoing plausibility analysis; Marslen-Wilson and Tyler (1981) have pioneered this sort of view, and they stress the immediacy with which syntactic, semantic, and pragmatic considerations can all be brought to bear on the course of sentence processing. On the other hand, Frazier and her colleagues (e.g., Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983) argue that the syntactic parser imposes its preferred structuring on the sentence based only on syntactic considerations, passing the results of this processing on quickly to a thematic interpreter that can reject the syntactic parse in favor of a thematically more appropriate reading.

Whichever view one holds, it is clear that a mechanism is needed in which all the constituents of a sentence can work simultaneously to influence the assignment of roles to constituents. While we ourselves tend to favor a highly interactive view, the model we will describe here takes as its input a partial surface parse (though it is one that leaves certain attachment decisions unspecified) and generates from it a case-level representation. Intended extensions of the model, which we will describe below, would incorporate feedback to the syntactic structure level; but most of the model's behavior is not dependent on this feedback, and so readers committed to a less interactive view of the relation between syntactic and thematic analyses may yet find the model to be of interest.

GOALS

The primary goal of our model is to provide a mechanism that can begin to account for the joint role of word order and semantic constraints on role assignment. We wanted the model to be able to *learn* to do this based on experience with sentences and their case representations. We wanted the model to be able to *generalize* what it learned to new sentences made up of novel combinations of words.

In addition, we had several other goals for the model:

- We wanted the model to be able to select contextually appropriate readings of ambiguous words.

- We wanted the model to select the appropriate verb frame based on the pattern of arguments and their semantic features.

- We wanted the model to fill in missing arguments in incomplete sentences with plausible default values.

- We wanted the model to be able to generalize its knowledge of correct role assignment to sentences containing a word it has never seen before, given only a specification of some of the semantic properties of the word.

The model succeeded in meeting all these goals, as we shall see.

The model also exhibits an additional property that we had not actually anticipated, even though it is a central characteristic of language understanding: The model exhibits an uncanny tendency to shade its representation of the constituents of a sentence in ways that are contextually appropriate. It does this without any explicit training to do so; in fact, it does this in spite of the fact that the training inputs it receives are not contextually shaded as they would be in reality. We will examine this aspect of the model's behavior through examples, and observe how it emerges naturally from the model's structure.

The model is, of course, very far from a complete or final model of sentence processing or even case role assignment. Perhaps it is best seen as a partial instantiation of one view of what some properties of the interface between syntactic and more conceptual levels of language representation might be like. We offer the model not because it "solves the problem of sentence comprehension." Rather, we offer it because it suggests new ways of thinking about several aspects of language and language representation. The simulation model that embodies these ideas will undoubtedly require substantial development and elaboration.

It is our belief, though, that the basic principles that it embodies will prove extremely valuable as cognitive science continues to try to come to grips with the problem of understanding natural language.

We have limited the model in several ways. Most importantly, we have considered only single clause sentences. We have also considered only a limited set of roles and a limited vocabulary. Since we have restricted the analysis to English, case inflectional morphology does not arise. Within these bounds, we will see that we have been able to meet the goals of the model quite successfully, using a very simple PDP architecture.

Previous, Related Work

Both Cottrell (1985; Cottrell & Small, 1983) and Waltz and Pollack (1985) have preceded us in noting the appeal of connectionism as a means of exploiting the multiple constraints that appear to influence both case role assignment and the contextual disambiguation of ambiguous noun phrases. Their models differ from ours in several ways, most notably in that both rely primarily on local representations (one-unit–one-concept) as opposed to distributed representations, although Waltz and Pollack (1985) do suggest ways that a distributed representation could be used to represent global contextual influences on word meaning disambiguation. Within the context of distributed models, ours builds on the work of J. A. Anderson (1983) and Kawamoto (1985): Both models show how context can be used to select the appropriate reading of an ambiguous word. Our work incorporates mechanisms quite like theirs to accomplish this and other goals. Finally, Hinton's (1981a) early discussion of the use of distributed representations to represent propositions played an important role in the development of the ideas described here.

ARCHITECTURE OF THE MODEL

The role-assignment model is a distributed model, and has many properties in common with the verb learning model described in Chapter 18. The model consists of two sets of units: one for representing the surface structure of the sentence and one for representing its case structure. The model learns through presentations of correct surface-structure/case-structure pairs; during testing, we simply present the surface-structure input and examine the output the model generates at the case-structure level.

