Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Psychology 47 (2003) 119–163
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Language learning and innateness: Some implications of *Compounds Research*

Todd R. Haskell,[a] Maryellen C. MacDonald,[b,*] and Mark S. Seidenberg[b]

[a] *University of Southern California, USA*
[b] *Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI 53706-1696, USA*

Accepted 28 October 2002

Abstract

In noun compounds in English, the modifying noun may be singular (*mouse-eater*) or an irregularly inflected plural (*mice-eater*), but regularly inflected plurals are dispreferred (**rats-eater*). This phenomenon has been taken as strong evidence for dual-mechanism theories of lexical representations, which hold that regular (rule-governed) and irregular (exception) items are generated by qualitatively different and innately specified mechanisms. Using corpus analyses, behavioral studies, and computational modeling, we show that the rule-versus-exceptions approach makes a number of incorrect predictions. We propose a new account in which the acceptability of modifiers is determined by a constraint satisfaction process modulated by semantic, phonological, and other factors. The constraints are acquired by the child via general purpose learning algorithms, based on noun compounds and other constructions in the input. The account obviates the regular/irregular dichotomy while simultaneously providing a superior account of the data.

1. Introduction

In English and many other languages, words are inflected to indicate properties such as tense and number. The English past tense is formed by adding the suffix

* Corresponding author. Fax: 1-608-262-4029.
E-mail address: mcmacdonald@facstaff.wisc.edu (M.C. MacDonald).

0010-0285/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0010-0285(03)00007-0

spelled *-ed* (the phonological realization of which is conditioned by the preceding phoneme, as in *baked*, *baled*, and *baited*), and the plural by adding the suffix spelled *-s* (as in *bucks*, *bales*, and *buses*). Although the majority of plurals and past tenses are formed this way, there are well-known exceptions: the plural of *mouse* is *mice*, not *mouses*, and the past tense of *take* is *took*, not *taked*. Research on inflectional morphology has played an important part in a larger debate concerning the nature of linguistic knowledge and how it is acquired, used in skilled performance, and represented in the brain. Based on extensive analyses of many aspects of the past tense, Pinker and his colleagues (e.g., Marcus, Brinkman, Clahsen, Wiese, & Pinker, 1995; Pinker, 1991, 1994, 1999) have argued that English inflectional morphology illustrates an important characteristic of language, that it involves the use of rules. The knowledge that underlies regular forms such as *cats* and *walked*, which obey the rule, is said to be categorically different from the knowledge that underlies exceptions such as *mice* and *ran*, which violate the rule. Regulars and irregulars are thought to involve different types of knowledge (a rule, a lexicon) that are acquired by different learning mechanisms (rule-induction, "rote learning"), and represented in different brain regions (Ullman et al., 1997).[1] This yields a dual-mechanism theory (Pinker, 1991; see Coltheart, Curtis, Atkins, & Haller, 1993; Davelaar, Coltheart, Besner, & Jonasson, 1978, for applications of this idea to reading). An alternative approach described by Rumelhart and McClelland (1986) emphasizes the similarities between rule-governed forms and exceptions, and suggests that both are learned and represented as part of a single connectionist lexical processing system (see Seidenberg & McClelland, 1989 for applications of this idea to reading). This debate has continued for some time (e.g., Clahsen, 1999; Joanisse & Seidenberg, 1999; Marchman, 1997; Marcus et al., 1995; Patterson, Lambon-Ralph, Hodges, & McClelland, 2001; Pinker & Prince, 1988; Plunkett & Marchman, 1991; Plunkett & Marchman, 1993; Prasada & Pinker, 1993; Ullman et al., 1997).

The focus of the present paper is on a part of English in which the inflectional system interacts with another aspect of grammar, the formation of compound words. Both pluralization and compounding are highly productive processes. However, there is an apparent restriction on the ways in which these two processes can be combined, as illustrated by the examples in (1) (the asterisk indicates an ungrammatical form):

(1) (a) rat-eater
    (b) mouse-eater
    (c) *rats-eater
    (d) mice-eater
    (e) rat/mouse/mice-eaters

An animal that eats rats could felicitously be called a *rat-eater* (i.e., a kind of meat-eater), and similarly for *mouse-eater*. An observation dating from at least Kiparsky (1982) is that whereas *rats-eater* is conspicuously ill-formed, *mice-eater* is not.

[1] We use the terms rule-governed and regular interchangeably, and the same for exception and irregular.

A major difference between the two, of course, is that *rats* is the rule-governed plural of *rat*, whereas *mice* is an exception. Also note that the plural may be added to the head noun *eater*, as in *rat/mouse/mice-eaters*. These examples suggest that regular plurals may not occur as the left-most word (commonly referred to as the *modifier*) in a noun-noun compound. This pattern of acceptability has been important in the rules-versus-connections debate for two reasons. First, it has been argued to provide strong converging evidence for the existence of inflectional rules from a non-obvious source (Pinker, 1991, 1999). Second, it presents a classic learnability problem: given the nature of children's experience, how could they arrive at the correct grammatical generalizations? The proposed answer is that acquiring this knowledge is possible only because the language-learner brings innate knowledge of grammar to bear on the learning process. Thus, this small aspect of the grammar of one language illustrates in a particularly clear manner the poverty of the stimulus argument (Chomsky, 1965).

