JOURNAL OF MEMORY AND LANGUAGE 28, 519–530 (1989)

Levels of Affixation in the Acquisition of English Morphology

PETER GORDON

University of Pittsburgh

The theory of level-ordering proposes that word formation processes are assigned to one of three levels within the lexicon. Level 1 processes are applied before Level 2, which are applied before Level 3. Such ordering constraints predict differences in acceptability between pairs such as *teeth marks* versus **claws marks*, and *Darwinianism* versus **Darwinismian*. An acquisition model is examined in which Level 1 forms are separately lexicalized, and Level 2 and 3 forms are more productive. Three untimed lexical-decision experiments were carried out with 5- through 9-year-olds and found general support for a systematic relation between productivity and level assignment. However, a number of serious problems with the model are pointed out with regard to both linguisitic and empirical data. A revision of the model is posposed which appears to deal effectively with these problems.

Recent studies of the acquisition of morphology have revealed that children have productive control of a number of word-formation devices from very early in their language learning careers. For example, Clark (1981, 1982) has documented that even 2-year-olds appear to use processes of derivational word-formation quite productively as evidenced by their overgeneralization errors. Similarly, Slobin (1985) notes that Turkish 2-year-olds demonstrate almost errorless learning of a complex set of agglutinative affixes in their language. In considering word formation in language development, there appear to be two central issues which can broadly be characterized as questions relating to (i) productivity, and (ii) constraints.

"Productivity," in the present sense, refers to the extent to which people actively create words, for example, by combining stems and affixes in much the same way that they generate sentences. Unlike sentence generation though, it is reasonable to argue that a large proportion of complex words that are comprehended or produced are simply frozen lexical forms with little or no decomposition within the lexicon. For example, it seems highly unlikely that *organization* is generated from its root, *organ* during on-line processing—although it may be linked in some way within the lexicon. On the other hand, the novelty of a word like *xeroxable* would suggest that the form is generated during language production. The notion of "constraints" deals with the fact that given that there is some productivity, how is that productivity restricted. For example, why do we find *Darwinianism* acceptable, but not **Darwinismian*.

With regard to productivity, Aronoff (1976) has suggested that the key to deciding whether a complex word is productively formed is in noting whether the meaning is predictable as a joint function of the meanings of the stem and affix. Thus, since the idiosyncratic meaning of *organization* does not meet this condition, then there would have to be a separate lexical entry for the complex form in order that the appropriate meaning be associated with it in the lexicon. In addition to semantic compositionality, there are also phonological properties that might determine analyticity, in turn affecting productivity. Processes such as stress shift and vowel change can have con-

This study was supported by N.S.F. Grant BNS-8811810 to the author. The author wishes to thank Jack Protetch and Sharon Serdula for running subjects. Requests for reprints should be sent to Peter Gordon, Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260.

0749-596X/89 $3.00

siderable consequences in making the relation between the stem and derived form opaque (e.g., *beast* → *bestial*). While such phonological changes could be characterized as rule-based (e.g., Chomsky & Halle, 1968), it is not always clear that such rules capture truly psychological processes of word formation rather than just linguistic generalizations based on diachronic rather than synchronic processes.

In general, affixes appear to differ in terms of whether they trigger phonological changes in the stem to which they attach. Those that cause such changes are usually referred to as "nonneutral" and those that do not are called "neutral." Actually, neutral affixes may sometimes cause changes at the juncture of the stem and affix such as in velar softening (*/k/*→*/s/* e.g., *critic* → *criticism*). However, such affixes do not alter internal aspects of the stem phonology such as vowel quality or relative stress assignment. For example, *Darwinism* has the same stem stress pattern as *Darwin*, whereas the nonneutral affix *-ian* shifts the stress to the second syllable as in *Darwinian*.

In a recent approach to lexical theory known as "level-ordering" (Allen, 1978; Siegel, 1977; Kiparsky, 1982, 1983), these differences between affix types have been extended to other word-formation processes such as compound formation, irregular inflections, and zero derivation. The theory proposes that word-formation rules are assigned to levels within the lexicon, and that there is an ordering of rules between levels. That is, Level 1 processes apply first, followed by Level 2 and so on.

Within this proposal, Level 1 rules include nonneutral derivational affixes (e.g., *-ion, -ian, -ity, -ous*), irregular inflections (e.g., *mice, geese, stood*), and certain zero derivations that cause stress shift (e.g., *to protest* → *a protest*). Level 1 processes thus have access to the internal phonology of the stems to which they attach, and are often not semantically predictable.

