Ambiguous Novel Compounds and Models
of Morphological Parsing

Gary Libben, Bruce L. Derwing, and Roberto G. de Almeida

University of Alberta, Edmonton, Alberta, Canada

This paper reports on two experiments that investigated the activation of mor-
phemes in English novel compounds. All experiments employed stimuli that we
have called ''ambiguous novel compounds.'' These words (e.g., *clamprod*) have
two interpretable parses (e.g., *clam + prod* or *clamp + rod*) and thus offer an
opportunity to investigate which parses are preferred, whether both possible parses
are computed, and whether parsing procedures ''divide'' words into their morpho-
logical constituents or ''extract'' constituent representations. The results suggest
that morphological parsing does not simply divide a word into its constituents, but
rather generates multiple representations that are subsequently evaluated. © 1999
Academic Press

Companies that create names for new products such as *Powerbar* and
Dryloft proceed from the assumption that the constituent morphemes of these
novel compounds are easily and automatically accessible to readers of En-
glish. While this is almost certainly a sound and uncontroversial assumption,
identifying the mechanisms that allow for such constituent activation is con-
siderably less straightforward (see Taft & Forster 1976; Taft 1981; Bergman,
Hudson, & Eiling 1988; Libben 1998). Because novel compounds do not
have whole-word representations in the mind, a prelexical parsing procedure

seems to be required to account for the recognition of their components. Furthermore, this decomposition cannot proceed by removing affixes because, by definition, these words are constructed through the concatenation of morphological roots.

In this paper, we report on an investigation of how prelexical parsing is achieved. Our study focuses on a set of ''tough cases''—ambiguous novel compounds—which were introduced by Libben (1994) as a stimulus type that offers a window into the mechanisms of prelexical parsing. Ambiguous novel compounds are strings such as *clamprod* which have two possible parses (in this case *clam-prod* and *clamp-rod*). Libben (1994) argued that if prelexical parsing were characterized by a simple left-to-right parsing procedure, then the preferred parse for such compounds would always isolate the first possible constituent of the ambiguous compound (i.e., *clam-prod*). We term this the ''first possible parse'' hypothesis and contrast it with the ''last possible parse'' hypothesis which, in the case of *clamprod,* would generate *clamp-rod.*

Libben (1994) concluded that neither of these alternatives characterize the prelexical parsing procedure. Rather, he claimed that ambiguous novel compounds are processed through a recursive parsing procedure that results in the creation of all possible morphologically-legal representations. In the experiments reported below, this claim was tested using a new experimental paradigm that was developed to probe both constituent activation and the mechanisms of prelexical parsing.

EXPERIMENT 1: PARSING AMBIGUOUS NOVEL COMPOUNDS

Our first experiment addressed the question of whether or not native speakers of English show a tendency to assign a ''first possible parse'' to ambiguous novel compounds. The experiment employed the *morpheme recall task* which was developed to be used with large groups of participants to uncover their spontaneous parsing preferences.

Participants

Fifty-two undergraduate students in a introductory course in linguistics participated as a single group in the experiment.

Procedure

All participants were tested in a single session. The experimental stimuli were presented on a large screen in a university amphitheater using an In Focus liquid crystal display that allowed the screen of a Macintosh PowerBook 520 to be displayed to a large group of participants.

In the *morpheme recall task,* each trial had four components: an alert component, a stimulus
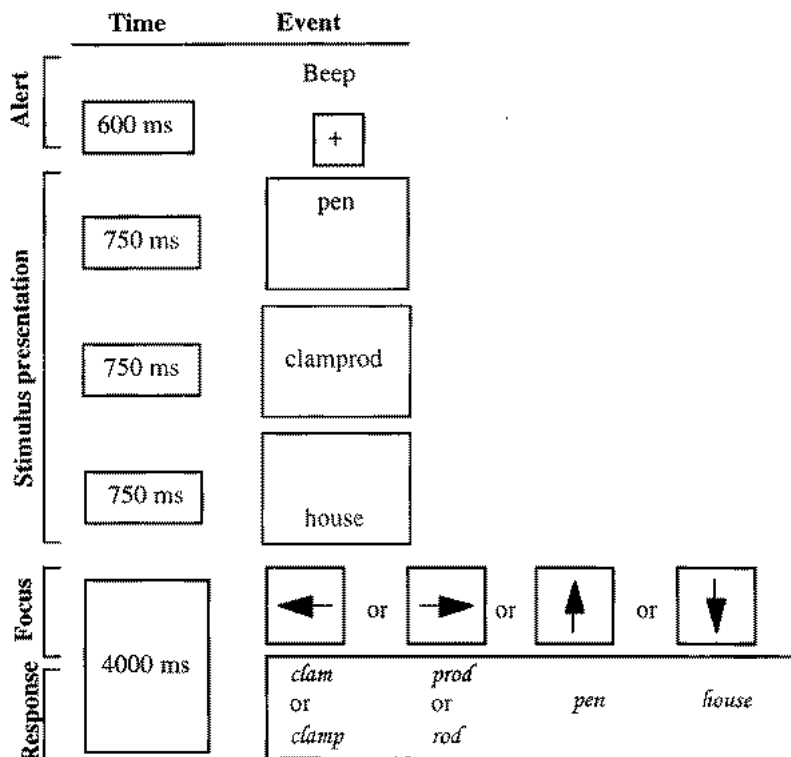
FIG. 1. A single trial in the morpheme recall task.

