THE HUNTING OF THE QUARK: THE PARTICLE IN ENGLISH*

CATHERINE P. BROWMAN
Haskins Laboratories

The verb-particle construction in English has two forms when occurring with a nominal direct object. The particle can precede the object, e.g., *look up the information,* or follow it, e.g., *look the information up.* Experimental evidence based on 224 subjects is presented showing that the particle tends to be contiguous to the verb when the verb-particle combination is semantically or phonologically cohesive, i.e., when the combination is idiomatic or when the particle begins with a vowel. A model within the framework of variable rules is proposed to handle the data.

Keywords: verb-particles, variable rules, discontinuous constituents, performance constraints in syntax

"The particles [are] thought to be made up of quarks . . . [which] are said to come in four flavors, and each flavor is said to come in three colors . . . One of the quark flavors is distinguished by . . . charm." From "Quarks with Color and Flavor," by Sheldon Lee Glashow, in *Scientific American,* October 1975, p. 38.

INTRODUCTION

The verb-particle construction in American English has two possible forms when combined with a non-pronominal direct object: The particle may be next to the verb, e.g., *look up the information,* or it may be separated from the verb by the direct object, e.g., *look the information up.* A number of factors affecting particle placement have been posited, including phonological complexity, semantic focus, and lexical/semantic cohesion (see Lindner, 1983, for a review). The length and complexity of the direct object has the clearest influence: In the presence of a lengthy (or "heavy") direct object the verb and particle are usually contiguous (Fraser, 1976). The effect of emphatic stress is equally clear: A stressed particle usually follows the direct object (Bolinger, 1971; Kennedy, 1920; Kruisinga and Erades, 1953; Van Dongen, 1919). Bolinger suggests this effect is part of a more general tendency for items with semantic focus to occur in final position; such a tendency also covers the observation by Kruisinga and Erades that semi-pronominal nouns tend to occur between the verb and particle, i.e., non-finally. In addition, Kruisinga and others suggest that lexically/semantically cohesive, i.e., "extra-close" or idiomatic, verb-particles tend to be contiguous.

* Thanks to Louis Goldstein, Marion Orr Harris, Larry Horn, Peter Ladefoged, Marian Macchi, David Sankoff, Vicki Tartter, and Sandra Thompson for reading and criticizing versions of this paper. The experiments and data analyses were performed at the University of California, Los Angeles between 1974 and 1977, and were supported by the National Science Foundation and National Institutes of Health.

As Butters (1973, p. 278) points out, "one would . . . need to examine an enormous amount of 'hard' data even to begin to suspect the right grammatical restrictions and performance constraints (affecting particle placement)." The above studies have begun this enormous task; but more work is needed to determine the extent to which the proposed (and other) factors affect actual particle placement for a variety of speakers in the linguistic community. The first goal of the present paper, then, is to extend the investigation of particle placement to large numbers of speakers and to conditioning factors other than extreme length and stress.

The second goal of this paper is to determine if the same conditioning factors appear in each speaker's individual grammar, and thus may be considered rules of English. Note that this is complicated by the possibility (confirmed in this paper) that individuals may differ in their overall preference for separating (or not separating) verb-particles. The solution to the problem of individual differences used in this paper is to contrast pairs of verb-particles that differ on a factor of interest – say, idiomatic vs. non-idiomatic usage. Here the hypothesis would be that idiomatic verb-particles tend to be contiguous (as in *put down Ronnie,* meaning 'to scorn,' vs. *put Ronnie down,* meaning 'to move to a low physical position'). This would be confirmed if each speaker were to separate more non-idiomatic verb-particles than the corresponding idiomatic verb-particles – even if all the verb-particles were separated more than 50% of the time. To help clarify this point, consider the following *gedanken* experiment. Suppose two speakers are tested on 20 sentences, 10 containing idiomatic verb-particles and 10 containing non-idiomatic verb-particles. Table 1 shows three possible results, all of which (if more speakers conform) support the hypothesis. Column I suggests that non-idiomatic verb-particles are obligatorily separated. Column II shows a regular but optional tendency for non-idiomatic verb-particles to be separated. Column III, like Column II, shows a tendency for non-idiomatic verb particles to be separated, but in addition there is a strong component of individual difference between the subjects. Speaker 2 rarely separates; nevertheless, s/he separates non-idiomatic verb-particles more often than idiomatic ones.

The third and final goal of the present paper is to propose a formalism for expressing the contextual conditioning. A number of different formalisms have been proposed to handle the placement of the particle. Transformational grammarians have suggested that either the contiguous form (Absalom, 1973; Chomsky, 1957; Fillmore, 1965) or the separated form (Emonds, 1972) is basic, with the other form derived through a movement transformation. Non-transformational phrase structure grammarians, such as those working within the general framework of GPSG (Generalized Phrase Structure Grammar: Gazdar, Klein, Pullum and Sag, 1985), retain the verb and particle as separate constituents, but propose an analysis that is relatively neutral with respect to placement (Jacobson, to appear). Still other grammarians view the verb and particle as forming a single constituent which may be discontinuous (Huck, 1985), or correlate positional tendencies with various regularities such as degree of analyzability of the verb-particle combination (Lindner, 1983). The type of formalism proposed in this paper is similar in spirit to these latter approaches, in that it does not presuppose that either form is necessarily more basic. Rather, it seeks to quantify the observable variation in particle placement using a formalism akin to that of variable rules (Labov, 1969; Cedergren and

TABLE 1

Possible results of hypothetical experiment

	Score (number of responses that separate verb and particle)					
	I		II		III	
	Obligatory		Optional		Variable with Individual Differences	
	Idiomatic		Idiomatic		Idiomatic	
	+	–	+	–	+	–
Subject 1	0	10	2	8	7	9
Subject 2	0	10	2	8	1	3

Sankoff, 1974; Rousseau and Sankoff, 1978). A possible alternative model is also proposed, in which the conditioning factors are used to predict the degree of lexicalization of the verb-particle combination.

EXPERIMENT 1

Factors hypothesized to affect particle placement

A number of binary factors conditioning particle placement were hypothesized on the basis of a preliminary investigation of particle placement in written texts and recorded spontaneous conversation. The 13 texts included two biographical essays from *The New Yorker,* one article and one essay from *Harper's,* one story from *Galaxy,* and eight articles from the "Travel" and "Theatre" sections of the *Los Angeles Times.* The recorded spontaneous conversations included nine hours of a 1968 New York City radio talk show, plus transcripts of seven conversations, each 15 to 30 minutes long, occurring in California in 1972-73. The radio talk show consisted primarily of the announcer plus a variety of unidentified callers. The seven conversations were all recorded by members of a graduate seminar in sociology at UCLA, under the aegis of Emmanuel Schegloff, and consisted of the students (of various ages) talking with their friends and, in one case, children.