Level 2 includes neutral derivational affixes (e.g., *-er, -ism, -ness, -able*), compounding, and zero derivations where there are no phonological alterations (e.g., *a pattern* → *to pattern*). Level 2 processes thus have no access to the internal phonology of the stem, and tend to be semantically transparent (although semantic idiosyncracies can arise). Correlated with these transparency properties, Level 2 processes also tend to be fairly unrestrictive in that there are fewer idiosyncratic exceptions to the application of the rule. For example, *-able* will attach to almost any main verb. This is in contrast to Level 1 processes that tend to be more restrictive in their application to particular stems. Finally, Level 3 forms include regular inflections (e.g., *-s, -ed, -ing*) which are semantically compositional, phonologically neutral, and completely unrestrictive unless blocked by an irregular form (e.g., **mouses*).

The purpose of level assignment is made clearer when one considers the constraints imposed by ordering of rules within the lexicon. For example, since *-ian* (Level 1) must apply before *-ism* (Level 2), it is predicted that *Darwinianism* is acceptable but not **Darwinismian*. Similarly, an irregular plural at Level 1 may occur inside a compound at Level 2 (e.g., *teeth-marks*). However, a regular plural at Level 3 may not (cf. **claws-marks*). Kiparsky (1982) has noted a large number of predictions of this type that are generated by level-ordering theory and appear to meet with a good degree of success. The theory thus appears to provide an elegant explanation for many constraints on word formation that might otherwise seem quite inexplicable.

The question to be addressed in the present paper concerns how children end up with the kinds of intuitions that appear to fall out from level-ordering theory. In a previous study, Gordon (1985) addressed the issue of whether children obey level-ordering constraints by examining the interaction of pluralization and compounding noted above. When children were encouraged to produce compounds of the form *X-eater,* they almost never used a regular plural inside a compound (e.g., **rats-eater*), but did so quite willingly for irregular plurals (e.g., *mice-eater*). It was also noted that since children almost never hear irregular plurals inside compounds, their willingness to include them probably arises from the systematic application of principles that are captured by level-ordering theory.

Gordon (1985), Tyler (1986), and Walsh (1984), have each proposed that ordering effects could be explained to some extent by the degree of productivity of the processes at different levels. Level 1 rules are those that tend to make the relation between the stem and complex form either semantically or phonologically opaque, thus making it harder for the child to form the appropriate generalizations. In such a case, he or she might simply acquire the complex form as a separate lexical entry rather than being related by rule to the stem. In the case of Level 2 and 3 forms, the child should be better able to discover the relationship between the two forms and eventually induce a productive rule.

The ordering of rules can be made to fall out of these differences in productivity quite naturally. If Level 1 derivations are acquired as separate lexical items then, like base forms, they are open to further Level 2 and 3 processes. On the other hand, once a Level 2 or 3 rule has applied, the Level 1 affix could not attach outside of these forms, because the affixation process is not specified as an independent rule. The ordering of Level 3 inflections as a final process could follow from a number of principles. Primarily, these would be based on the fact that inflection reflects the interface between morphology and syntax (e.g., inflections are required for agreement between sentence constituents—see Anderson, 1982). If we assume that syntactic processes are in some sense applied after lexical processes,[1] then this entails the required late application of inflections after Level 1 and 2 derivations. Notice however, that irregular inflections such as *mice* cannot be rule generated and must be listed in the lexicon.

Evidence for the existence of productive rules in children standardly comes from inflectional overregularizations (e.g., *foots, goed*) and novel uses of rules such as in Berko's (1958) *Wug* study. Spontaneous novel derivational forms have been noted by Clark (1981, 1982), Clark & Clark (1979), and Clark & Hecht (1982). These include forms such as *tell-wind* [= weather vane], *lessoner, toothachy,* and *flyable*. In Clark's extensive searches for children's lexical innovations, nearly all of the derivational examples involved Level 2 processes. She provides no examples of Level 1 affixes used productively and comments on their absence (Clark, 1981, p. 319).[2] Such differences in innovative usage clearly provide prima facie support for the model.

The specific predictions to be tested in the present study concern the relationship between a complex form and its stem. This relationship should be most significant in the case of the highly productive rules (Level 3, and to some extent, Level 2), and least significant for lexicalized, Level 1

[1] This assumption does not appear to be very controversial. If one assumes that there is a syntactic component, it is hard to imagine how it could operate before there are any words inserted into the constituents.