Sentences. The sentences processed by the model consist of a verb and from one to three NPs. There is always a Subject NP, and optionally there may be an Object NP. If this is present, there may also be a *with-NP*; that is, a NP in a sentence-final prepositional phrase beginning with the word *with*. All of the numbered sentences considered in the introduction are examples of sentence types that might be presented to the model.

Input format of sentences. What the model actually sees as input is not the raw sentence but a canonical representation of the constituent structure of the sentence, in a form that could be produced by a simple surface parser and a simple lexicon. Such a parser and lexicon are not, in fact, parts of the model in its present form—the sentences are simply presented to the model in this canonical format. We discuss ways such a parser could be implemented in a PDP model in the discussion section.

Semantic Microfeatures

In the canonical input format, words are represented as lists of semantic microfeatures (Hinton, 1981a; see Chapter 3; Waltz & Pollack, 1985, also make some use of a microfeature representation). For both nouns and verbs, the features are grouped into several dimensions. Each dimension consists of a set of mutually exclusive values, and, in general, each word is represented by a vector in which one and only one value on each dimension is ON for the word and all of the other values are OFF. Values that are set to be ON are represented in the feature vectors as 1s. Values that are set to be OFF are represented as dots (".").

We chose the dimensions and the values on each dimension to capture what we felt were important dimensions of semantic variation in the meanings of words that had implications for the role assignments of the words. We should be very clear about one point, though, which is that we do not want to suggest that the full range of the phenomena that are described under the rubric of the "meanings" of the words are captured by these semantic microfeatures. Indeed, we do not think of words as actually having some fixed meaning at all. Exactly how we do think of meanings will become clear after we examine the behavior of the model, so we postpone a fuller consideration of this issue until the discussion.

The full set of dimensions used in the feature sets are given in Table 1. The noun dimensions are largely self-explanatory, but the

TABLE 1

FEATURE DIMENSIONS AND VALUES

Nouns	
HUMAN	human nonhuman
SOFTNESS	soft hard
GENDER	male female neuter
VOLUME	small medium large
FORM	compact 1-D 2-D 3-D
POINTINESS	pointed rounded
BREAKABILITY	fragile unbreakable
OBJ-TYPE	food toy tool utensil furniture animate nat-inan
Verbs	
DOER	yes no
CAUSE	yes no-cause no-change
TOUCH	agent inst both none AisP
NAT_CHNG	pieces shreds chemical none unused
AGT_MVMT	trans part none NA
PT_MVMT	trans part none NA
INTENSITY	low high

Note: nat-inan = natural inanimate, AisP = Agent is Patient, NA = not applicable.

different dimensions of the verbs may need some explication. Basically, these dimensions are seen as capturing properties of the scenario specified by the verb. Thus, the DOER dimension indicates whether there is an Agent instigating the event. The CAUSE dimension specifies whether the verb is causal. If not, it indicates whether this is because there is no cause specified (as in the case of *the window broke*) or whether it is because there is no change (as in the case of *the boy* *touched the girl*). The TOUCH dimension indicates whether the Agent, the Instrument, both, or neither touches the Patient; the "AisP" value simply indicates that the Agent and the Patient are the same (as in *the cat moved*). The NAT_CHNG dimension specifies the nature of the change that takes place in the Patient. The AGT_MVMT and PT_MVMT specify the movement of the Agent and the Patient, respectively; and INTENSITY simply indicates the forcefulness of the action. The labels given to the dimensions are, of course, only for reference; they were chosen so that each noun or verb dimension would have a unique first letter that could be used to designate the dimension.

It must be stressed that we are not strongly wedded to this particular choice of features, and that other features would need to be included to extend the model to larger sets of nouns and verbs. On the other hand, the features that we did include were carefully chosen because they seemed highly relevant to determining the case role assignments. For example, the DOER dimension directly specifies whether there is or is not an Agent. Thus, the features of the verb, in particular, often have direct case-structural implications. (We would prefer a model that constructed its own semantic microfeatures using back propagation [Chapter 8] or a related method for learning, but this extension has not yet been implemented.)

Figures 1 and 2 give the vectors that we assigned to each of the words used in the model. It will be immediately noted that some of our encoding decisions were arbitrary, and that sometimes we seem to be forcing words into molds that they do not perfectly fit. Further, each feature has the same weight as all the others, and is as definite as all the others. Reality is not nearly so definite or evenhanded, of course. Balls are round, but may be soft or hard; paperweights are generally compact in shape but need not be, etc. The definiteness of the input used in the simulations is a simplification that we have adopted to make the initial coding of the input patterns as straightforward as possible. A more realistic coding would allow some features to be more definite than others. We will see that the model tends to correct this deficiency on its own accord.