component, a focus component, and a response component (see Fig. 1). The alert component consisted of an auditory beep and the presentation of a ''+'' in the center of the screen. In the stimulus component, participants saw three stimuli presented in succession. The first was a monomorphemic word presented at the top of the screen. The second was a compound presented in the center of the screen, and the third was another monomorphemic word presented at the bottom of the screen. Each stimulus was centered with respect to the left and right screen margins and remained visible for 750 ms.

The stimulus component was followed by a focus indication consisting of an arrow pointing in one of four directions. In the final response component, participants were required to write down the morpheme that corresponded to the arrow direction. If the arrow pointed upward, they were to write the word presented at the top of the screen. If it pointed downward, they were required to write down the word that appeared at the bottom. In the critical ''left arrow'' or ''right arrow'' conditions, they were required to write down either the first or last morpheme of the compound word.

The experiment consisted of a ten-trial practice session and an experimental session consisting of 124 trials. The experimental session lasted 15 min and was conducted in a single block of trials. Because each trial involved the presentation of one compound and two monomorphemic words, participants were exposed to 248 monomorphemic words and 124 compounds in total. Sixty-two of these were the critical ambiguous novel compounds and each was randomly assigned to either the left or right arrow conditions. Participants were not in-

TABLE 1
Parsing Preferences for Ambiguous Novel
Compounds

	Focus	First parse	Last parse
←	(clam vs clamp)	56%	44%
→	(prod vs rod)	48%	52%
Mean		52%	48%

formed of the ambiguity of the critical stimuli until the debriefing session that followed the experiment.

Results

Thirty-one of the 62 ambiguous novel compounds were accompanied by a left arrow and 31 were accompanied by a right arrow. Parsing preferences were calculated by counting the number of first parses vs. the number of last parses for each compound. This yielded the values in Table 1.

Although overall parsing choices showed a tendency to favor first parses, this difference was not significant and analyses by items revealed substantial variation among individual ambiguous novel compounds. We explored the source of this variation in a subanalysis in which the number of first parses for each item was correlated to the semantic plausibility rating for the first parse of that item. The semantic plausibility rating was obtained from a separate group of 32 undergraduate students registered in a different section of the same introductory course in linguistics. In that task, participants were shown both ''readings'' of all ambiguous novel compounds and were asked to rate each reading on a five-point scale of plausibility. The resulting correlation between parsing preference and semantic plausibility was .74.

We conducted a similar analysis to investigate the relation between parsing choices and constituent frequency. This yielded a correlation of .42 for the number of first parses and the frequency of first parse constituents as rated by 28 undergraduate linguistics students. The correlation of parsing preference with the Kucera & Francis (1967) frequency count yielded $r = .45$.

Discussion

The results suggest that prelexical morphological parsing does not proceed in a simple left-to-right manner that is terminated by the creation of a legal and interpretable parse. As was discussed above, this predicts that participants would show consistent first possible parses, which is not what we observed. Moreover, our second finding in this experiment—that parsing choices are highly correlated with semantic plausibility—suggests that both parses were generated by the prelexical decomposition procedure. Only in this case would it be possible for the participant to ''know'' which parse

was the more plausible. We are led, therefore, to the view that the recognition of novel compounds such as *clamprod* results in the activation of the constituents of all parses (in this case, *clam, clamp, prod,* and *rod*).