In this paper we present data that call into question the traditional characterization of these phenomena and develop an alternative account within the probabilistic constraints framework discussed by Seidenberg (1997) and Seidenberg and MacDonald (1999). In this approach, the well-formedness of constructions such as compounds is a graded function of several probabilistic constraints involving different types of information (e.g., semantics, phonology, and syntax). We identify two of the major constraints and show that they account for a variety of behavioral data of the sort previously taken as evidence for separate rule and exception mechanisms. The learnability implications of this theory are also discussed; the theory suggests how the constraints could be learned from information available to the child. These findings cast a somewhat different light on the nature of the innate capacities that make learning this and other aspects of language possible.

1.1. Plural modifiers and level-ordering

The examples in (1) present a puzzle: Out of several superficially similar cases, why is only *rats-eater* disallowed? Kiparsky (1982) offered an answer to this puzzle based on the notion of level-ordering (Allen, 1978; Siegel, 1974). In this view, morphologically complex words are formed by the application of rules, like the plural and past tense rules mentioned above. These rules are organized into several ordered levels, such that the output of rules in later levels cannot serve as the input to rules in earlier levels. Kiparsky (1982) assumed that regular forms such as *rats* are formed in a different way than irregular forms such as *mice*, an early example of a dual-mechanism approach. Singulars like *rat* and *mouse* are assumed to be stored in the lexicon, while irregular plurals like *mice* are generated at Level 1. In contrast, regular plurals like *rats* are formed by application of a rule farther downstream at Level 3. This approach is in keeping with a longstanding assumption within morphological theory that the lexicon contains only the idiosyncratic aspects of words, i.e., those not predicted by rules (Halle, 1990). Kiparsky's (1982) further assumption was that the rule governing the formation of compounds applies at Level 2, before the rule for generating the regular plural. Given these assumptions, the facts in (1) follow directly.

More recently, Pinker (1994, 1999) has discussed a more psycholinguistically oriented theory based on level-ordering which incorporates the following main assumptions:

(a) uninflected, base forms of words (e.g., *take*, *bake*) and morphologically irregular forms (e.g., *took*) are stored in the mental lexicon;
(b) regular, inflected words (such as *baked*) are generated by rule rather than stored;
(c) the rule governing the formation of inflected words applies after the rule governing compound formation.

This account can explain the facts in (1), and it is consistent with other evidence that the English inflectional system involves two components, a rule system and a lexicon.

Although the Kiparsky and Pinker accounts differ in detail, they both assume that whereas uninflected and irregular forms are available to enter into compounding, regularly inflected forms are not. The theories also share the idea that regular forms are generated by rule and that word formation rules are ordered. Pinker's theory stresses the idea that singulars and irregular plurals are both stored in the mental lexicon and thus should pattern together; Kiparsky's theory includes further claims about how the lexicon is structured (e.g., into multiple levels associated with different rules and morphological elements). References to the "level-ordering theory" in this paper are intended to include both the original Kiparsky account and the Pinker theory, which incorporates major components of it. The larger contrast is between these theories and the probabilistic constraints approach, which does not incorporate morphological rules or levels, or distinguish between a rule component and a lexicon, and uses the same constraint satisfaction mechanism for all forms.

The implications of these phenomena concerning language learning have also been noted. Gordon (1985) found that young children are sensitive to the prohibition against plural modifiers. He asked 3- to 5-year-olds what they would call someone who eats, say, rats. In this case they almost always said *rat-eater*. When the question was posed about *mice*, however, the children readily produced *mice-eater*. Gordon found very few examples of plural-containing compounds in a corpus of English. The early age at which the constraint against regular plurals was observed, taken with the apparent lack of relevant examples or parental instruction, makes it hard to see how the constraint could be learned from experience. This suggested to Gordon that the fact that word formation rules apply at distinct levels must be innate (see Pinker, 1994, 1999, for discussion).

1.2. Problems with level-ordering

Level-ordering provided a seemingly elegant account of data such as (1); moreover, the account of these data was attractive because it fell out of a general theory that applied to many other phenomena. However, it is important to note that the level-ordering theory itself subsequently fell out of favor. As Spencer (1991) noted, "Almost as soon as [level-ordering] was proposed a series of difficulties were exposed, which have ultimately caused many morphologists and even certain Lexical

Phonologists to reject the idea'' (p. 179). In the years since Spencer's comment, level-ordering has faded further from active consideration. Perhaps the only context in which it has continued to receive serious attention is the debate about rule-based versus connectionist theories of morphology. Because level-ordering has been an important component of rule-based accounts, we briefly review some of the problems with this approach here.