One of our goals for the model is to show how it can select the contextually appropriate meaning for an ambiguous word. For ambiguous words (*bat*, flying or baseball, and *chicken*, living or cooked) the input pattern is the average of the feature patterns of each of the two readings of the word. This means that in cases where the two agree on the value of a particular input dimension, that dimension has the agreed value in the input representation. In cases where the two disagree, the feature has a value of .5 (represented by "?") in the input representation. A goal of the simulations is to see if the model can correctly fill in these unspecified values, effectively retrieving the contextually

	HU	SO	GND	VOL	FORM	PO	BR	OBJ_TYP
ball	. 1	1 .	. . 1	1 . .	1 . . .	. 1	. 1	. 1 . . . . .
fl-bat	. 1	1 .	1 . .	1 . .	. . . 1	1 .	. 1	. . . . . 1 .
bb-bat	. 1	. 1	. . 1	1 . .	. 1 . .	. 1	. 1	. 1 . . . . .
bat	. 1	? ?	? . ?	1 . .	. ? . ?	? ?	. 1	. ? . . . ? .
boy	1 .	1 .	1 . .	. 1 .	. . . 1	. 1	. 1	. . . . . 1 .
paperwt	. 1	. 1	. . 1	1 . .	1 . . .	1 .	. 1	. . . . 1 . .
cheese	. 1	1 .	. . 1	1 . .	. . 1 .	. 1	1 .	1 . . . . . .
li-chicken	. 1	1 .	. 1 .	1 . .	. . . 1	. 1	1 .	. . . . . 1 .
co-chicken	. 1	1 .	. . 1	1 . .	1 . . .	. 1	1 .	1 . . . . . .
chicken	. 1	1 .	. ? ?	1 . .	? . . ?	. 1	? ?	? . . . . ? .
curtain	. 1	1 .	. . 1	. 1 .	. . 1 .	. 1	1 .	. . . . 1 . .
desk	. 1	. 1	. . 1	. . 1	. . . 1	1 .	. 1	. . . . 1 . .
doll	. 1	1 .	. 1 .	1 . .	. . . 1	. 1	1 .	. 1 . . . . .
food	. 1	1 .	. . 1	1 . .	? ? ? ?	. 1	1 .	1 . . . . . .
fork	. 1	. 1	. . 1	1 . .	. 1 . .	1 .	. 1	. . . 1 . . .
girl	1 .	1 .	. 1 .	. 1 .	. . . 1	. 1	. 1	. . . . . 1 .
hatchet	. 1	. 1	. . 1	1 . .	. 1 . .	1 .	. 1	. . 1 . . . .
hammer	. 1	. 1	. . 1	1 . .	. 1 . .	. 1	. 1	. . 1 . . . .
man	1 .	1 .	1 . .	. . 1	. . . 1	. 1	. 1	. . . . . 1 .
woman	1 .	1 .	. 1 .	. . 1	. . . 1	. 1	. 1	. . . . . 1 .
plate	. 1	. 1	. . 1	1 . .	. . 1 .	. 1	1 .	. . . . 1 . .
rock	. 1	. 1	. . 1	1 . .	. . . 1	1 .	. 1	 1
potato	. 1	1 .	. . 1	1 . .	1. . .	. 1	1 .	1 . . . . . .
pasta	. 1	1 .	. . 1	1 . .	. 1 . .	. 1	1 .	1 . . . . . .
spoon	. 1	. 1	. . 1	1 . .	. 1 . .	. 1	. 1	. . . 1 . . .
carrot	. 1	. 1	. . 1	1 . .	. 1 . .	1 .	1 .	1 . . . . . .
vase	. 1	. 1	. . 1	1 . .	. 1 . .	. 1	1 .	. . . . 1 . .
window	. 1	. 1	. . 1	. 1 .	. . 1 .	1 .	1 .	. . . . 1 . .
dog	. 1	1 .	1 . .	. 1 .	. . . 1	. 1	.1	. . . . . 1 .
wolf	. 1	1 .	1 . .	. 1 .	. . . 1	1 .	. 1	. . . . . 1 .
sheep	. 1	1 .	. 1 .	. 1 .	. . . 1	. 1	1 .	. . . . . 1 .
lion	. 1	1 .	1 . .	. . 1	. . . 1	1 .	. 1	. . . . . 1 .