In this corpus, separation of the verb and particle appeared to be favored by:

1) pronominal direct object, e.g., *each other, someone* (vs. nominal direct object). Personal pronouns were not included here since separation is obligatory in their presence (although, in some dialects, this may be true only when they are unstressed;

see, for example, Fischer, 1976).

2) animate direct object (vs. inanimate direct object).

3) contrastive usage, e.g., *turn the light off,* where *off* can contrast with *on* (vs. non-contrastive usage).

4) completive particles, e.g., *burn the house down* (vs. non-completive usage, e.g., *pick up the hitchhiker*).

Contiguity of the verb and particle appeared to be favored by:

5) the particle *up* (vs. any other particle). Bolinger (1971, p. 10) also notes that "the particle *up* gives trouble in phrase-terminal position even when highly literal."

6) the verb-particle combination *pick up* (vs. all other combinations).

7) verb-particle combinations used idiomatically, e.g., *put down the book,* meaning 'to criticize the book' (vs. the same combination with directional reading, e.g., *put the book down* [on the table]).

Method

Twenty-six sentences were constructed to test the various factors hypothesized to affect particle placement. A 26-item written test, designed to elicit these sentences, was administered during class to students in introductory linguistics courses at UCLA (prior to discussion of particle placement). In order to check for subject reliability, the experiment was rerun the following week on the same group. A total of 64 students completed all items on both presentations.

Each of the 26 items in the experiment comprised a context followed by a scrambled set of words that included a verb-particle combination with a direct object (see Table 2). The subjects' task was to rearrange the scrambled words into a sentence that fitted the preceding context. For the first presentation of the experiment, each set of scrambled words consisted of one phrase from Table 3 (in the order particle–direct object–verb) plus a subject and modals and/or adverbials inserted pseudo-randomly in the phrase to provide variety. Nouns generally had penultimate stress; nominal direct objects were either a proper noun or a noun plus quantifier or indefinite article (*some, a*). In those items that differed on only one factor, and thus would be compared in the analysis, the direct objects had similar morphological and phonological patterns. Seven verb-particle combinations were used; each verb and each particle were monosyllables.

To reduce the possibility that similarity of responses for the two presentations was due to subjects' remembering the first presentation, the second presentation was a week later, and differed from the first in several ways. For 25 of the 26 items, the wording (but not the relevant meaning) of the context was changed. In the set of scrambled words the particular lexical items used for the subjects, adverbials, and direct objects were also changed, while retaining the same order of scrambling, number of syllables, stress, morphology, and values of the hypothesized factors. Thus, for example, as can be seen in Table 2, for both presentations of item no. 16, the direct object was second to last in the scrambled words, had two syllables and penultimate stress, was monomorphemic, and was an animate nominal (in fact, a proper noun).

TABLE 2

Experiment 1: Sample items

Example of idiomatic usage:

Presentation 1	*Presentation 2*
16. Situation: Jane hates Harry's guts.	16. Situation: Bart thinks that Barry makes terrible spaghetti.
Words: always Jane down Harry puts	Words: always Bart down Barry puts
Your sentence:________________	Your sentence:________________

Example of adverbial usage (-idiomatic):

24. Situation: Sam hates to carry his daughter Sally on his shoulders for very long.	24. Situation: Sue's daughter Joni cries as soon as she is taken from her crib and held.
Words: always Sam down Sally puts quickly	Words: always Sue down Joni puts quickly
Your sentence:________________	Your sentence:________________

Each response was coded as 0 if the particle and verb were contiguous, or 1 if the particle and verb were separated by the direct object.

Results

First, each subject received two scores, one for each presentation, where a subject score was the sum of responses to the 26 items for that subject on that presentation. Thus a subject who consistently separated verbs and particles received a score close to 26 (for one presentation). The distribution of the subject scores for presentations 1 and 2 is shown in Figure 1. Note that both curves (drawn by hand) are unimodal and display relatively normal distributions. That is, a few subjects consistently separated the verb and particle, while a few subjects consistently placed the particle next to the verb. But most of the subjects showed variability in their responses, i.e., responded differently to different items, in response to variations in the linguistic environment. The group of subjects as a whole showed a slight preference for not separating the verb and the particle on both presentations, although the mode shifted slightly toward separation on the second presentation.

Then each item received two scores, again one for each presentation, where an item

TABLE 3

Experiment 1: Verb–particle–noun phrase combinations

	Factors						
	1. nominal DO	2. inanimate DO	3. noncontrastive particle	4. noncompletive particle	5. *up*	6. *pick up*	7. idiomatic usage
1. pick up Samantha	+	–	+	+	+	+	+
2. hold up someone	–	–	+	+	+	–	–
3. turn off Sarah	+	–	–	+	–	–	+
4. leave out something	–	+	+	+	–	–	–
5. hold up Sarah	+	–	+	+	+	–	+
6. burn down an apartment	+	+	+	–	–	–	
7. pick up an adult	+	–	+	+	+	+	+
8. turn off something	–	+	–	+	–	–	+
9. hold up someone	–	–	+	+	+	–	+
10. put down the Bible	+	+	+	+	–	–	+
11. pick up paper	+	+	+	+	+	+	–
12. turn off their lights	+	+	–	+	–	–	+
13. burn up a playhouse	+	+	+	–	+	–	
14. leave out Indians	+	–	+	+	–	–	+
15. pick up some marijuana	+	+	+	+	+	+	+
16. put down Harry	+	–	+	+	–	–	+
17. pick up Tommy	+	–	+	+	+	+	–

18. leave out something	–	+	+	+	–	–	+
19. hold up Wells Fargo	+	+	+	+	+	–	+
20. leave out someone	–	–	+	+	–	–	+
21. put down Harper's	+	+	+	+	–	–	–
22. burn up Patricia	+	–	+	+	+	–	+
23. pick up an adult	+	–	+	+	+	+	–
24. put down Sally	+	–	+	+	–	–	–
25. hold up Sammy	+	–	+	+	+	–	–
26. pick up a hitchhiker	+	–	+	+	+	+	+

score was the sum of responses by 64 subjects to that item for that presentation. The 26 item scores (for both presentations), shown in Figure 2, ranged from zero to 63, and were extremely reliable from presentation 1 to presentation 2. For example, 63 of the 64 subjects separated the verb and particle in item no. 24 in both presentations. This item reliability was due to subject reliability (not shown on the graph). On the average, 91% of the subjects gave the same response to an item during presentation 2 as during presentation 1. For the single least reliable item (no. 21), 78% of the subjects responded the same on presentations 1 and 2; on the most reliable items (no. 3, no. 15, and no. 26), 98% of the subjects responded consistently to the two presentations. Thus, in general, subjects' responses varied if and only if the environment varied.