The major problem for level-ordering with respect to compounding is the existence of exceptions such as *awards ceremony*, *pilots union*, *sports announcer*, *weapons inspector*, and many others. English seems to readily accommodate such exceptions: our paper is about compounds research, not compound research; the experiments were conducted in the Neurosciences Building at USC, not the Neuroscience Building. These kinds of examples, which were recognized early in the development of the level-ordering theory, raise questions about its descriptive adequacy.

The main approach to the exceptions over the years has been to assume that something like the level-ordering account is essentially correct for all of the cases to which it applies, but that one or more other factors license the exceptions. Kiparsky (1982) suggested that the exceptions might be explained by a semantic factor. He observed that the regular plurals that appear as modifiers tend to have idiosyncratic meanings that cannot be derived in a straightforward way from the meaning of the singular. Because these meanings cannot be computed by rule, such words must be stored in the lexicon, causing them to pattern with the irregulars. Although this analysis was consistent with some of the exceptions, it was not specific enough to explain why plurals such as *pilots* qualify as idiosyncratic, whereas plurals such as *rats* do not.

A more promising approach was developed by Alegre and Gordon (1996b), who also attempted to identify semantic bases for the exceptions. Rather than treating the acceptable plural modifiers as having idiosyncratic meanings, they focused on systematic semantic properties that might cause them to pattern together. Consider the contrast between *that store carries paint* and *that store carries several different paints*; rather than indicating multiple instances, this use of *paints* actually indicates multiple kinds (e.g., several different colors). Alegre and Gordon (1996b) used the term *heterogeneous* for plurals used in a way that, like *paints* in the example above, highlights diversity among the things being referred to. Alegre and Gordon (1996b) noticed that heterogeneity seems to be a necessary condition for regular plural modifiers to be acceptable. However, they also noticed that acceptable regular plural modifiers tend to be abstract. They proposed that a regular plural must be both heterogeneous and abstract (in the context of a particular compound) to be acceptable. An example would be *publications catalog*. In this compound, *publications* refers to many different publications, rather than multiple instances of the same one. Furthermore, a superordinate term like *publications* is highly abstract relative to more specific terms like *book* or *magazine*. Since this compound satisfies both criteria, it should be acceptable. Alegre and Gordon (1996b) reported ratings concerning compound acceptability that were consistent with this account. Specifically, the acceptability of plural modifiers was related to how abstract and heterogeneous they are, as indicated by independent ratings: regular plural modifiers were acceptable only if they were both heterogeneous and abstract.

We think it likely that something like the semantic factor studied by Alegre and Gordon (1996b) affects the well-formedness of plural modifiers and discuss their proposal further below. However, their proposal is successful in large part because it introduces powerful new mechanisms to account for the exception cases. This raises the question of whether such mechanisms might be sufficient to account for the entire range of cases, not just the exceptions, thus making level-ordering superfluous. In fact, this is precisely what we suggest below (see also Fabb, 1988).

Other problems for level-ordering are created by phrasal modifiers such as *employee-of-the-month competition* (Lieber, 1988), expressions that should be disallowed because compounding precedes phrase-level syntax. One possibility is to add a loop from syntax (where the phrase is formed) back into morphology (where the compound is formed), but this seems to defeat the purpose of the sequential ordering idea that was central to the level-ordering approach; as Spencer (1991) noted, "Many linguists regard this as an admission that level ordering is not the right way to approach the problem of morpheme ordering" (p. 115; see also Selkirk, 1982).

Interestingly, Alegre and Gordon (1996a) observed that many examples of regular plural modifiers involve phrasal modification. For example, *red rats eater*, meaning an eater of red rats, seems somewhat more acceptable than *rats eater*. They suggested that the hypothesized loop from syntax back to morphology provides one possible means for regular plurals to surface as modifiers. As Alegre and Gordon (1996a) acknowledge, this loop circumvents level-ordering. Since *rats* is just as good a phrase as *red rats*, the application of the loop needs to be constrained in some fashion to account for the fact that one is acceptable as a modifier while the other is not. To our knowledge, there has been no explicit proposal about how this could be achieved.

As serious as these problems are, level-ordering faces a slew of others. Several examples involve the attachment of multiple affixes to words, as in *standard–standardize–standardization*, and *book–bookish–bookishness*. Not all affixes can combine freely, and level-ordering was initially thought to provide the necessary constraints on multiple affixation. However, this account was subsequently shown both to permit many affix combinations that don't occur in English (Fabb, 1988), and also to predict the ungrammaticality of forms that do readily occur in the language (Aronoff, 1974).

To summarize, a variety of facts about compounding and other aspects of word formation have called into question the adequacy of the original level-ordering theory. As long as attention is focused on examples such as (1), level-ordering appears to provide an elegant account. When a broader range of data is considered, however, the theory runs into difficulties that seem to require fundamental changes to the theory or an entirely different type of account. In recognition of some of these limitations, Pinker (1999) recommended that level-ordering not be taken literally, but rather "as laying out the *logic* of word formation" [p. 181, emphasis in original]. However, his discussion does not make clear how the data that are problematic for level-ordering can be explained within a theory that retains its "logic" but not its substantive proposals. The validity of the level-ordering account has to be considered with respect to all the phenomena it was intended to address, not merely the restricted subset involving regular plural modifiers in compounds.