FIGURE 1. The nouns used in the model and their features. For ambiguous noun constituents, the correct, fully specified reading was used in specifying what the case role representation of the constituent should be, but the underspecified, ambiguous forms were used in the sentence-level input representation. See text for a full discussion.

appropriate missing values in the process of assigning the word to the appropriate case role. Figure 1 indicates both "full" readings of *bat* and *chicken*, as well as the ambiguous forms used as inputs.[2]

Another goal for the model is to show how it can select the contextually appropriate reading of a verb. This is handled in much the same

[2] For the concept *food*, which is taken to be the implied Patient in sentences like *The boy ate*, no particular shape seems appropriate. Therefore the intended output representation is assumed to be unspecified (as indicated by the "?") for all values on the shape dimension. For all other dimensions, *food* has what we take to be the typical values for foods.

	DO	CAU	TOUCH	N_CHG	A_MV	P_MV	IN
ate	1 .	1 . .	. . 1 . .	. . 1 . .	. 1 . .	1 . . .	1 .
ateAVP	1 .	1 . .	. . 1 . .	. . 1 . .	. 1 . .	1 . . .	1 .
ateAVPI	1 .	1 . .	. . 1 . .	. . 1 . .	. 1 . .	1 . . .	1 .
ateAVF	1 .	1 . .	. . 1 . .	. . 1 . .	. 1 . .	1 . . .	1 .
broke	1 .	1 . .	. 1 . . .	1	. 1 . .	. . 1 .	. 1
brokeAVPI	1 .	1 . .	. 1 . . .	1	. 1 . .	. . 1 .	. 1
brokeAVP	1 .	1 . .	1	1	. 1 . .	. . 1 .	. 1
brokeIVP	1 .	. 1 .	. 1 . . .	1	. . . 1	. . 1 .	. 1
brokePV	1 .	. 1 .	. . . 1 .	1	. 1 . .	. . 1 .	. 1
hit	1 .	. . 1	. 1 . . .	. . . 1 .	. 1 . .	. . 1 .	. 1
hitAVPI	1 .	. . 1	. 1 . . .	. . . 1 .	. 1 . .	. . 1 .	. 1
hitAVP	1 .	. . 1	1	. . . 1 .	. 1 . .	. . 1 .	. 1
hitIVP	. 1	. . 1	. 1 . . .	. . . 1 .	. . . 1	. . 1 .	. 1
moved	1 .	1 . .	1	. . . 1 .	1 . . .	1 . . .	1 .
movedAVP	1 .	1 . .	1	. . . 1 .	1 . . .	1 . . .	1 .
movedAVS	1 .	1 . .	 1	. . . 1 .	1 . . .	1 . . .	1 .
movedPV	. 1	. 1 .	. . . 1 .	. . . 1 .	. . . 1	1 . . .	1 .
touched	1 .	. . 1	. 1 . . .	. . . 1 .	. 1 . .	. . 1 .	1 .
touchedAVPI	1 .	. . 1	. 1 . . .	. . . 1 .	. 1 . .	. . 1 .	1 .
touchedAVP	1 .	. . 1	1	. . . 1 .	. 1 . .	. . 1 .	1 .
touchedIVP	. 1	. . 1	. 1 . . .	. . . 1 .	. . . 1	. . 1 .	1 .

FIGURE 2. The verbs used in the model and their microfeature representations. The forms followed by strings of uppercase letters (e.g., AVPI) represent the alternative feature patterns that the model must choose between as its way of specifying the contextually appropriate reading of the verb. These alternative feature patterns correspond to the semantic features of the verb appropriate for particular configurations of case roles, as indicated by the uppercase letters: A = Agent, V = Verb, P = Patient, I = Instrument, M = Modifier, S = Self, F = implied Food. The position of the letter indicates the position of the corresponding constituent in the input sentence. The patterns given with the generic verb unadorned by uppercase letters were used in the sentence-level, input representations.

way as noun ambiguity resolution. The different readings are represented by (potentially) different sets of semantic microfeatures; for example, the Agent/No-Instrument reading of *broke* (brokeAVP) involves contact between the Agent and the Patient, while the Instrument/No-Agent version (brokeIVP) and the Agent/Instrument version (brokeAVPI) involve contact between the Instrument and the Patient. The input representation of the features of a given verb is the same, regardless of context, and the task given to the model is to activate the set of features for the sentence-appropriate version. Rather than use the average pattern based on all of the different possible readings of the verb, we used a "generic" pattern for each verb, which is the pattern for what we took to be the verb's most typical case frame.

This is indicated in Figure 2 by the pattern of features next to the plain verb.[3]

The feature patterns corresponding to the different case frames the model must choose among are indicated on the lines in the table following its generic pattern. (The labels on these lines are used simply to designate the feature patterns. They indicate the roles the various arguments in the surface structure of the sentence play. Thus, brokeAVPI specifies the case frame in which the surface subject is the Agent, the surface object is the Patient, and the with-NP is the Instrument.) Note that the microfeatures of two different readings of the same verb may or may not differ, depending on whether the features of the scenario do or do not change in different case frames.