The effects of the hypothesized factors were determined by comparing the item scores for selected pairs of items, where for any factor, the items in each selected pair differed only in the value of that factor. The results of the comparisons are shown in Table 4, where significance was computed using a chi-square test. Three conditions significantly favored contiguity of the verb-particle: the particle *up*, the combination *pick up*, and the idiomatic usage of the combinations. The effects of the other four hypothesized factors were limited to only a few pairs, suggesting that any possible effects are specific to the lexical item (e.g., contrastive usage in "turn Sarah off"; completive usage in "burn a playhouse up").

Discussion

This experiment demonstrated that the factors conditioning particle placement operate with surprising consistency within each individual's grammar (although individuals differ in their overall preference for separation or contiguity). Three factors, in particular, appeared to favor contiguity of the verb and particle–idiomatic meaning of the verb-particle, the particle *up*, and the combination *pick up*.

There were several problems, however, in interpreting these results. First of all, the length and stress pattern of the direct objects were not completely controlled, although all the direct objects were between two and five syllables long. It is possible that this lack of control contributed to the non-significance of several of the factors, since seven of the twelve non-significant pairs did not have identical lengths, and of these, five pairs

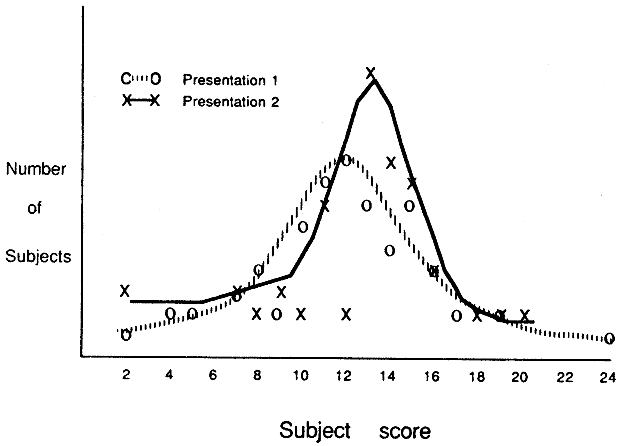

Fig. 1. Distribution of separated items (Experiment 1). The subject score is the number of items with separated verb-particles.

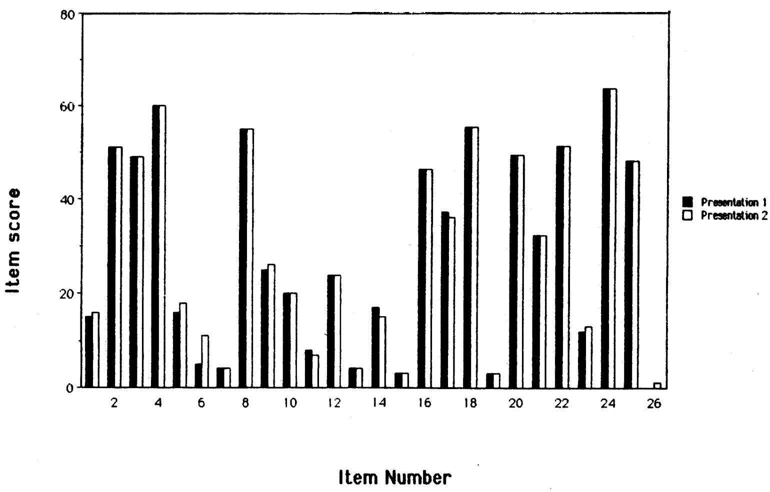

Fig. 2. Item reliability (Experiment 1). The item score is the number of subjects separating the verb and particle.

TABLE 4

Experiment 1: Effects of factors using pairwise comparison

Factors

1. nominal/pronominal DO	2. inanimate/animate DO	3. noncontrastive/contrastive particle	4. noncompletive/completive particle	5. "up"/not "up"	6. "pick up"/not "pick up"	7. idiomatic/adverbial usage
Significant Difference ($p < 0.01$)						
14/20	21/24	1/3	13/12	1/3	1/22	1/17
		5/3		1/16	1/3	7/23
				5/3	1/16	5/25
				5/16	15/12	9/2
				9/20	17/15	16/24
				25/24	17/24	26/23
No Significant Difference						
5/9	15/7	16/3	6/12		1/5	18/4
25/2	15/26	18/8	6/15			
		22/3	13/15			

had the length difference in the "wrong" direction, although the difference was only one or two syllables in each case. Therefore, a second experiment was designed in which the factors of length (and stress pattern, for completeness) were rigorously controlled in order to determine if minimal differences in these factors might indeed affect particle placement.

Secondly, two of the significant factors, *up* and *pick up*, were idiosyncratic in nature. It clearly would be desirable to extrapolate from these idiosyncratic factors to more general factors, if possible. Therefore, the second experiment was also designed to test possible candidates for more general factors.

One candidate for a more general factor is the phonological shape of the particle, especially its initial phoneme. Most verbs in verb-particle combinations (using Makkai, 1972) end in consonants; perhaps then vowel-initial particles such as *up* are potentially more cohesive with the consonant-final verb than are consonant-initial particles, due to syllable-structure constraints. That is, consonant-consonant sequences will in many cases be obligatorily heterosyllabic, whereas consonant-vowel sequences are potentially ambisyllabic (Kahn, 1976). This possibility is particularly inviting because it has also been noted that "it is the phonological shape of a verb that determines to a large extent whether or not it can combine with a particle" (Fraser, 1976, p. 13).

A second candidate for a more general factor is frequency of usage. Those verb-particle combinations such as *pick up* that are frequently used may tend to be more cohesive. A third candidate is the existence of a corresponding nominal, e.g., *I have to make a pickup,* which may influence the verbal form through either indirect or direct lexical linkage.