[2] An exception here is the case of zero-derivation. Clark (1981) records four cases of deverbal nouns which are postulated to be Level 1, since they can trigger stress-shift, are quite restrictive (Kiparsky, 1982). By comparison, from the same corpus, Clark (1982) records 224 cases of novel denominal verbs, which are formed at Level 2. Both of the examples of deverbal nouns cited by Clark (*a rub* [= eraser], and *a chop* [= axe]) are monosyllabic and hence not subject to stress shift. Thus, even though novel Level 1 zero derivations may occur, they are very infrequent and probably do not involve phonological changes characteristic of Level 1 rules.

forms. In order to examine the relatedness of stem and complex forms, an untimed lexical decision task was developed, with error rate being the relevant dependent variable. Children were presented with a (spoken) word or nonword string and asked to decide if what they heard was a word or not. In such a situation, if the child hears a Level 1 affixed form (e.g., *visitation*), affirmation or denial should be a function of whether the child has heard that individual complex word before and created a lexical entry for it. The response should be relatively independent of whether the child knows the stem form (*visit*). At the opposite extreme, Level 3 inflected forms (e.g., *imagining*) should show responses that reflect the child's knowledge of the stem form (*imagine*). Responses to Level 2 derivatives will depend on how productive Level 2 affixation really is. In any case, they should be more productive than Level 1 forms and probably somewhat less so than for Level 3 forms.

Experiment 1

In order to examine the influence of the stem on responses to complex forms, it is necessary to find a characteristic of the stem that could influence recognition rate. One candidate is a frequency effect. That is, children should be more likely to recognize high frequency words than low frequency words. Consider a complex word that has a very low item frequency, but the frequency of the stem is very high. If the stem influences responding to the complex form, then acceptance rates should be elevated with respect to the expected rate for the low frequency complex form. Such elevation in responding should therefore be greatest for Level 3 forms, followed by Level 2, and then Level 1. While a frequency effect would seem to be intuitively plausible, child lexical decision tasks have hardly ever been used, and no frequency effect has been established in the literature. The first experiment therefore sought to establish a frequency effect as a preliminary to examining the effects of stem frequencies on responses to complex forms in the second experiment.

Method

Subjects. 36 children consisting of twelve 5-, 7-, and 9-year-olds were included. All were native speakers of English and attended a university research school.

Materials and procedure. The words used in the lexical decision task were simple (nonaffixed) words chosen from the Francis and Kucera (1982) word count. The aim of the study was to establish a frequency effect and two specific ranges of frequencies reflecting the upper and lower bounds of correct responding. Pilot testing appeared to show that such ranges could be established by grouping words by natural log frequencies. Stimulus materials thus consisted of 54 words with 9 in each of the frequency ranges from ln 0 to ln 5 (see Appendix). Frequencies were calculated in terms of "cluster" frequencies (i.e., the combined frequency of the word, plus those of its inflections—see Experiment 2 for justification). To avoid presenting too many items, children were divided into three groups, each of which received a subset of three items from each of the frequency ranges plus a set of 18 nonword distractors.

Words and distractors were written on 3″ × 5″ cards and placed in a pile on a board. The board resembled a colorful game board with three rectangles outlined. The cards were placed in a lower middle rectangle. Above this were two rectangles, one to the left and one to the right. One had YES and a check written above it, and the other had NO and a cross. Children were told that the experimenter would read what was on the card and the child had to decide if it was a word or not. Words would go in the YES pile and nonwords in the NO pile. Training consisted of three highly recognizable words (*truck, baby, chair*) and three distractors that appeared to be distinctly recognizable as nonwords (*friblib, pleabus,* *zigglipob*). Feedback was given in the training; however, no corrections were necessary due to 100% correct performance by all subjects. Feedback was not given for the main items, and the distractors were chosen to sound more like real words. One distractor, *wult,* was dropped from the analysis since a number of children confused it with *Walt* (Disney).

Results and Discussion

The acceptance rates by age and frequency, shown in Fig. 1, demonstrate a clear frequency effect for all three age groups (min $F'(5,72) = 10.4$, $p < .001$). Baseline rates for correct responding occurred at around ln 0 and ln 1 frequency ranges. Ceiling was around ln 4 and ln 5. Figure 1 also shows children's erroneous acceptance rates for nonwords. Children had significantly higher acceptance rates for real words (min $F'(1,77) = 56.3$, $p < .001$), which was true even at the lowest frequency range (min $F'(1,35) = 13.4$, $p < .001$). Developmentally, there was a significant effect of age (min $F'(2,88) = 7.6$, $p < .001$). Examining Fig. 1, one can see that there were no developmental changes at the higher frequency ranges. Advances between 5 and 7 years appear to be primarily in the middle frequency ranges (ln 2 and ln 3), and between 7 and 9 years in the lower frequency ranges (ln 0 and ln 1). The establishment of a frequency effect with baseline and ceiling levels provides the necessary response characteristics for examining the nature of the representation of affixed forms in the next experiment.