This situation leaves an important question unanswered: What is the explanation for the facts? It is clear that regular plurals cannot freely occur as modifiers in compounds; if the level-ordering account is inadequate, what other kind of theory can do better? In the remainder of this paper we pursue this question, addressing two issues. The first concerns the data. Theorizing about the interaction of inflectional morphology and compounding has relied heavily on acceptability judgments for a small number of examples such as the ones in (1) and exceptions such as *parks department*. We obtained independent data concerning the frequency of occurrence of different forms and their relative acceptability, as rated by native speakers. These data provide information about the general character of compounding that is not apparent from isolated examples. Second, we develop and assess an alternative account of the factors that affect the acceptability of different types of modifiers. We view these phenomena within the probabilistic constraints approach (Seidenberg, 1997; Seidenberg & MacDonald, 1999) in which the acceptability of linguistic forms is a graded function of multiple probabilistic constraints. These constraints obviate the need for level-ordering and the distinction between rule-governed cases and exceptions. In this paper we investigate two major constraints, involving semantic and phonological information that underlies important aspects of the phenomena.

2. Study 1

Our first step was to obtain basic data about how often regular and irregular plurals are used as modifiers in compounds, relative to non-modifier uses. There is little existing evidence about how often plural modifiers occur in English. It is therefore unclear whether the exceptions represent a small number of accidental cases or a broader tendency in the language.

A baseline against which the frequency of plural modifiers can be assessed is provided by the behavior of nouns in non-modifier contexts. To obtain a quantitative measure of this behavior, we used the parsed version of the Brown corpus produced by the Penn Treebank project (Francis, 1964; Marcus, Santorini, & Marcinkiewicz, 1993). This corpus contains approximately 1 million words of English, drawn from a variety of written (and some spoken) sources. In this corpus, about 70% of tokens of non-modifier nouns are singular, and 30% are plural. Thus, the percentage of plural uses (hereafter plural percentage), is 30%. In modifier contexts, level-ordering suggests that nouns that would ordinarily be produced in their regular plural form will be singular instead. Thus, if the prohibition against plural modifiers is real, the plural percentage for regular nouns in modifier contexts should be zero, or close to zero if the grammar tolerates a few highly marked exceptions. For irregular plurals, level-ordering allows either the singular or the plural form to surface as modifier. In the absence of any reason to believe the contrary, one might therefore expect the plural percentage for irregular nouns to be about the same for modifier and non-modifier contexts.

We tested these predictions by placing all nouns in the corpus into one of four categories created by crossing syntactic context (modifier or non-modifier) and plural

type (regular or irregular). A noun was classified as a modifier if it was a sister to and immediately preceded another noun in the same noun phrase. A noun was classified as a non-modifier if it did not immediately precede another noun in the same noun phrase. Mass nouns and pluralia tanta (e.g., *scissors*) were excluded; for purposes of this analysis, a noun was considered a mass noun if it appeared in the corpus only in singular form, and a pluralia tantum if it appeared only in plural form (these criteria are only approximate; the goal was to perform the analysis using clear cases). The plural percentages for these categories are shown in Fig. 1; the raw counts are presented in Table 1.

The plural percentage was lower for regular than irregular nouns in modifier contexts [$\chi^2(1) = 21.08, p < .001$]. Similarly, for regular nouns, the plural percentage was markedly lower in modifier than non-modifier contexts [$\chi^2(1) = 4388, p < .001$]. These findings are consistent with level-ordering in that regular plural modifiers appear to be strongly dispreferred relative to both singulars and irregular plurals. However, for irregular nouns, the plural percentage was also markedly lower in

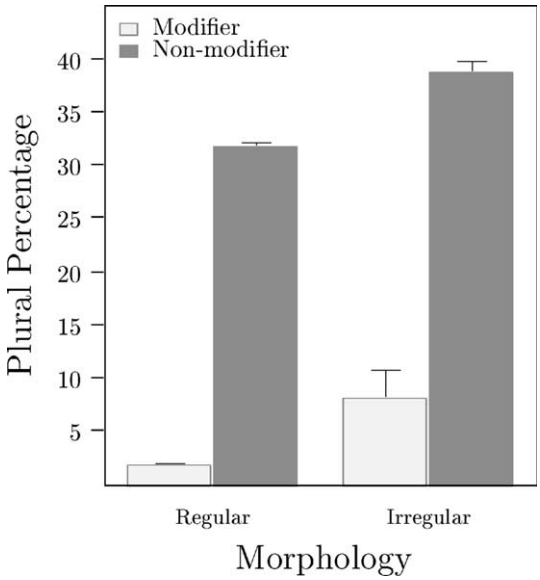

Fig. 1. Plural percentage for regular and irregular nouns in modifier and non-modifier contexts for the Brown corpus.