These three candidates for more general factors, together with the previously established factor of lexical/semantic cohesion (idiomaticity) and the controls for length and stress pattern of the direct object, were tested in a second experiment with tighter experimental design.

EXPERIMENT 2

Hypothesized factors

Factors of the verb-particle combination hypothesized to favor contiguity of the verb and particle were:

1) idiomatic (vs. adverbial) where idiomaticity was generally determined by Makkai (1972, pp. 213-253).
2) vowel-initial particle (vs. consonant-initial particle).
3) corresponding nominal (vs. no corresponding nominal), where the status of the corresponding nominal was generally determined using Makkai (1972, pp. 213-253).
4) high frequency (vs. low frequency), where frequency of usage was determined by summing the individual frequencies of the verb, the particle, and the verb-particle combination from Kučera and Francis (1967).

Properties of the direct object hypothesized to favor contiguity of the verb and particle were:

5) three syllables (vs. two syllables).
6) first syllable stress (vs. second syllable stress).

Method

The same task as in the first experiment was used, although without the replication. The experiment was administered in another introductory linguistics class at UCLA, prior to discussion of verb-particles. One hundred and sixty subjects completed the task.

Sixty-four items were constructed, each consisting as before of a context and a set of scrambled words, where the scrambled words included one of the direct objects listed in Table 5. Because phonology and stress were being tested, the rhythm and phonological pattern of the desired response were strictly controlled. Each verb and each particle

TABLE 5

Experiment 2: Verb-particle combinations and direct objects

Verb-particles

		Idiomatic	Initial Vowel	Corresponding Nominal	High Frequency
1.	pick up	+	+	+	+
2.	pick up	–	+	+	+
3.	work out	+	+	+	–
4.	work out	–	+	+	–
5.	put down	+	–	+	+
6.	put down	–	–	+	+
7.	set back	+	–	+	–
8.	set back	–	–	+	–
9.	show up	+	+	–	+
10.	show up	–	+	–	+
11.	put out	+	+	–	–
12.	put out	–	+	–	–
13.	keep down	+	–	–	+
14.	keep down	–	–	–	+
15.	hold back	+	–	–	–
16.	hold back	–	–	–	–

Direct Objects

2 syllables		3 syllables	
1st syll. stressed	2nd syll. stressed	1st syll. stressed	2nd syll. stressed
Arnold	Irene	Oliver	Amanda
Albert	Annette	Anthony	Amelia
Allen	Eileen	Emily	Ophelia
Ellen	Elaine	Agatha	Anita

were monosyllabic; the direct objects were all persons' names beginning with a vowel, and ending with a consonant if two syllables or a vowel if three syllables. In the sets of scrambled words the verb-particles and direct objects were in the same order as in the previous experiment. Interspersed among them were the subject (always a two-syllable person's name), a one-syllable modal, and the word *to.*

The sets of scrambled words comprised all possible combinations of the six binary factors being tested, i.e., $2^6 = 64$ sets. In order to minimize the chances that the subjects would figure out the purpose of the experiment, the subjects were divided into four groups of 40, with each group receiving one fourth of the items. In addition, six dummy items were inserted into each group's set of items to further obfuscate the purpose. Thus each group responded to 16 test items and 6 dummy items. Each of the 16 direct objects and each of the 16 verb-particle combinations occurred only once in each group. Nevertheless, the particular experimental design used[1] enabled 61 of the 64 combinations of factors to be generalized to all subjects, even though each subject responded to only 16 combinations.

As before, the subjects' responses were coded as 0 if the verb and particle were contiguous and as 1 if separated.

Results

The main effects of the six hypothesized factors were computed using analysis of variance.[2] The results are shown in Table 6. Factors that significantly favored contiguity were idiomatic usage, vowel-initial particles, and corresponding nominals (vs. adverbial usage, consonant-initial particles, and no corresponding nominal, respectively). High-frequency combinations tended to be contiguous more often than low-frequency combinations. Neither the stress nor the number of syllables in the direct object affected particle placement and therefore neither was considered further.

The analysis showed that most of the two-way interactions for the non-direct-object factors were significant. Therefore, each significant main factor was tested individually against each other significant main factor in order to determine the simple main effects, shown in Table 6. The significance of the main effect was maintained across the other

[1] The experimental design was partially confounded, factorial, and block (Kirk, 1968). The "block," in this experiment, means that each subject responded to a block of items (16, in this case) rather than to just one item. The "factorial" means that all possible combinations of the hypothesized factors were tested. Six binary factors imply $2^6 = 64$ combinations. The "partially confounded" means that the effects of certain combinations of factors cannot be determined because they are confounded with, or cannot be separated from, the effects of group membership. The specific design used was plan 6.3 of Cochran and Cox (1957, p. 234) (where a = idiomaticity, b = frequency, c = initial phoneme, d = length, e = stress, and f = corresponding nominal). The effects of the three four-way combinations abcd, abef, and cdef were confounded with the group effects. None of these combinations showed a significant effect in an analysis of variance.

[2] The analysis of variance was computed on the Campus Computing Network at UCLA using the BMD05V program. The significance of each of the 64 combinations was tested separately.

TABLE 6

Experiment 2: Significance of Factors

Factors	Main Effects	Simple Main Effects							
		Idiomatic	Adverbial	Initial vowel	Initial consonant	Corresponding nominal	No corresponding nominal	High frequency	Low frequency
1. idiomatic/adverbial	$p<.01$	–	–	.01	.01	.01	.01	.01	.01
2. initial vowel/consonant	$p<.01$	.01	.01	–	–	.01	.01	.01	.05
3. corresponding nominal/none	$p<.01$	.01	ns	.01	ns	–	–	.01	ns
4. high/low frequency	$p<.05$	ns	ns	.01	ns	.01	ns	–	–
5. 3/2 syllables	ns								
6. first/second syllable stress	ns								

Note: ns = not significant ($p > 0.05$)

factors for the idiomatic/adverbial and initial vowel/consonant factors. However, the simple main effects for the corresponding nominal/none factor were significant across only half the other factors; for the high/low frequency factor only one third of the simple main effects were significant.