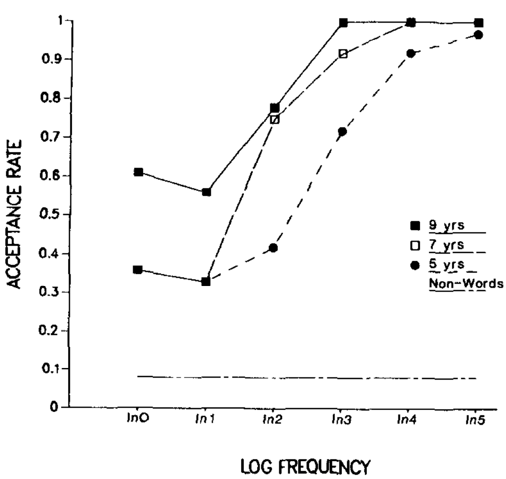

FIG. 1. The frequency function for children's lexical decision.

EXPERIMENT 2

The basic method for examining the relationship between affixed forms and their related stems was to choose affixed words that were very low in individual word frequency, but whose stems had high cluster frequencies. Cluster frequencies include the frequency of the stem and its related inflectional forms. These are included on the assumption that inflected forms are probably not represented separately, or at least that they contribute to the frequency characteristics of the stem. The present methodology borrows heavily from previous adult lexical decision studies in which item versus cluster frequency is manipulated (e.g., Taft, 1979; Bradley, 1979). If the frequency of the stem influences the child's decision, then a high frequency stem cluster should elevate correct responding to the affixed form. From the present model we can predict that affixes from Levels 2 and 3 will have higher rates of correct affirmative responding than Level 1. To the extent that inflection is the most productive form of affixation, one might also predict that Level 3 should have higher affirmative responses than Level 2.

Method

Subjects. Ten 5-year-olds, eleven 7-year-olds, and twelve 9-year-olds who had participated in Experiment 1 were retested in Experiment 2 in the following week.

Materials and procedure. The procedure was identical to Experiment 1, except that there was no training. Children were simply reminded of the basic procedure. Affixed forms from Level 1 (*-ous, -(at)ion, -ity*), Level 2 (*-ness, -er, -able*), and Level 3 (*-s, -ed, -ing*) were chosen from the Francis and

Kucera (1982) word count. The affixed forms had word frequencies of 1 to 2 per million. Their related stem-cluster frequencies were in the ln 4 to ln 5 frequency range. These represent, respectively, the baseline and ceiling levels from Experiment 1. A sample of 36 words was chosen with 12 examples from each level including 4 of each affix type. One level 3 form (*bridges*) was dropped from the analysis because of an error in obtaining the correct frequency count. Subjects were divided into two groups, each receiving half of the items. Of the 18 distractors, nine were nonce stems each with one of the above affixes attached to it (e.g., *tostil-ness*). The other nine were function words with the same set of affixes attached (e.g., *than-ness*). Such distractors control for the possibility that children might simply respond positively to anything with a recognizable affix on it or any real word with an affix on it. All stimuli are listed in the Appendix. Again, one distractor, *more-ous,* was dropped from the analysis since many children confused it with *Morris* (the cat).

Results and Discussion

Results revealed a strong main effect for words versus nonwords (min $F'(1,68) = 263.3, p < .001$), showing a clear preference for children to accept real words and reject distractors. Unlike Experiment 1, there was no main effect for age by either a subject or item analysis. The effect of lexical levels was significant by subjects ($F1(2,29) = 13.9, p < .001$), but marginally failed to reach significance by items ($F2(2,32) = 3.01, p = .06$). In addition, the ordering of acceptance rates for levels was not completely as predicted by the model. Level 2 items were accepted more often than Level 3, although both were accepted more than Level 1 items. These data are shown in Fig. 2.

On reexamining the original set of Level 3 stimuli, it was noticed that many of them

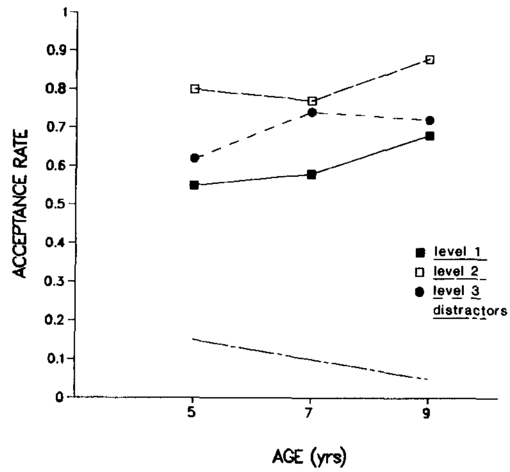

FIG. 2. Affirmation rates for affixed words.