EXPERIMENT 2: SEMANTIC PRIMING IN AN ACCURACY PARADIGM

The results of Experiment 1 lead to a straightforward prediction regarding semantic priming in the morpheme recall task. If indeed both *clam* and *clamp* are activated during the processing of the ambiguous novel compound *clamprod,* then this compound should prime semantic associates of each of these constituents (i.e., *sea* for *clam* and *hold* for *clamp*). This prediction was tested by modifying the morpheme recall task in the following manner: whereas in Experiment 1, the monomorphemic words that appeared above and below the critical compounds were fillers that made the recall task sufficiently difficult to yield an error rate, in Experiment 2 they were used instead as pre- and postprimes (because they either preceded or followed the critical compounds by 750 ms.). This overall design allowed us to investigate two separate questions in the same experiment: (1) whether ambiguous novel compounds facilitate recall accuracy to monomorphemic associates of both parsing choices and (2) whether the presence of semantically associated words in the presentation affects participants' choice of initial morpheme for ambiguous novel compounds (i.e., whether it influences participants' morphological parsing of those compounds).

Participants

One hundred fourteen native speakers of English registered in an introductory course in linguistics participated as volunteers in the experiment. None had participated in Experiment 1.

Procedure

Stimuli were presented to the group using the same apparatus as in Experiment 1. In this experiment, however, the participants were randomly assigned to four groups labeled by the card suits *clubs, spades, hearts,* and *diamonds.*

The alert and stimulus presentation components of the experiment were identical to those in Experiment 1. However, the focus component differed in the following manner: Instead of seeing a single arrow pointing in one of four directions, participants saw a screen such as that presented in Fig. 2. In this example, participants in the ''hearts'' group would be required to write down the monomorphemic word that appeared at the bottom of the screen and participants in the ''spades'' group would be required to write down the first ''part'' of the compound word. The position of the card suits was rerandomized for each trial so that subjects in each of the card suit groups supplied data for all morpheme positions (above, left, right, and below).

As in Experiment 1, participants completed a ten-trial practice session and an experimental session consisting of 124 trials of which 62 involved the critical ambiguous novel compounds.
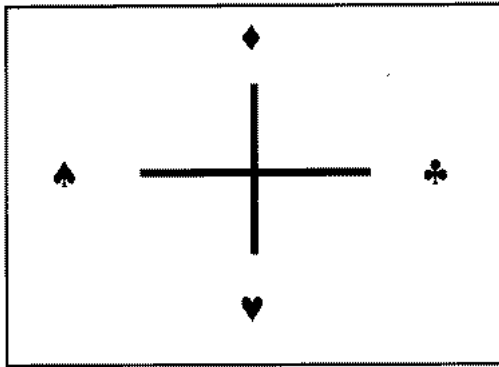
FIG. 2. The ''focus'' screen for Experiment 2.

Results

The recall of constituent associates. For each participant, accuracy scores were computed for the monomorphemic words that occurred above and below the ambiguous novel compounds. The monomorphemic words were constructed to be either semantically unrelated to any of the compound constituents, related to the first parse initial constituent (e.g., *sea* for *clam* in *clamprod*), or related to the second parse initial constituent (e.g., *hold* for *clamp* in *clamprod*). Our analysis centered on these conditions as well as the two additional conditions in which the monomorphemic words were related to the same parse (e.g., *sea-clamprod-shell*) or conflicting parses (e.g., *sea-clamprod-hold*). Table 2 summarizes these accuracy scores.

As can be seen in Table 2, accuracy scores were essentially parallel for monomorphemic words occurring above (i.e., before) and below (i.e., after) the compounds.

For both monomorphemic positions, a main effect of condition was found

TABLE 2
Accuracy Scores (Proportion Correct) for Monomorphemic Words Shown Above and Below Ambiguous Novel Compounds

Monomorphemic word	Above	Below
Unrelated to either parse	.75 (.35)	.74 (.21)
Related to first parse (other is unrelated)	.87 (.18)	.82 (.32)
Related to second parse (other is unrelated)	.85 (.19)	.85 (.27)
Related to first parse (other is related to second)	.87 (.27)	.86 (.27)
Related to first parse (other is supporting)	.90 (.30)	.83 (.32)

Note. Standard deviations are provided in brackets.

TABLE 3
Parsing Choices (Proportion of the First Possible Parses) for Ambiguous Novel
Compounds in Relation to Priming Conditions

Presentation condition	Proportion of first parses	
Unrelated to either parse	.70	(.10)
Related to first parse (other is unrelated)	.52	(.27)
Related to second parse (other is unrelated)	.81	(.15)
Related (other is contradictory)	.34	(.64)
Related to first parse (other is supporting)	.83	(.11)
Related to second parse (other is supporting)	.79	(.19)

Note. Standard deviations are provided in brackets.