The above analysis of variance used subjects as the random factor. That is, even though the subjects had significant individual preferences with respect to particle placement (for all of the four groups, $F(39, 562) > 3.75$, $p < 0.01$), the significant factors should be significant for another set of subjects selected at random, and therefore for the population as a whole – but only for the particular set of verb-particle combinations tested. In order to determine how widely the factors affect the lexicon, it is necessary to also estimate how representative the words selected are of all the verb-particle combinations relevant to the factor, using the minimum quasi-F ratio (min F').[3] If

[3]

$$\min F' = \frac{\text{variance of factor}}{\text{variance of interaction of factor with subjects + variance of words}}$$

$$F = \frac{\text{variance of factor}}{\text{variance of interaction of factor with subjects}}$$

The minimum quasi-F ratio, and the whole question of making general claims on the basis of a limited sample, is discussed in Forster and Dickinson (1976) and Clark (1973).

TABLE 7

Experiment 2: Generalizability of Factor Effects

Factors	Groups 1	2	3	4
Idiomatic/ adverbial	min F' (1,13)=6.21 $p<.05$	min F' (1,4)=2.95 $p<.25$	min F' (1,11)=5.56 $p<.05$	min F' (1,15)=3.88 $p<.10$
Initial vowel/ consonant	min F' (1,8)=3.39 $p<.25$	min F' (1,9)=5.19 $p<.05$	min F' (1,13)=4.67 $p<.05$	min F' (1,12)=4.8 $p<.05$
Corresponding nominal/none	min F' (1,15)=2.8 $p<.25$	min F' (1,9)=1.16 $p>.25$	min F' (1,12)=1.3 $p>.25$	min F' (1,11)=3.07 $p>.25$

min F' is significant for a factor, then the effect of that factor generalizes beyond the particular verb-particles and individuals used to all verb-particles relevant to the factor for all users of the language. [Note that the membership of the set of "all verb-particles relevant to the factor" is unclear. Adverbial usage, for example, might include any verb-particle used adverbially, or might be restricted to that subset of verb-particles that participate in the specific binary contrast of idiomatic vs. adverbial. In the former case, a significant min F' would indicate that the adverbial factor generalizes to all verb-particles used adverbially. In the latter, a significant min F' would indicate that the adverbial factor generalizes to all other verb-particle combinations participating in the binary contrast.]

Min F' was computed for one-fourth of the subjects at a time, because of computational limitations. (There were no significant differences among the four sets of subjects used.) The results are shown in Table 7. Both the initial vowel/consonant factor and the idiomatic/adverbial factor should have a significant effect if a different set of verb-particle combinations were used; the corresponding nominal/none factor is significant only for the particular set of verb-particles used in the experiment. The interaction effects also did not generalize beyond this experiment. Note that the initial vowel/consonant and idiomatic/adverbial factors should generalize even though particle placement was also significantly affected by the specific verb-particle combinations used (for all of the four groups $F(8, 562) > 3.75, p < 0.01$).

The above analyses were based on the linguistic behavior of the group as a whole. In order to test whether all (or most) of the individuals showed the same linguistic behavior, a type of implicational scaling (Guttman scaling[4]) was used. In implicational scaling, a hierarchy (scale) of expected behavior is tested to see whether items higher on the scale show a particular behavior if and only if items lower on the scale show the same behavior. If such is the case, the items are said to be implicationally scaled. And if a pair of items is implicationally scaled, then it is true that most of the subjects are behaving the same way – have the same grammar – for that set of items. Note this

[4] The SPSS program "Guttman Scale" on the Campus Computing Network at UCLA was used to analyze the data.

is true regardless of the overall preference of individuals for (in this experiment) either contiguity or separation of the verb and particle. Thus, for example, the group as a whole tended to separate adverbial verb-particles more than idiomatic ones. Implicational scaling tests whether each individual in the group maintains this preference, by testing whether individuals separate idiomatic verb-particles only if they also separate the adverbial ones. Implicational scaling would fail if, for example, there were two subgroups, a larger one consisting of individuals who separated adverbial verb-particles more often than the corresponding idiomatic verb-particles, and a smaller one consisting of individuals who separated idiomatic verb-particles more often than the corresponding adverbial verb-particles.

Three binary implicational scales were tested, one for each of the three most significant factors (idiomatic/adverbial, initial vowel/consonant, corresponding nominal/none). Since each factor was represented by eight pairs of verb-particles, individual behavior was tested pair by pair. The results are shown in Table 8. For the idiomatic/adverbial and initial vowel/consonant factors, most of the subjects behaved predictably on six of the eight pairs. For the corresponding nominal/none factor, subjects behaved as predicted on half of the pairs.

Discussion

Particle placement was shown to be affected by the phonology of the particle and by the lexical/semantic cohesion of the verb-particle combination. While there were individual differences based on both the particular verb-particle combination and the particular subject, nevertheless the influence of the linguistic factors was shown to generalize (statistically) to other verb-particle combinations and to other speakers. Moreover, most of the subjects appeared to have the same rules in their grammars.

The only phonological factor that affected particle placement was the initial segment of the particle (vowel vs. consonant). The number of syllables and stress pattern of the direct object failed to affect particle placement. Apparently the difference between two and three syllables is not enough to stimulate separation, and any stress-pattern or rhythmic effects are not elicited in a written experiment of this sort. But the fact that vowel-initial particles tended to be contiguous confirms the hypothesis that syllable-structure constraints affect particle placement, so that verbs and particles without clear (phonotactically defined) syllable boundaries between them are more cohesive.

Phonologically defined cohesion interacted interestingly with the semantic and lexical cohesion factors. The only robust lexical/semantic cohesion effect was that of the idiomatic/adverbial contrast. That is, the existence of a corresponding nominal facilitated contiguity of the verb and particle only for certain verb-particle combinations, i.e., this result was not general. And frequency of the verb-particle combination, by itself, did not affect particle placement. However, frequency did affect particle placement for verb-particle combinations with corresponding nominals or with vowel-initial particles – both situations that independently favor contiguity. A similar situation obtains for the corresponding-nominal factor – it appeared to have an effect in only those situations that favored contiguity (idiomatic, vowel-initial, or high frequency). These interactions suggest that there may be different underlying forms for combinations that tend to be contiguous

TABLE 8

Experiment 2: Guttman Scaling

%Subjects behaving as predicted

100%

			reversed 95%	85%	ns--> 70%	weak sig[b]--> 80%	strong sig[a]--> 90%
		13/14					x
		3/4					x
		7/8					x
	idiomatic/	9/10					x
	adverbial	11/12					x
	factor	5/6				x	
		1/2				x	
		15/16			x		
		9/13					x
		10/14					x
Pairs	initial	2/6					x
of	vowel/	3/7					x
items	consonant	1/5					x
	factor	4/8				x	
		11/15			x		
		12/16	x				
		5/13					x
		2/10					x
	correspon-	6/14					x
	ding nomi-	3/11					x
	nal/none	7/15				x	
	factor	1/9			x		
		4/12		x			
		8/16	x				

[a]Coefficient of reproducibility > 0.9 and coefficient of scalability > 0.7.
[b]Only 1 coefficient above threshold.

and those that tend to be separated, with the contiguous combinations being a single lexical item and the separated combinations being two lexical items. Lexical factors such as frequency of occurrence and corresponding nominal, which are based primarily on the combined verb and particle, appear to be relevant only for the single lexical items, in which the verb and particle tend to be contiguous.