Table 1
Raw counts for modifier and non-modifier uses of nouns

	Regular		Irregular	
	Modifier	Non-modifier	Modifier	Non-modifier
Singular	10,801	93,160	120	3154
Plural	233	44,084	11	2029

modifier than non-modifier contexts [$\chi^2(1) = 49.79, p < .001$], a finding not predicted by level-ordering. Furthermore, regular plural modifiers, while proportionally rare, occurred at a non-trivial rate: there were about twice as many tokens of regular plural modifiers as irregular *singular* modifiers (233 vs. 120).

These data present a somewhat different picture than that suggested by previous research: There appears to be a bias (but not a prohibition) against plural modifiers of all types, and this bias seems to be stronger for regular than irregular plurals. Thus, the level-ordering account misses two important generalizations: that there is a dispreference for all plural modifiers (irregular as well as regular), and that this dispreference is probabilistic rather than absolute.

Corpus data such as these must be interpreted cautiously. While these results begin to call into question the assumptions on which the level-ordering account is based, the frequencies with which particular forms occur depend in part on non-grammatical factors such as properties of the sample of texts that constitute the corpus. In addition, our interpretation of the results depends on the assumption that, in the absence of any constraints, people would use plural nouns just as often in modifier contexts as in non-modifier contexts. This assumption might not be valid; there may be semantic or pragmatic reasons why speakers would be more likely to use plurals in one case than in the other. To address these concerns, Study 2 examined whether similar results obtained when native speakers judged compound acceptability.

3. Study 2

The purpose of Study 2 was to further investigate the patterns observed in the corpus analysis, particularly the finding that both regular and irregular plurals appear to be underrepresented as modifiers. In Study 2 we collected graded grammaticality judgments for novel compounds in which the modifier was either singular or plural, and either regular or irregular. If the level-ordering account is correct, compounds containing singular or irregular plural modifiers should be judged as completely acceptable, while compounds containing regular plural modifiers should be judged as completely unacceptable. If, however, the corpus data accurately reflect the underlying acceptability of different modifiers, then irregular plural modifiers should actually be rated worse than singulars.

3.1. Method

3.1.1. Participants

Seventy two undergraduate college students participated in the experiment. All were native speakers of English and either received course credit or were paid a nominal sum for their participation.

3.1.2. Materials

Fourteen nouns were chosen to serve as modifiers, seven irregulars (e.g., *mouse*) and seven regulars (e.g., *rat*). In choosing the irregular nouns, we excluded 'Latinate'

(e.g., *phenomena*) and no-change plurals, nouns with ambiguous plural forms (e.g., *people* vs. *persons*), and nouns whose singular or plural was unlikely to be familiar to the typical undergraduate. Words were also excluded if they were ambiguous for part of speech and the noun sense was not the most common (based on Francis & Kucera, 1982). For each irregular noun item, a regular control was chosen, as similar as possible in meaning to and having the same number of syllables as the experimental item. For each of these seven irregular/regular pairs, we chose two head nouns that could be meaningfully combined with either member of the pair to form a compound (e.g., *rat/rats/mouse/mice observation*). Two head nouns were chosen, rather than just one, so that each stimulus list could contain both the regular and irregular nouns without repeating a head noun (e.g., one list contained both *mice observation* and *rats condos*, while another contained *rats observation* and *mice condos*).

To further increase the sensibility of the compounds, a short context paragraph (one to three sentences) was generated for each head noun. The test compound was always at the end of this context. The plural form of the test noun always appeared at some earlier point in the context, regardless of whether the test item appeared in singular or plural form in the test compound (see Gordon, 1985). An example item is given in Table 2. The full set of experimental items and their controls is listed in Appendix A.

Four counterbalanced lists were generated from these items. Each list contained each modifier and each head exactly once. The order of the heads was randomized once and used for all lists. Across lists, each modifier appeared in singular and plural forms with both heads. In addition to these items, we also generated three example items and seven fillers which appeared in all the lists. An additional seven items appeared in each list as part of a separate manipulation (see Study 4).

3.1.3. Procedure

The compounds were presented to participants in a written questionnaire. Each item consisted of the context paragraph, ending in the underlined test compound. Participants were instructed to read the context, and rate on a seven-point scale

Table 2
Example stimulus for each condition in Study 2

Regular Singular[a]	Amy's **toes** had been hurting for several days, and she wanted someone to check them out. She called the clinic to schedule a **toe** examination.
Regular Plural	Amy's **toes** had been hurting for several days, and she wanted someone to check them out. She called the clinic to schedule a **toes** examination.
Irregular Singular[a]	Amy's **teeth** had been hurting for several days, and she wanted someone to check them out. She called the clinic to schedule a **tooth** examination.
Irregular Plural	Amy's **teeth** had been hurting for several days, and she wanted someone to check them out. She called the clinic to schedule a **teeth** examination.

[a] Throughout this paper, the term "regular singular" will be used to refer to the singular form of nouns with regular plurals. Similarly, "irregular singular" will be used to refer to the singular form of nouns with irregular plurals.

how good the compound sounded to them, with seven indicating the best rating. Participants were told that a 'bad' compound was one that seemed awkward or sounded 'off'; to minimize influences of prescriptivist teaching, no explicit mention of grammaticality was made. The questionnaire began with three example items with illustrative ratings, followed by three fillers. The remaining fillers were interspersed with the test items. Equal numbers of participants saw each of the four forms.