were not, in fact, straightforward inflectional forms. In constructing the materials it was very hard to find inflected words that had the right frequency characteristics. That is, it is almost impossible to find words with very high frequency stem clusters, where each of the inflected forms is not also very high in frequency. One exception is when that inflection attaches to a category that is not normally assigned to the stem. For example, if a high frequency noun changes category to a verb, it may not be used often as a verb, and hence the frequency of the progressive *-ing* form will be low. On inspecting the stimulus materials, it was noticed that a number of items were of this type (*fours, middles, filming, bonding, modeling, fielded, pained, radioed*).[3]

[3] The criteria for deciding on a category change included: (i) The meaning of the base category was entailed in the meaning of the derivation, but not vice versa. For example, the verb, *to model,* means roughly "to provide a model," whereas *a model* does not entail a verb component: (ii) The frequency of the alternate (base) category was substantially greater than that of the inflected form. For example, the frequency of *middle* is roughly double for the adjective compared to the noun. However, this did not apply in the case of *typed,* where the noun is more frequent, but the meaning (*a type* = "a kind") is unrelated to the verb.

The problem here is that the category change would require a zero-derivational process at Level 1 or 2, in addition to the inflectional (Level 3) process.

In order to evaluate the possible contribution of zero derivtion, Level 3 forms were divided into those that involved category change (see above), and those that did not (*incomes, imagining, typed*). The results of this partitioning are shown in Fig. 3. It can be seen in this figure that the zero-derivation forms now resemble Level 1 derivations, and the straightforward inflectional forms now show elevations greater than Level 2 forms, as originally predicted. With this partitioning, there was now a significant effect of levels (min$F'(3,59)$ = 2.71, p = .05)—although some caution is warranted in interpreting data from only three items.

Another problem with the original materials was that they contained some words whose frequency appeared to reflect the biases of the Francis and Kucera (1982) word count, as opposed to actual frequencies in children's language input. For example, *incomes* was often rejected by children, probably because they failed to recognize the stem. Thus, if we consider the group of children that did not receive *incomes* as a Level 3 stimulus, responding was almost perfect for the 5- and 7-year-olds (97%) and was perfect for the 9-year-olds.

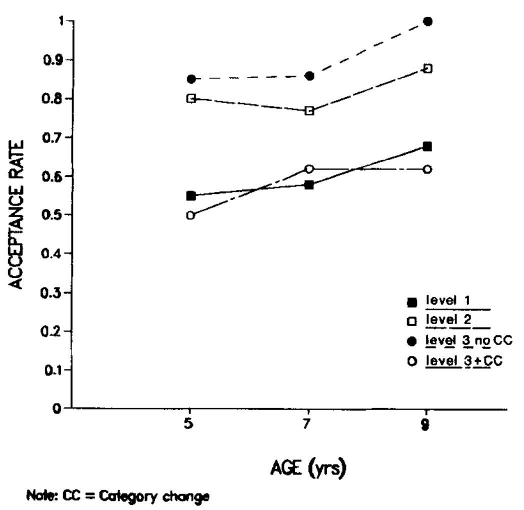

FIG. 3. Affirmation rates factoring zero derivation.

In addition to comparing the relative elevation effects for different level affixes in Experiment 2, one can also compare the results of Experiment 2 with those of Experiment 1. That is, to what extent do responses to the low frequency complex affixed forms in Experiment 2 resemble responses to either the low frequency (ln 0 to ln 1) or high frequency (ln 4 to ln 5) simple forms in Experiment 1? By hypothesis, one should find that productively affixed forms should have responses comparable to the high frequency group, whereas lexicalized complex forms should have acceptance levels comparable to the low frequency group. The relevant data are presented in Table 1. For all ages, Level 3 forms showed responses comparable to the high frequency group indicative of truly productive rule forms. However, Level 1 forms which are expected to be lexicalized, tended to show responses that were quite elevated with respect to the low frequency group. This was

TABLE 1
ACCEPTANCE RATES FOR EXPERIMENTS 1 AND 2

		Experiment 1	
Age/Level	Expt. 2 Affixed	Low freq.	High freq.
5 years		.35	.95
1	.53		
2	.82		
3	.85 (.97)		
7 years		.35	1.00
1	.58		
2	.77		
3	.86 (.97)		
9 years		.58	1.00
1	.68		
2	.88		
3	1.00 (1.00)		

Note. Level 3 does not include zero-deviation forms. Parentheses indicate results for the group not tested on *incomes* (see text). "Low freq." = mean for ln 0 and ln 1. "High freq." = mean for ln 4 and ln 5.

especially true for the younger 5- and 7-year-olds whose responses were elevated by some 20%. Level 2 forms showed much clearer elevation, although they did not suggest the complete productivity of Level 3 forms. In summary, there appears to be some influence of the high frequency stem even in the case of Level 1 forms and certainly in the case of Level 2 forms. Level 3 forms appear to have response characteristics indistinguishable from those of the stem when various confounding factors are eliminated.