The feature vectors for the constituents of the sentence *The boy broke the window with the hammer* are shown just below the corresponding constituents at the top of Figure 3. Note that these are displayed in the order: Verb, Subject NP, Object NP, and With-NP. The row of letters below the feature vectors indicates the first letter of the name of the dimension on which each feature represents a value. For example, the first two elements of the verb feature vector are labeled *d* for the DOER dimension; the first two values of each of the three noun feature vectors are labeled *h* for the HUMAN dimension.

Sentence-structure units. The sentence-structure level representation of an input sentence is not actually the set of constituent feature vectors; rather, it is the pattern of activation these vectors produce over units that correspond to *pairs* of features. These units are called sentence-structure (SS) units.[4]

---

[3] The different handling of nouns and verbs is not a principled distinction, but an exploration of two endpoints on a continuum ranging from underspecification of the input for ambiguous words to complete specification of an input representation, regardless of the fact that the features used in the case role representation will differ as a function of context. Perhaps the idea that the features will be altered as a function of context is the best way of putting things in this case. We imagine that the true state of affairs is intermediate between these two extremes, for both nouns and verbs. In any case, the model does not have any prior commitment to the idea that the features in the input representation should be preserved in the output representation; the full prespecification simply gives the model a fuller description to work from, thereby allowing greater differentiation of the different verbs.

[4] An alternative name for these units would be "surface-structure" units, to indicate that they do not capture the notion of underlying subject, object, etc. However, we have chosen the term "sentence-structure" because, for present purposes, the information they capture is not even a full surface-structure parse of the sentence; in particular, it does not specify the attachment of the with-NP.

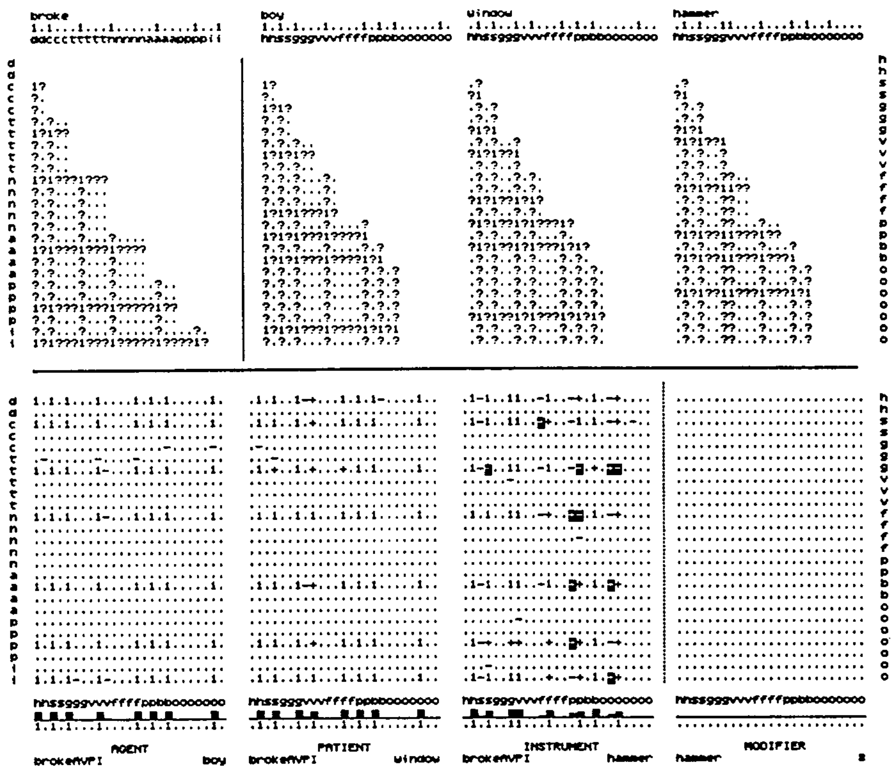

FIGURE 3. The top line of this figure displays the constituents of the sentence *The boy broke the window with the hammer* in the order: Verb, Subject NP, Object NP, and With-NP. Below these are the microfeatures of each constituent, and below these are conjunctive sentence-structure units for each constituent. Below the horizontal line are the blocks of case-structure units for the Agent, Patient, Instrument, and Modifier roles. Below these is an indication of the pattern of noun features the model is activating for each slot (represented by the vertical black bars), followed by a representation of the microfeatures of the correct filler of each slot. The last line gives the label of the correct head (verb frame or modified NP) and tail (slot filler) for each slot. See text for further explanation.