3.2. Results

The mean acceptability ratings in each condition are plotted in Fig. 2. An ANOVA revealed a significant main effect of number: plural modifiers were rated worse than singular modifiers [$F_1(1,71) = 120.8, p < .001$; $F_2(1,6) = 33.7, p < .01$]. The main effect of morphological regularity did not reach significance [$F_1(1,71) = 1.8$, n.s.; $F_2(1,6) = 1.4$, n.s.]. There was also a significant interaction between number and morphology [$F_1(1,71) = 20.3, p < .001$; $F_2(1,6) = 6.9, p = .04$]. To explore the interaction more fully, pairwise comparisons were conducted. Regular plurals were rated worse than matched irregular plurals, as level-ordering would predict, an effect that was significant by participants and marginal by items [$t_1(71) = 4.1, p < .001$; $t_2(6) = 2.2, p = .07$]. Consistent with the corpus data, irregular plurals were also rated as significantly less acceptable than their corresponding singulars [$t_1(71) = 5.2, p < .001$; $t_2(6) = 2.8, p = .03$]. Finally, irregular singulars (i.e., singulars with irregular plural forms) were actually slightly less acceptable than regular singulars, although this effect was marginal in the items analysis [$t_1(71) = 2.5, p = .01$; $t_2(6) = 2.4, p = .06$].

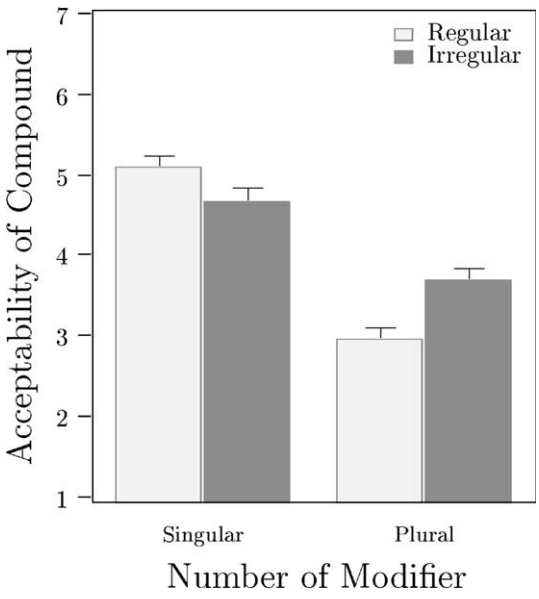

Fig. 2. Mean acceptability ratings for regular and irregular nouns as modifiers in Study 2.

3.3. Discussion

The ratings confirm that regular plural modifiers are dispreferred, consistent with the level-ordering account. However, two other aspects of the data are noteworthy. First, although level-ordering predicts that singulars and irregular plurals should pattern together and that both should be better than regular plurals, the irregular plurals actually fall between the singulars and regular plurals. The ratings data are consistent with the corpus analysis in providing evidence for a bias against both types of plural modifiers, not merely regular ones. Second, singulars that have irregular plurals were rated worse than singulars with regular plurals. This is the complement of the findings for plural modifiers, for which irregulars are preferred over regulars. This result should be considered cautiously because it was only marginally significant by items. However, one interpretation is that the acceptability judgments for the singular and plural forms of a word are linked: the degree of preference for one affects the degree of dispreference for the other. Thus, regular singulars are highly preferred and regular plurals highly dispreferred; irregular singulars are somewhat less preferred (compared to regular singulars) and irregular plurals less dispreferred.

Some potential insight into this effect comes from considering connectionist networks in which a single set of weights encodes all forms. Settings of the weights that are beneficial for one form may disfavor a related form. Thus, evidence that the irregular plural is preferred pushes the weights toward values that simultaneously act against the related singular form; similarly, evidence that the singular is preferred acts against the related regular plural. Thus, competition between similar forms arises due to their aggregate effects on the weights. This type of effect is seen in the triangle reading model (Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996; Harm & Seidenberg, in press). Experience with the spelling-sound patterns of *save*, *gave* and *pave* in these models results in weight settings that are no longer optimal for the pronunciation of *have*, even though all these forms can be produced correctly.

4. Toward an alternative theory

The results of the first two studies indicate that the level-ordering account provides a poor fit to the data: it makes incorrect predictions (e.g., about the relative acceptability of irregular plural modifiers) and it misses an important generalization (that there is a bias against plural modifiers in general, not merely regular ones). That we have discovered additional problems with level-ordering is perhaps unsurprising given other known limitations of the theory. In this section we begin developing an alternative account of these phenomena and then describe additional studies that bear on it.