To summarize the results of these experiments, then, the placement of a particle is not completely optional but is affected by semantic and phonological factors. Moreover, while individuals appear to be affected similarly by these factors, they also have their own general preferences for particles and verbs to be contiguous or not. And each verb-particle combination has its associated tendency towards contiguity or separation, in addition to the general effects of the linguistic environment. All of this regular variation is part of the linguistic behavior of a community of language users, and hence must be included in a description of the language. A statement describing particle placement must minimally contain some references to all of the factors affecting particle placement. It would be useful if the statement also related the factors to each other, both by determining the relative importance of the factors and by determining how they interact, if at all. In the next section, a model for linguistic description is proposed that fulfills the above requirements.

MODELING THE DATA

The proposed model is a generalized linear model that 1) is additive and 2) permits any interactions among factors to be modeled. (It is, in effect, a regression equation using binary variables in such a way as to equate regression and analysis of variance.[5]) As in analysis of variance, it describes linguistic behavior in terms of averages and deviations from average caused by the relevant factors.

Thus, in the second experiment, the verb and particle were separated 74% of the time. Since a separated response was indicated by 1 and a contiguous response by 0, the overall average response was 0.740. The average response for adverbial verb-particles (0.83) was higher than the overall average of 0.74, while the average response for idiomatic verb-particles (0.65) was lower than the overall average. If the only factor that affected particle placement were the adverbial/idiomatic factor, then the following equation would model the data:

$$\text{expected value} = 0.74 \pm 0.09$$

or, in symbols,

$$Y = M + b_A A$$

[5] The variables in the equation have been coded in "effect coding." Effect coding differs from the more familiar "dummy coding" (which uses only 1's and 0's) in that it uses -1 as well as $+1$ and 0 to code the variables; the consequence is that the *b*-weights in the regression equation are deviations from the mean, and hence represent treatment effects in an analysis of variance. Thus, a major advantage of effect coding is that it functionally equates analysis of variance and regression (Kerlinger and Pedhazur, 1973). Another advantage is that the factors are independent of each other in the equation. (The variables within the subject factor are correlated with each other, that is, S_1 is not independent of S_{10}. But every variable of the subject factor is independent of every variable of non-subject factors. For example, S_1 is independent of A, C, and the eight V-P combination variables.)

where

Y = the expected value for particle placement
M = 0.74 (= overall average)
b_A = 0.09 (= weight of adverbial/idiomatic factor)
A = +1 for adverbial
−1 for idiomatic.

In fact, of course, the factor of idiomatic/adverbial was not the only factor affecting particle placement: The other significant factors were initial vowel/consonant, the particular verb-particle combination used, and subject preference. Therefore these factors must also be included in the equation. However, the analysis of variance that determined the significance of these factors did not indicate the relative importance of the factors. Therefore, the statistical technique of regression[6] was used to order the factors in terms of their relative importance of prediction of particle placement. The ordered equation, a kind of variable rule, is shown in Table 9, with the associated *b*-weights. The initial phoneme of the particle is the most important factor, followed closely by the adverbial/ idiomatic factor. The word and subject variables, when entered individually, are the least important in predicting particle placement (see the Appendix).

The above model is relatively theory-neutral, in that it simply predicts particle placement based on the conditioning factors, and, as such, can be used within any syntactic framework. The question remains as to whether the data described in this paper can in fact help us to select among various syntactic frameworks. To help answer this question, consider the various approaches to the verb-particle construction mentioned in the introduction:

(1) *Transformational,* in which the verb and particle are separate constituents, with either the contiguous form or the separated form being basic;

(2) *Phrase structure* (non-transformational), in which the verb and particle remain separate, participating in different syntactic constituents depending on the placement of the particle, but with particle placement determined by rules of approximately equal cost (for example, meta-rules or rules using the slash feature);

(3) *Discontinuous constituents,* in which the verb and particle form a single constituent that can be discontinuous;

(4) *Analyzability,* in which positional tendencies are correlated with the analyzability of the verb-particle combination.

The data in this paper are completely consistent with either the discontinuous approach (3) or the analyzability approach (4); in the alternative model to be proposed shortly, they may be consistent with the phrase-structure approach (2). They are not consistent with the transformational approach (1).

Let us first consider the lack of support in the data for the transformational approach (1).

6 The SPSS program "Regression" on the UCLA Campus Computing Network was used to provide the *b*-weights and the ordering via forward stepwise regression.

TABLE 9

Combination of factors for determining particle placement

$$\text{placement } Y_{ij} = M + b_C C + b_A A + b_{W_i} + b_{S_j}$$

where placement = 1 ——> separation

= 0 ——> contiguity

M = mean (= 0.74)

b_C = initial consonant/vowel weight (= 0.089)

b_A = adverbial/idiomatic weight (= 0.085)

b_{W_i} = verb-particle combination tendency ($-0.117 \leq b_{W_i} \leq 0.183$)

b_{S_j} = subject preference ($-0.60 \leq b_{S_j} \leq 0.25$)

and

C = +1 ——> consonant initial

= –1 ——> vowel initial

A = +1 ——> adverbial

= –1 ——> idiomatic

The type of data analysis performed in this paper might be considered to provide support for a transformational approach if there were a strong preference for one placement of the particle over the other. In such a case, this placement would be considered to be the basic one, with the other one being derived. Such a preference is not observed in the data. The average percentage of separation of the verb-particles differed categorically between the two experiments, with the overall percentage in Experiment 2 being about 75%, and in Experiment 1 being less than 50%. Thus, on the basis of these experiments, there is no motivation for saying that one or the other form is basic.