Each SS unit represents the conjunction of two microfeatures of the filler of a particular surface role. Since there are four sentence-structure roles, there are four sets of SS units. Within each set there is a unit that stands for the conjunction of every microfeature value on each dimension with every microfeature value on every other dimension. For example, for nouns there are units for:

HUMAN = yes / GENDER = male
SOLIDITY = hard / BREAKABILITY = fragile

among many others; for the verb, one of the units corresponds to

DOER = yes / TOUCH = instrument

(i.e., there is a doer—the Instrument touches the Patient).

The sentence-structure units are displayed in Figure 3 in four roughly triangular arrays. The verb array is separated from the arrays for the three NPs to indicate that different features are conjoined in the verb and NP representations.

Each array contains the conjunctive units for the constituent immediately above it. There is a unit wherever there is a 1, a "?", or a ".". Within each array, the units are laid out in such a way that the column a unit is in indicates one of the microfeatures that it stands for, and the row it is in indicates the other microfeature. Rows and columns are both ordered in the same way as the microfeature vectors at the top of the figure. The dimensions are indicated by the row of letters across the top of each array and along the left (for the verb units) or right (for the three sets of NP units). Note that the set of units in each array fills less than half of each block for two reasons. First, there are only $[n(n-1)]/2$ distinct pairs of n features; second, pairs of values on the same dimension are not included.

We considered various schemes for activating the sentence-structure units. One possible scheme would be to use a strict deterministic activation rule, so that a particular SS unit would be turned on only if both of the features the unit stands for were on in the feature vector. This use of the SS units would allow the model to learn to respond in a finely tuned way to particular conjunctions of microfeatures. However, we wished to see how well the model could function using an inherently noisy input representation. Furthermore, as discussed in Chapter 18, we knew that generalization is facilitated when units that only partially match the input have some chance of being activated. In the present case, we considered it important to be able to generalize to words with similar meanings. Therefore, the SS units were treated as stochastic binary units, like the units used in Chapter 18. Each SS unit received excitatory input from each of the two features that it stands for, and we set the bias and variance of the units so that when both of a SS unit's features were active, the unit came on with probability .85; and when neither was active, it came on with probability .15. These cases are represented in the figure by "1" and ".", respectively. Units receiving one excitatory input came on with probability .5; these units are represented in Figure 3 by "?".

The use of the SS units in conjunction with these particular activation assumptions means that the input representation the model must use as the basis for assigning words to case roles is both noisy and redundant. Each feature of the input is represented in the activation of many of the SS units, and no one of these is crucial to the representation. A drawback of these particular activation assumptions, however, is that they do not allow the model to learn to respond to specific conjunctions of inputs. While the model does well in our present simulations, we presume that simulations using a larger lexicon would require greater differentiation of some of the noun and verb representations. To handle such cases, we believe it would be necessary to allow tuning of the input connections to the SS units via back propagation (Chapter 8) so that greater differentiation can be obtained when necessary. In principle, also, higher-order conjunctions of microfeatures might sometimes be required. Our use of broadly tuned, pair-wise conjunctive units illustrates the *style* of representation that we think is appropriate for the input, but the present version is only an approximation to what we would expect a model with a tunable input representation to build for itself.

Case role representation. The case role representation takes a slightly different form than the sentence-structure representation. To understand this representation, it is useful to drop back to a more abstract viewpoint, and consider more generally how we might represent a structural description in a distributed representation. In general, a structural description can be represented by a set of triples of the form (A R B) where A and B correspond to nodes in the structural description, and R stands for the relation between the nodes. For example, a class-inclusion hierarchy can be represented by triples of the form (X IS-A Y), where X and Y are category names. Any other structural description, be it a syntactic constituent structure, a semantic constituent structure, or anything else, can be represented in just this way. Specifically, the case role assignment of the constituents of the sentence *The boy broke the window with the hammer* can be represented as:

Broke Agent Boy
Broke Patient Window
Broke Instrument Hammer

The constituent structure of a sentence such as *The boy ate the pasta with the sauce* would be represented by:

Ate Agent Boy
Ate Patient Pasta
Pasta Modifier Sauce

In a localist representation, we might represent each of these triples by a single unit. Each such unit would then represent the conjunction of a particular head or left-hand side of a triple, a particular relation, and a particular tail or right-hand side. Our more distributed approach is to allocate groups of units to stand for each of the possible relations (or roles), namely, Agent, Patient, Instrument, and Modifier, and to have units within each group stand for conjunctions of microfeatures of the first and third arguments (the head and the tail) of the triple. Thus, the triple is represented not by a single active unit, but by a pattern of activation over a set of units.