As a working hypothesis we assume that the well-formedness of compounds, like other linguistic structures, is a graded function of multiple probabilistic constraints (Allen & Seidenberg, 1999; Seidenberg, 1997; Seidenberg & MacDonald, 1999). In

this view, linguistic elements (e.g., phonemes, morphemes, and words) are defined in terms of correlations between different types of information (e.g., sound, meaning, and contexts). In the course of learning to comprehend and produce utterances, language users acquire information about the properties of these different elements (e.g., Dell, Reed, Adams, & Meyer, 2000). For example, they may acquire information regarding what nouns tend to sound like, what meanings they tend to have, and what contexts they occur in. This information can be viewed as constraints which define what a typical or acceptable noun is. The acceptability of a given use of a noun is then defined in terms of the extent to which it satisfies these constraints. The constraints are graded (a noun may have a meaning which is very typical, somewhat typical, or highly atypical of nouns in general), and the effect of one constraint may depend on the value of other constraints. For example, a word which has frequently occurred in noun contexts may be considered a good noun regardless of its phonology. However, words that have typically been used as verbs are more readily extended to noun uses when they have phonological properties typical of nouns, compared to items without typical noun phonology (Kelly, 1988). In applying this general framework to phenomena such as compounding, the goal is to identify what the relevant constraints are, how they are learned, how they give rise to observed behavior, and (more ambitiously) why languages exhibit the particular constraints they do and not others that could be imagined.

Descriptively, the results of Study 2 can be accounted for in terms of two factors. One is whether a potential modifier is plural in meaning. This will be termed the semantic constraint. *Rats* and *mice* pattern together on this dimension and differ from *rat* and *mouse*. The second factor is whether a potential modifier has the phonological structure typical of a regular plural, hereafter the phonological constraint. On this dimension, *rat*, *mouse*, and *mice* pattern together and differ from *rats*. Only the latter contains the /s/ that is a proper realization of the plural inflection. Note that *mice* could not be a regular plural, e.g., of *mi* because the [s] allomorph must be preceded by a voiceless consonant (as in *cats*), not a vowel (as in *mice*). Assume that each potential modifier loads either positively or negatively on each factor (Table 3). This yields the outcomes in Table 3, which are consistent with the results of Study 2, ignoring for the moment the difference between the two types of singulars. The idea that semantic and phonological factors are behind the differing acceptability of regular and irregular plural modifiers was also part of a somewhat different proposal put forward by Chapman (1996). Pinker (1999) also mentioned the possibility that certain plural modifiers are disallowed because they sound strange, but quickly discarded it.

Table 3
Prediction of modifier acceptability by semantic and phonological factors

Example	Semantically plural?	Phonologically plural?	Acceptability
Rat, mouse	No	No	Acceptable
Mice	Yes	No	Marginal
Rats	Yes	Yes	Dispreferred

This sort of account differs from level-ordering in an important way: It explains the data in terms of semantic and phonological properties of words—information that children have to learn anyway—rather than levels of structure in the mental lexicon or ordered application of rules. Level-ordering emphasized morphological distinctions among words rather than semantic or phonological ones. Minimal pairs such as *mice-eater*/**rats-eater* were taken to indicate that the constraint against plural modifiers is grammatical rather than semantic; *mice* and *rats* are semantically similar but only one is a grammatical modifier (Pinker, 1999). The semantic factors proposed by Kiparsky (1982) and Alegre and Gordon (1996b) were intended to supplement level-ordering, and were only relevant to the exceptional cases. On our view, the semantics of the modifier does matter, as one of several properties of words that are weighed by the constraint satisfaction mechanism. The semantic and phonological properties discussed above are assigned the greatest weight, but overall goodness of fit may be modulated by other constraints (this issue is discussed further below). The competing theories then make different empirical predictions, for example with respect to whether exceptions should pattern with regulars.

Although it has some descriptive validity, our analysis of modifier preferences was developed post hoc; an important question is whether there is any converging evidence for it. The next two studies address this question by examining whether this account makes correct predictions about cases other than the ones used in developing it.

5. Study 3

The two-factor account presented above generates an additional case not considered in Table 3: nouns that are semantically singular, but resemble regular plurals phonologically. Nouns of this sort are predominantly terms for bifurcate objects, i.e., objects with joined symmetrical parts, such as *pants*, *scissors*, and *binoculars*. These nouns form a subset of the larger class of *pluralia tantum* nouns, or nouns appearing only in the plural form. Although phonologically and grammatically plural, ratings data collected by Bock, Eberhard, Cutting, Meyer, and Schriefers (2001) indicated that English speakers consider bifurcate pluralia tanta to be semantically singular. This leads to a novel prediction: bifurcate pluralia tanta, which are phonologically plural but semantically singular, should be the complement of irregular plurals such as *mice* (phonologically singular but semantically plural). Thus, like irregulars, bifurcate pluralia tanta have one strike against them, and therefore should be intermediate in acceptability between singulars and regular plurals. According to the level-ordering account, because these nouns are idiosyncratic forms, they should be stored in the lexicon, and therefore pattern with singulars. Indeed, Pinker (1999) refers to pluralia tanta as "irregular regulars," and suggests that they are quite acceptable as modifiers. The literature contains little data relevant to this issue. Gordon (1985) included a few pluralia tanta among his stimuli, but the pattern of results was inconclusive. Study 3 was conducted to carry out

direct comparisons among singulars, regular plurals, and bifurcate pluralia tantum nouns.