($F_{above}(4,452) = 3.8$, $p = .004$; $F_{below}(4,200) = 2.7$, $p = .03$).[1] No significant differences were found among any of the related conditions. However, for both positions, single-*df* comparisons showed accuracy scores in the unrelated condition to be significantly lower than all others ($p < .002$ in both analyses).

In summary, then, ambiguous compounds primed semantic associates of both their parses in both conditions, suggesting that both parses were actually conducted. This effect was not changed by the presence of contradictory information in the stimulus presentation.

The effect of constituent associates on compound parsing. The observation that contradictory information did not inhibit the priming effects leads to the expectation that the parsing process itself is not affected by the presence of the monomorphemic primes. This is exactly what our analysis of parsing choices revealed (this analysis is essentially the same as that conducted for Experiment 1). Our analysis focused on 23 ambiguous novel compounds that showed the greatest likelihood of being affected by monomorphemic primes because they did not contain a graphemic or phonological structure that would constrain their parses. Thus ambiguous novel compounds containing digraphs (e.g., *seathorn*) and those in which different parses possessed different syllable structures (e.g., *planetrail*) were excluded from the analysis.

As can be seen in Table 3, the data do not fall into a consistent pattern in relation to the presence, absence, or combinations of monomorphemic primes. Because in this paradigm relatively few compounds were involved and because participants responded to different compounds under different conditions, we conclude that the response variance is attributable to semantic properties of compounds and the frequencies of their constituents (because phonological and graphemic factors were controlled). In any case, it seems

[1] Data points were included in the within-subjects ANOVA only if participants responded to all five conditions.

clear that the idiosyncratic properties of individual compounds cannot be overridden by priming effects in this paradigm.

GENERAL DISCUSSION

In this study we have introduced a new experimental technique that targets morphological processing in a nonchronometric, mass presentation paradigm.

The results of the two experiments that we have reported using this paradigm can be summarized in the following manner: Parsing choices for ambiguous novel compounds do not seem to be determined by the operation of the prelexical parser. Thus, readers are not uniformly led by the parser to either a first possible parse or last possible parse of an ambiguous compound. Rather, the primary function of the prelexical parser seems to be to supply all possible parses of a string. This conclusion is supported by the finding in this study that ambiguous compounds prime semantic associates of all constituents. This conclusion also follows from the results of Libben, Derwing, and de Almeida (1999) who found that naming latencies to ambiguous novel compounds such as *clamprod* were longer than those to reversals of these compounds (e.g., *prodclamp*) which contain the identical morphemes, but are rendered unambiguous as a result of the constituent reversal. Libben, Derwing, and de Almeida reasoned that this increased response latency reflects additional activity at the level of prelexical parsing as well as additional activity required at a later stage of processing in which one parse is ''chosen'' over the other.

We are left, then, with converging evidence for a view of morphological parsing in which relatively autonomous procedures interact to yield final interpretations for novel compounds. This parsing modularity was particularly evident in Experiment 2 of this study, in which we found *extended* semantic priming effects result as an outcome of prelexical parsing, but did not find that semantic priming could affect the prelexical parsing procedure itself.

This latter finding forces us, in our view, toward a re-analysis of the role of morphological parsing in the lexical processing system as a whole. From the outset, a key assumption in the literature on morphological parsing has been that a mental lexicon that is organized by morphemes would have the advantage of storage efficiency (Aitchison, 1994). It has also been assumed that the morphological parser functions as the ''front-end'' of this scheme to optimize storage efficiency. Yet in this study, we do not find evidence for a prelexical parser that is terribly concerned with efficiency. Rather, we find evidence that the parser passes on to the lexical level all possible morphemes that can be uncovered by left-to-right parsing. As was shown in Experiment 2, this parser seems not to be constructed to make use of hints (i.e., primes) that in most natural settings would also lead toward processing efficiency.

In short, our study suggests that if prelexical parsing is indeed guided by any design principle, efficiency is not it.

REFERENCES

Aitchison, J. 1994. *Words in the mind.* Oxford: Blackwell.

Berman, M. W., Hudson, P. T. W., & Eiling, P. 1988. How simple complex words can be: morphological processing and word representations. *Quarterly Journal of Experimental Psychology,* **40,** 41–72.

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

Libben, G. 1994. How is morphological decomposition achieved? *Language and Cognitive Processes,* **9**(3), 369–391.

Libben, G. 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language* **61,** 30–44.

Libben, G., Derwing, B. L., & de Almeida, R. G. (1999). Ambiguous novel compounds and prelexical parsing. Paper submitted for publication.

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior,* **20,** 289–297.

Taft, M., & Forster, K. I. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior,* **15,** 607–620.