In our implementation, there is a group of units for each of the four relations allowed in the case structure. In Figure 3, the Agent, Patient, Instrument, and Modifier groups are laid out from left to right. Within each group, individual units stand for conjunctions of one microfeature of the head of each relation with a microfeature of the tail of each relation. Thus, for example, Broke–Agent–Boy is represented by a pattern of activation over the left-most square block of units. The unit in the ith row and jth column stands for the conjunction of feature i of the verb with feature j of the noun. Thus all the units with the same verb feature are lined up together on the same row, while all the units with the same noun feature are lined up together in the same column. For the Modifier group, the unit in the ith row and jth column stands for the conjunction of feature i of the modified NP and feature j of the modifier NP. Letters indicating the dimension specifications of the units are provided along the side and bottom edges.

The figure indicates the net input to each case role unit produced at the end of the training described below, in response to the sentence *The boy broke the window with the hammer.* (We will see very shortly how these net inputs are produced.) As before, a 1 indicates that the net input would tend to turn the unit on with probability (p) greater than or equal to .85, and a "." indicates that the net input would tend to turn it on with probability of .15 or less. A "+" indicates that the net input has a tendency to turn the unit on ($.85 > p > .5$), and a "−" indicates that the net input has a tendency to turn the unit off ($.5 > p > .15$).

The correct case-frame interpretation of the sentence is provided to the model by a specification that lists, for each of the four possible case roles, the label corresponding to the head and tail of the role. These are shown below each of the four blocks of case role units. The "#" is used to indicate a null slot filler, as in the Modifier role in the present example. From this it is possible to compute which units should be on in the case role representation. Here we simply assume that all the correct conjunctions should be turned on and all other units should be off.

In this example, the pattern of net inputs to the case role units corresponds quite closely to the correct case role representation of the sentence. The features of *boy* may be seen in the columns of the block of Agent units; the features of *window* in the columns of the block of Patient units; and the features of *hammer* in the columns of the block of Instrument units. The features of the Agent–Verb–Patient reading of the verb *broke* can be seen in the rows of each of these three sets of units. There are no features active in the fourth set of units, the Modifier units, because there is no Modifier in this case. In both the Agent and the Patient slots, the model tends to turn on ($p > .5$) all the units that should be on, and tends to turn off ($p < .5$) all the units that should be off. In the Instrument slot, there are some discrepancies; these are indicated by blackening the background for the offending units. All of the discrepancies are relatively mild in that the unit has either a weak tendency to go on when it should not (+ on a black background) or to go off when it should be on (– on a black background).

Several things should be said about the case-frame representations. The first thing is that the slots should not be seen as containing lexical items. Rather, they should be seen as containing patterns that specify some of the semantic properties assigned by the model to the *entities* designated by the words in the sentences. Thus, the pattern of feature values for the verb *break* specifies that in this instance there is contact between the Instrument and the Patient. This would also be the case in a sentence like *The hammer broke the window.* However, in a sentence like *The boy broke the window*, with no Instrument specified, the pattern of feature values specifies contact between the Agent and the Patient. Thus, the verb features provide a partial description of the scenario described by the sentence. The noun features, likewise, provide a partial description of the players (to use Fillmore's analogy) in the scenario, and these descriptions, as we will see later on, may actually be modulated by the model to take on attributes appropriate for the scenario in question.

Details of Sentence Processing and Learning

The model is very much like the verb learning model (Chapter 18). When a sentence is presented, a conventional computer program front-end determines the net input to each of the sentence-structure units, based on the feature vectors of the words. Each of these units is then turned on probabilistically, as described above. Each surface-structure unit has a modifiable connection to each of the case-structure units. In addition, each case-structure unit has a modifiable bias (equivalent to a connection from a special unit that is always on). Based on the sentence-structure pattern and the current values of the weights, a net input to each case-structure unit is computed; this is just the sum of the weights of the active inputs to each unit plus the bias term. Case-structure units take on activation values of 0 and 1, and activation is a probabilistic function of the net input, as in the verb learning model.