While the reanalysis regarding zero derivation was enlightening, this reduced the straightforward inflectional forms down to a very small group. Therefore, a further experiment was carried out in order to test the hypothesis more clearly. In addition, it was possible to eliminate words where the frequencies did not appear to reflect the language that children hear. In addition to eliminating words such as *incomes*, it was also possible to eliminate words like *loser*, which are rare in the word count, but probably less rare in children's language input.

Experiment 3

Method

Subjects. 17 kindergartners (aged 4;10 to 6;2, M = 5;8), 20 second-graders (aged 7;2 to 8;4, M = 7;8), and 20 fourth-graders (aged 9;1 to 10;7, M = 9;7) participated in the experiment. All children were native English speakers from a university research school. None had participated in experiments 1 or 2.

Materials and design. Exhaustive searches of Francis and Kucera (1982) for Level 3 items with high (ln 4 to 5) stem-cluster frequency and low (ln 0 to 1) item frequency revealed that they simply do not exist. Therefore, it was necessary to relax the frequency criteria for the level 3 stem-cluster measure. New Level 3 items included words whose stem-cluster frequencies were between ln 2.4 and ln 4.3 (item frequencies were still within ln 0 to 1). Initially, new items were chosen if the stems seemed to be words that children would probably know. All items were then submitted to a test that required that the stem appear in either Hall, Nagy, and Linn (1984) or Moe, Hopkins, and Rush (1982). These are word counts based on children's speech. Some of the Level 1 and 2 items were also replaced on this basis (see Appendix).

Since children's knowledge of the related stems of complex forms could no longer be predicted on the basis of the frequency effect, the acceptance rates for the stems had to be assessed directly. Thus, one group of children received the affixed form for half the items (n = 18), and the stem form for the other half (n = 18). A second group received the opposite arrangement of stem and affixed forms for those items. To avoid overburdening the children, only 9 distractors (affixed nonce forms) were included. Since the previous experiments revealed a larger number of false negatives than false positives, the reduction in distractors would not cause a great imbalance of *Yes* and *No* responses. All other aspects of training and procedure were identical to Experiment 2.

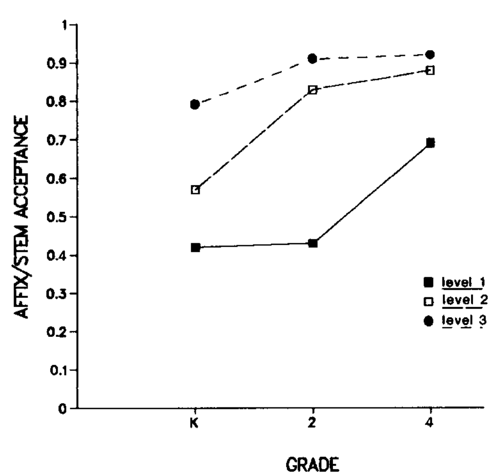

FIG. 4. Proportion of affixed forms accepted in relation to stem acceptances.

Results

The results of Experiment 3 are presented in Fig. 4. The data points in the figure represent averaged acceptance rates for affixed forms, divided by the average acceptance rates for the stems of the same items (collected from children of the same grade but in the other condition.) The curves show the predicted order of acceptance with Level 1 forms considerably less well recognized than Level 2 and 3 forms. Analyses revealed significant effects of lexical level (min $F'(2,48) = 6.57, p = .003$), and grade (min $F'(2,92) = 3.41, p < .05$). While the level *x* grade interaction was significant in a subject analysis ($F1(4,53) = 3.6, p < .01$), the item analysis was not significant ($p = .4$).

General Discussion

The present set of studies has shown a general pattern of support for the lexical model proposed by Gordon (1985), Tyler (1986), and Walsh (1984) as an attempt to explain level-ordering effects in terms of developmental differences in productivity of affix types. Comparisons of neutral and nonneutral affixes have also been investigated by Tyler (1986) where 4th, 6th, and 8th graders were asked whether they knew the meanings of ill-formed words (e.g., **cabbage-ous*, **destroy-ism*.) When the affix was neutral (Level 2) there was a late pattern of overacceptance exhibited by 6th graders. However, there was no such pattern for nonneutral (Level 1) affixes. Such overgeneralization further supports the inference of greater productivity for the neutral affixes.

In adult studies, Bradley (1979) has compared lexical decision times for pairs of complex derivational forms that were matched on item frequency but differed in stem-cluster frequency. Effects of stem influence were found for the neutral affixes, *-ness*, *-er*, and *-ment*, but not for the nonneutral affix, *-ion*. Such results similarly suggest that only neutral affixes are represented with their stems in a compositional relationship. Other adult studies of morphological representation have tended to focus on issues of compositionality and costorage but not specifically differentiated between types of affixes (for a review see Cutler, 1983).