In considering the implications of the data for the other three approaches, it is worth considering another model for the data. This alternative model uses the conditioning factors to indicate the degree of lexicalization of the verb-particle combination, as can be seen in Table 10; the degree of lexicalization then predicts the particle placement (in conjunction with individual differences associated with each lexical item, and, of course, the individual speaker). The two models differ in their directness of prediction. That is, the variable-rule type model in Table 9 uses the conditioning factors directly to predict particle placement, while the lexicalization model in Table 10 uses the conditioning factors indirectly, first determining lexicalization and then predicting particle placement.

TABLE 10

Derivation of lexical status of verb-particle combinations

	idiom	vowel	nominal	hifreq	Observed % of Separation	Predictions of Lexicalization	
						Statistical	Binary
pickup	1	1	1	1	.45	4	1
workout	1	1	1	0	.46	3	1
showup	1	1	0	1	.49	2	1
pickup_a	0	1	1	1	.52	2	1
putdown	1	0	1	1	.70	1	1
holdback_a	0	0	0	0	.75	0	0
holdback	1	0	0	0	.76	1	0
putout	1	1	0	0	.76	2	0
setback	1	0	1	0	.77	1	0
putdown_a	0	0	1	1	.86	0	0
putout_a	0	1	0	0	.86	1	0
workout_a	0	1	1	0	.89	1	0
showup_a	0	1	0	1	.90	1	0
setback_a	0	0	1	0	.93	0	0
keepdown	1	0	0	1	.94	1	0
keepdown_a	0	0	0	1	.98	0	0
Correlation between observed and predicted						$R = 0.82$	$R = 0.88$

The lexicalization model originated in the observations about the limited effects of the lexical factors of corresponding nominal and frequency, noted in the *Discussion* of EXPERIMENT 2 (p. 325). The fact that these lexical factors were significant only when the verb and particle tended to be contiguous suggested a possible correspondence between lexicalization and contiguous placement. Thus, the lexicalization model tests whether some prediction of lexicalization could be derived that would in turn predict the placement of the particles.

As seen in Table 10, two predictions of lexicalization were tested. The one on the left, labeled "statistical," was determined using the first two interaction terms in the regression equation,[7] which were three-way interactions between (1) initial consonant/vowel, corresponding nominal/none, and high/low frequency, and (2) idiomatic/adverbial, initial consonant/vowel, and corresponding nominal/none. In each case, the value associated with contiguity was assigned a 1, and the other value a 0. Thus, initial vowel, idiomatic, corresponding nominal, and high frequency were all assigned the value of 1; the interaction terms simply multiplied the 1's and 0's, resulting in an interaction value of 1 only if all the factors were 1. The prediction of lexicalization resulted from adding four terms: the two main effects of initial consonant/vowel and idiomatic/adverbial, and the two interaction terms. The resulting five degrees of lexicalization correlated quite well with the observed percentages of separation ($R = 0.82$).

While this prediction of lexicalization worked quite well, and has the advantage of being statistically based, it is somewhat cumbersome to use. Therefore, a simpler method was tested, one which simply adds the values associated with each conditioning factor, and determines a binary measure of lexicalization. In this method, shown on the right of Table 10 under the column labeled "binary," a verb-particle was assumed to be lexicalized if three or more of its conditioning factors had a value of 1. This binary determination of lexicalization also correlated well with the observed percentages of separation ($R = 0.88$). Thus, a verb-particle combination is assumed to be a single lexical item if three out of the four conditioning factors are true: vowel-initial particle, idiomatic usage, corresponding nominal, and high frequency.

In its strictly binary form, the lexicalization model retains an aspect of the phrase-structure approach (2), in which there is a structural difference between contiguous and non-contiguous verb-particles. It is also compatible with the analyzability approach (4), in which the separation between the verb and particle is associated with the analyzability of the verb-particle combination. Note, however, that it assigns the analyzability difference to the lexical component rather than to the syntactic component. Thus, the lexicalization model requires some form of the discontinuous constituent approach (3), since the verb and particle in the single lexical items may be separated, resulting in a discontinuous constituent.

The binary lexicalization model suggests that there is a categorical difference between cohesive and non-cohesive verb-particle combinations. To the extent that this captures the facts, it is an advantage of the model. Thus, a compositional analysis of *put down,*

[7] The interactions were determined using forward stepwise regression (program 2R in the BMDP statistical package). The two main effects of initial consonant/vowel and idiomatic/adverbial were forcibly included in level 1; all the interactions of the four conditioning factors (initial consonant/vowel, idiomatic/adverbial, corresponding nominal/none, high/low frequency), but not the main effects of the other two main effects, were optionally included in level 2. The first two interaction terms in the stepwise regression were selected based on F-to-enter, change in R-Square, and their usefulness in the lexicalization model. On statistical grounds, only the first interaction, the three-way interaction among initial consonant/vowel, corresponding nominal/none, and high/low frequency, would have been included. However, in order to capture the interactions with idiomaticity as well, the next term, the three-way interaction among idiomatic/adverbial, initial consonant/vowel, and corresponding nominal/none, was also included. The regression was performed on the percentage of separated verb-particle combinations.

meaning 'to scorn,' is much more difficult that one of *put down* meaning 'to move to a low physical position.' But this difference may be more one of degree than of categoricalness. Certainly, the fairly wide distribution of percentage of separation within the two lexical categories in Table 10, especially the 0 category, suggests that a strict categorical model is missing something. In addition, the binary lexicalization model results in a more complicated description of verb-particle placement than models that say all verb-particles have the same lexical description. In particular, the formalism will no longer reflect the fact that verb-particles have something in common in their placement behavior, regardless of their overall preference for contiguity or separation.

With all these factors taken into consideration, then, the syntactic framework that provides the simplest statement of verb-particle placement appears to be the discontinuous constituent approach (3), combined with the type of variable rule in Table 9. The predictions of lexicalization in Table 10 are probably best seen as an additional characterization of the data, one that reflects the degree of cohesiveness between the verb and particle, and one that perhaps indexes the diachronic move towards lexicalization of the verb-particle combinations.

REFERENCES

ABSALOM, D.J. (1973). An investigation of verb-particle combinations in English. *Kivung*, **6**, 48-60.

BOLINGER, D. (1971). *The Phrasal Verb in English.* Cambridge, MA: Harvard University Press.