During learning, the resulting activation of each case-structure unit is compared to the value it should have in the correct reading of the sentence. The correct reading is supplied as a "teaching input" specifying which of the case role units should be on. The idea is that this teaching input is analogous to the representation a real language learner would construct of the situation in which the sentence might have occurred. Learning simply amounts to adjusting connection strengths to make the output generated by the model correspond more closely to the teaching input. As in the verb learning model, if a unit should be active and it is not, the weights on all the active input lines are incremented and the threshold is decremented. If a unit should not be active but it is, the weights on all the active output lines are decremented and the threshold is incremented. This is, of course, just the perceptron convergence procedure (Rosenblatt, 1962), whose strengths and weaknesses have been examined and relied upon throughout the book.

SIMULATION EXPERIMENTS

The most important thing about the model is the fact that its response to new inputs is strictly dependent upon its experience. In evaluating its behavior, then, it is important to have a clear understanding of what it has been exposed to during learning. We have done a number of different experiments with the model, but we will focus primarily on one main experiment.

The main experiment consisted of generating a corpus of sentences derived from the sentence frames listed in Table 2. It must be emphasized that these sentence frames were simply used to generate a set of legal sentences. Each frame specifies a verb, a set of roles, and a list of possible fillers of each role. Thus, the sentence frame *The human broke the fragile_object with the breaker* is simply a generator for all the sentences in which *human* is replaced with one of the words on the list of humans in Table 3, *fragile_object* is replaced with one of the words on the list of fragile objects in Table 3, and *breaker* is replaced with one of the words on the list of breakers in the table. It is clear that these generators do not capture all of the subtle distributional

TABLE 2

GENERATORS FOR SENTENCES USED IN TRAINING AND TESTS

Sentence Frame	Argument Assignment
The human ate.	AVF
The human ate the food.	AVP
The human ate the food with the food.	AVPM
The human ate the food with the utensil.	AVPI
The animal ate.	AVF
The predator ate the prey.	AVP
The human broke the fragile_object.	AVP
The human broke the fragile_object with the breaker.	AVPI
The breaker broke the fragile_object.	IVP
The animal broke the fragile_object.	AVP
The fragile_object broke.	PV
The human hit the thing.	AVP
The human hit the human with the possession.	AVPM
The human hit the thing with the hitter.	AVPI
The hitter hit the thing.	IVP
The human moved.	AVS
The human moved the object.	AVP
The animal moved.	AVS
The object moved.	PV

Note: Argument assignments specify the case role assignment of the constituents of a sentence from left to right. A = Agent, V = Verb, P = Patient, I = Instrument, M = Modifier, F = (implied) Food, S = Self.

properties of referents in real scenarios (e.g., the model is completely sex and age neutral when it comes to hitting and breaking things, contrary to reality), and so we cannot expect the model to capture all these subtleties. However, there are certain distributional facts implicit in the full ensemble of sentences encompassed by the generators. For example, all the breakers but one are hard, not soft (only *ball* is coded as *soft* in the feature patterns); only the humans enter as Agents into scenarios involving Instrument use; etc.

The "target" case-frame representations of the sentences were generated along with the sentences themselves. The case role assignments

TABLE 3

NOUN CATEGORIES

human	man woman boy girl
animal	fl-bat li-chicken dog wolf sheep lion
object	ball bb-bat paperwt cheese co-chicken curtain desk doll fork hatchet hammer plate rock pasta spoon carrot vase window
thing	human animal object
predator	wolf lion
prey	li-chicken sheep
food	co-chicken cheese spaghetti carrot
utensil	fork spoon
fragile_object	plate window vase
hitter	ball bb-bat paperwt hatchet hammer rock vase
breaker	paperwt ball bb-bat hatchet hammer rock
possession	ball dog bb-bat doll hatchet hammer vase

are indicated in Table 2 by the sequence of capital letters. These indicate the assignment of arguments from the sentences to the roles of Agent, Verb, Patient, Instrument, and Modifier (of the Patient).[5] Note that there are some sentences that could be generated by more than one generator. Thus, *The boy hit the girl with the ball* can be generated by the generator *The human hit the human with the possession*, in which case the ball is treated as a Modifier of the Patient. Alternatively, it may be generated by the generator *The human hit the thing with the hitter*. In this case, the ball is treated as the Instrument. Similarly, *The bat broke the vase* can be generated by *The breaker broke the fragile_object*, in which case its case-frame representation contains a

[5] Two special cases should be noted: For *The human ate*, the case frame contains a specification (F) that designates an implied Patient that is the generic food with unspecified shape, as indicated in the feature patterns displayed in Figure 1. For *The human moved* and *The animal moved*, the case frame contains a specification (S) that indicates that there is an implied Patient who is the same as the Agent (note that the sense of the verb *move* used here involves moving oneself and not one's possessions).