In evaluating the present model, it would be quite easy to accept the data as supporting the basic predictions. However, we must also ask the question of whether the model fully deals with the phenomena that it initially sets out to explain. It is here that we run into problems at both the theoretical and empirical levels. On the theoretical side, the model makes the wrong linguistic predictions in fairly crucial cases. It requires that Level 1 forms only occur within frozen lexicalized items. While this does prevent Level 1 forms from applying outside Level 2 or 3 forms, it also has the consequence of preventing Level 1 affixes from stacking outside other Level 1 affixes. This is clearly violated in words such as *direct-ion-al-ity* which contains three Level 1 affixes.[4] In addition, the model predicts that Level 1 affixes should not be productively applied in novel cases. Yet there are Level 1 affixes that have very few restrictions on their application. For example, *-ian* attaches to almost any proper name. If productivity were all that mattered in determining ordering constraints, then *-ian* should be an exception and should be allowed outside Level 2 forms. This would lead to the prediction that **Darwin-ism-ian* should be acceptable, which it clearly is not.

In addition to these linguistic arguments, we can also find evidence for some Level 1 productivity in the current data. If accep-

[4] It might be the case that *-ality* could be acquired and represented as a single, rather than a double morpheme. However, even if this is the case, we still need to know what forces constrain such morphemic combinations such that they obey ordering constraints in the first place.

tance rates for Level 1 complex forms (Experiment 2) were not influenced by their high frequency stems, then we would predict acceptance rates comparable to the low frequency unaffixed forms in Experiment 1. The relevant comparisons in Table 1 revealed elevations in acceptance rates by as much as 28% for the Level 1 forms compared to the low frequency words in Experiment 1.

Clearly there is more going on here than simple differences in lexicalization and productivity. While the present data show the right direction of effects, there are too many troublesome aspects to allow us to accept the model at face value. In this final section, we will consider some possibilities for making the model work better and account more fully for the linguistic and empirical data.

Consider the domain in which the different processes operate. The domain of Level 1 processes is the *word.* This is evidenced by the internal phonological changes such as stress shift, vowel reduction etc. that accompany Level 1 rules. For Level 2 and 3 affixes the domain is the *boundary,* which is indicated by the restriction of phonological effects to final consonants. Let us suppose that when the child acquires Level 2 and 3 forms, the juncture between the stem and affix is maintained in lexical representation. For example, when *-ism* is attached to *Darwin,* the representation would be something like: *#Darwin#ism#.*

On the other hand, since Level 1 processes require access to the whole word for internal phonological changes, affixation cannot maintain the boundary and would require a merger of the stem and affix into a single morph. Hence, *-ion* added to *direct* would be represented as *#direction#*—where the stem and affix become merged into a simple lexical form. Such merger, while making the morphology opaque, does not prohibit recursive application of intralevel morphological processes. Hence, the application of other Level 1 affixes such as *-al* and *-ity* would not be blocked in forming *directionality.* Neither would Level 2 and 3 affixes be prevented from applying further. However, the Level 1 affixes could not attach to a complex word formed by Level 2 affixation. For example, if one tried to attach *-ian* to *#Darwin#ism#,* there would be a boundary marker blocking off the stem. In essence then, *-ian* would end up attaching to *#ism#,* not to *#Darwinism#.* This would violate the requirement that Level 1 rules operate over the domain of the *word,* and hence would not be acceptable.

How might such proposals fit with the current data? First, the representation for a Level 1 complex form such as *#directionality#* is no different from a primitive base form. Thus, *directionality* might be acquired as an unanalyzed lexical item, or else through a derivational process—the result would be the same. Relating this to the empirical data, it is no longer a problem that the Level 1 forms in Experiment 2 showed elevated acceptance rates compared to the low frequency forms of experiment 1 (see Table 1). Some productivity for Level 1 forms is allowed under these proposals. However, the phonological deformations of stress shift and vowel reduction would not make them as transparent, and hence productive as Level 2 and 3 forms.

In summary, it has been shown that the initial model for the acquisition of ordering constraints was fatally flawed for both linguistic and empirical reasons. By adding the notion of *merger* and *domain of application,* the revised model fits both with the empirical data and appears to make better predictions for judgements of acceptability of complex forms. There are clearly many forces, both diachronic and synchronic, that constrain the structure of the lexicon and the set of allowable words of English. What the present results suggest is that children might have a significant part in this process.