BUTTERS, R.R. (1973). Acceptability judgments for double modals in southern dialects. In C-J.N. Bailey and R. Shuy (eds.), *New Ways of Analyzing Variation in English* (pp. 276-286). Washington, D.C.: Georgetown University Press.

CEDERGREN, H.J. and SANKOFF, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, **50**, 333-355.

CHOMSKY, N. (1957). *Syntactic Structures* (Janua Linguarum, **4**). The Hague: Mouton.

CLARK, H.H. (1973). The language-as-fixed effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335-359.

COCHRAN, W.G. and COX, G.M. (1957). *Experimental Designs.* New York: John Wiley and Sons.

EMONDS, J. (1972). Evidence that indirect object movement is a structure-preserving rule. *Foundations of Language*, **8**, 546-561.

FILLMORE, C.J. (1965). *Indirect Object Constructions in English and the Ordering of Transformations.* The Hague: Mouton.

FISCHER, S. (1976). Child language as a predictor of language change, a case study. *Working Papers in Linguistics* (University of Hawaii), **8**, 71-104.

FORSTER, K.I. and DICKINSON, R.G. (1976). More on the language-as-fixed effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F^1. *Journal of Verbal Learning and Verbal Behavior*, **15**, 135-142.

FRASER, B. (1976). *The Verb-Particle Combination in English.* New York: Academic Press.

GAZDAR, G., KLEIN, E., PULLUM, G. and SAG, I. (1985). *Generalized Phrase Structure Grammar.* Cambridge, MA: Harvard University Press.

HUCK, G.J. (1985). Discontinuity and word order in categorial grammar. *Indiana University Linguistics Club.*

JACOBSON, P. (to appear). Phrase structure, grammatical relations and discontinuous constituents. In G. Huck and A. Ojeda (eds.), *Discontinuous Constituency (Syntax and Semantics,* **20***).* New York: Academic Press.

KAHN, D. (1976). Syllable-based generalizations in English phonology. *Indiana University Linguistics Club.*

KENNEDY, A.G. (1920). The modern English verb-adverb combination. *Stanford University Publications in Language and Literature,* **1.**

KERLINGER, F.N. and PEDHAZUR, E.J. (1973). *Multiple Regression in Behavioral Research.* New York: Holt, Rinehart, and Winston.

KIRK, R.E. (1968). *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, CA: Brooks/Cole Publishing Co.

KRUISINGA, E. and ERADES, P.A. (1953). *An English Grammar* (first published in 1911). **1.** Groningen: P. Noordhoff.

KUČERA, H. and FRANCIS, W.N. (1967). *Computational Analysis of Present-Day American English.* Providence, RI: Brown University Press.

LABOV, W. (1969). Contraction, deletion and inherent variability of the English copula. *Language,* **45,** 715-762.

LINDNER, S.J. (1983). A lexico-semantic analysis of English verb particle constructions with *out* and *up. Indiana University Linguistics Club.*

MAKKAI, A. (1972). *Idiom Structure in English.* The Hague: Mouton.

ROUSSEAU, P. and SANKOFF, D. (1978). Advances in Variable Rule Methodology. In D. Sankoff (ed.), *Linguistic Variation* (pp. 57-69). New York: Academic Press.

VAN DONGEN, W.A., Sr. (1919). 'He put on his hat' and 'he put his hat on.' *Neophilologus: A Modern Language Quarterly,* **4,** 322-353.

APPENDIX

Determining *b*-Weights

In order to determine the *b*-weights for each individual subject and verb-particle combination, there were separate vectors for each in the regression equation (see below). This was done in order to model the behavior of the individual subjects and verb-particle combinations as exactly as possible. However, it means that the variables with the largest *b*-weight are not necessarily the most significant in explaining the variation. That is, the *b*-weights for the subjects, in particular, are much larger than those for the two linguistic factors. But a variable with a large *b*-weight will have a relatively small impact on the prediction of particle placement over a whole collection of cases if that variable occurs with only a few of the cases. Since each case was either idiomatic or adverbial, and had either a consonant- or vowel-initial particle, these two factors affect particle placement the most even with relatively small *b*-weights. But any one subject only provided 1/160 of the cases, so even a large *b*-weight for that subject is relatively unimportant in predicting the overall collection of cases. The fact that the subject variables are correlated with each other also affects the relationship between the subject *b*-weights and their relative contribution to the prediction of particle placement. Thus, out of 67 subject variables that differ significantly from the mean, only 4 are ordered before the two linguistic factors (and variable W_5). Of the other significant subject variables, 22 were ordered prior to W_2, and 29 before variables W_8 and W_4. (The other four verb-particle combination *b*-weights did not reach significance in the regression.)

The exact coding (effect coding[5]) for the verb-particle combinations follows. Two V-P combinations were used for each of eight binary variables, W_1 through W_8, corresponding to the eight linguistic conditions ($8 = 2^3$: 3 binary factors, i.e., adverbial/idiomatic, consonant/vowel, corresponding nominal/none). For any condition i, ($1 \leq i \leq 8$), variable $W_i = +1$ for the first verb-particle combination nested in the condition, $W_i = -1$ for the second V-P combination, b_{W_i} = weight associated with these combinations, and variables $W_j = 0$ (for all $j \neq i$, $1 \leq j \leq 8$).

The individual subject preferences were included in the equation as follows. To facilitate computation the subjects were split into 4 groups of 40 subjects each. For each group, 39 subject variables are necessary to express the effect of the subject factor. For any subject i ($1 \leq i \leq 39$), variable $S_i = +1$, b_{S_i} = weight indicating subject i's particle placement preference, and variable $S_j = 0$ (for all $j \neq i$, $1 \leq j \leq 39$). For subject 40, all 39 variables are set to -1;

$$b_{S_{40}} = -b_{S_1} - \ldots - b_{S_{39}}$$

Note that, as with other variables, any single subject *b*-weight is added once and subtracted once; that is, the subject weights are also defined in terms of how they differ from the mean. The resulting equation is a regression equation,

$$Y = M + b_A A + b_C C + b_{W_1} W_1 + \ldots + b_{W_8} W_8 + b_{S_1} S_1 + \ldots + b_{S_{39}} S_{39}$$

The equation can be simplified by considering only one verb-particle combination and one subject:

$$Y_{ij} = M + b_A A + b_C C + b_{W_i} W_i + b_{S_j} S_j$$