Appendix

Experiment 1

ln 0	ln 1	ln 2	ln 3	ln 4	ln 5
Log frequency					
taunt	hoist	pious	debt	hero	tree
adorn	equator	hazard	generous	cousin	win
bicep	gloat	null	statue	fashion	difficult
limerick	axle	cruel	territory	kitchen	rise
parsley	drab	gasoline	passenger	hill	sit
jinx	tariff	charcoal	protein	suffer	party
polio	meteor	humble	fiber	glass	early
smog	masculine	ethic	proud	poem	car
vortex	discrete	lens	clerk	join	example
Distractors					
quead	murn	keer	friggle	grote	bewlet
enton	cront	suthrane	prenkle	jord	grumult
naltom	pobel	wult	uverk	sabaign	muit

Experiment 2

Level 1	Level 2	Level 3
coloration	commonness	incomes
visitation	thinness	*fours
progression	greenness	*middles
conduction	roundness	*filming
officious	answerable	*bonding
wondrous	expectable	*modeling
gaseous	believable	imagining
studious	chargeable	*fielded
evidential	spreader	*pained
relational	walker	typed
motional	loser	*radioed
figural	grower	
Nonwords		
glantion	tostilness	drimacles
zelpinous	siebomable	vornting
fyssomal	componter	crofested
Function words		
when-ation	than-ness	from-s
more-ous	this-able	that-ing
become-al	very-er	about-ed

* Category change.

Experiment 3

Level 1	Level 2	Level 3
coloration	clearness	shuts
visitation	thinness	leans
desertion	greenness	rushes
suction	roundness	freezes
sensuous	answerable	releasing
wondrous	readable	repairing
thunderous	believable	capturing
studious	chargeable	flushing
familial	picker	ached
bridal	walker	swelled
facial	talker	hugged
figural	grower	dipped

References

ALLEN M. (1978). *Morphological investigations.* Unpublished doctoral dissertation, Department of Linguistics, University of Connecticut, Storrs, Conn.

ANDERSON, S. R. (1982). Where's morphology? *Linguistic Inquiry,* **13,** 571–612.

ARONOFF, M. (1976). *Word formation in generative grammar.* Linguistic Inquiry Monograph, Vol. 1, Cambridge, MA: MIT Press.

BERKO, J. (1958). The child's learning of English morphology. *Word,* **14,** 150–177.

BRADLEY, D. C. (1979). Lexical representation of derivational relation. In M. Aronoff and M. L. Kean (Eds.). *Juncture.* Cambridge, MA: MIT Press.

CHOMSKY, N., & HALLE, M. (1968). *The sound pattern of English.* The Hague: Mouton.

CLARK, E. V. (1981). Lexical innovations: How children learn to create new words. In W. Deutsch (Ed.). *The child's construction of language.* New York: Academic Press.

CLARK, E. V. (1982). The young word maker: A case study of innovation in the child's lexicon. In E. Wanner & L. R. Gleitman (Eds.) *Language acquisition: The state of the art*. Cambridge, England: Cambridge University Press.

CLARK, E. V., & CLARK, H. H. (1979). When nouns surface as verbs. *Language*, **55**, 767–811.

CLARK, E. V., & HECHT, B. F. (1982). Learning to coin agent and instrument nouns. *Cognition*, **12**, 1–24.

CUTLER, A. (1983). Lexical complexity and sentence processing. In G. B. Flores d'Arcais and R. J. Jarvella (Eds.). *The process of language understanding*. New York: Wiley.

FRANCIS, W. N., & KUCERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

GORDON, P. (1985). Level-ordering in lexical development. *Cognition*, **21**, 73–93.

HALL, W. S., NAGY, W. E., & LINN, R. (1984). *Spoken words: Effects of situation and social group on oral word usage and frequency*. Hillsdale, NJ: Erlbaum.

KIPARSKY, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst and N. Smith (Eds.). *The structure of phonological representations*. Dordrecht: Foris.

KIPARSKY, P. (1983). Word-formation and the lexicon. In F. Ingemann (Ed.) *Proceedings of the 1982 Mid-America Linguistics Conference*. University of Kansas, Lawrence, Kansas.

MOE, A. J., HOPKINS, C. J., & RUSH, R. T. (1982). *The vocabulary of first-grade children*. Springfield, IL: Thomas.

SIEGEL, D. (1977). *Topics in English morphology*. Unpublished doctoral dissertation, Department of Linguistics, MIT, Cambridge, MA.

SLOBIN, D. I. (Ed.). (1985). *The cross-linguistic study of language acquisition. Vol. 1: The data*. Hillsdale, NJ: Erlbaum.

TAFT, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, **7**, 263–272.

TYLER, A. E. (1986). *Acquistion and use of English derivational morphology: An experimental investigation*. Unpublished Ph.D. dissertation, Department of Linguistics, University of Iowa.

WALSH, L. (1984). *Possible words*. Paper presented at the M.I.T. Morphology Workshop, Jan. 26–27, 1983, Cambridge, MA.