5.1. Method

5.1.1. Participants

Twenty seven native speakers of English participated in the experiment. Participants volunteered or were paid a small sum for their participation.

5.1.2. Materials

The stimuli consisted of 11 triples of nouns. Each triple consisted of the singular and plural versions of a regular noun, and a semantically similar bifurcate pluralia tantum noun (e.g., *hammer*, *hammers*, and *pliers*). All members of a triple had the same number of syllables and stress pattern. Each triple was paired with a context sentence. The experimental items appeared in the context sentence as the modifier of a noun-noun compound at the end of the sentence (e.g., *Put that back on the hammers rack.*). Three lists were generated, each containing one randomly selected member of each triple. The order of the experimental items was randomized once, and then used for all three lists. Three practice items were also included at the beginning of each list. The experimental items are listed in Appendix B.

5.1.3. Procedure

For each list, the items were embedded in their context sentences and presented as a paper-and-pencil survey. Following each sentence was a 7-point rating scale. Participants were instructed to circle a number on the rating scale to indicate how 'good' the compound sounded to them. Circling a '7' indicated that the compound sounded very good, while circling a '1' indicated that the compound sounded very bad. The survey took approximately five minutes to complete. Equal numbers of participants completed each of the three list versions.

5.2. Results and discussion

The mean ratings in each condition are presented in Fig. 3. The analysis of variance yielded a significant main effect of morphological type (singular, pluralia tantum, or plural) [$F_1(2, 52) = 22.6$, $p < .001$; $F_2(2, 20) = 14.3$, $p < .001$]. Planned comparisons indicated that singulars were significantly more acceptable than pluralia tanta [$t_1(26) = 4.6$, $p < .001$; $t_2(10) = 5.5$, $p < .001$], and also that pluralia tanta were more acceptable than plurals, although only by participants [$t_1(26) = 2.2$, $p = .04$; $t_2(10) < 1$].

The results indicate that pluralia tanta are significantly less acceptable modifiers than singulars, contrary to level-ordering and Pinker's (1999) claims for these nouns. As with the irregular plurals studied in the previous experiment, the pluralia tanta fell between the singular and regular plural baselines, consistent with the claim that preferences are jointly determined by the semantic and phonological factors discussed above.

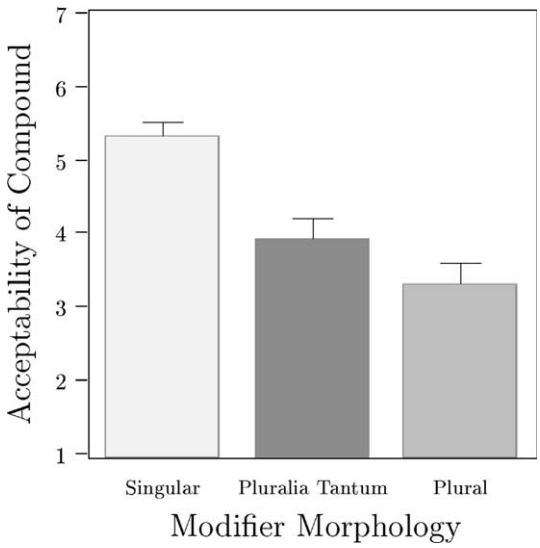

Fig. 3. Mean acceptability ratings for singulars, bifurcate pluralia tanta, and regular plurals as modifiers in Study 3.

6. Study 4

The data presented to this point are consistent with the constraint-based theory; whether other theories could be modified in order to account for them is considered in the general discussion. Next we consider whether our account extends to yet another type of plural. In so-called voicing-change (VC) plurals such as *loaf–loaves* and *wolf–wolves*, the unvoiced stem-final consonant /f/ in the singular becomes the voiced /v/ in the plural. Although some nouns undergo this change, others do not: compare *loaf–loaves* with *oaf–oafs*. The question then is how these items fare as modifiers. Because the voicing change is an idiosyncratic property of certain nouns, the basic level-ordering approach predicts that these plurals should be listed in the lexicon like other irregulars and be acceptable modifiers. Alternatively, Pinker (1999) treats voicing-change plurals as a variant of the ordinary regular plural. Under this treatment, they should be unacceptable as modifiers.

For the constraint-based theory, the important issue is how these words fare with respect to the hypothesized constraints. Voicing-change plurals are semantically plural and they have the phonological suffix typical of regular plurals; on this basis they should pattern with regular plurals and both should be strongly dispreferred compared to singulars. However, it is also true that the VC plurals differ from regular plurals in one respect: the extent to which the singular form is preserved in the plural. Regular plurals preserve the base word (*rat* occurs in *rats*) whereas *wolf* is deformed in creating *wolves*. Study 4 also assessed whether this aspect of VC plural phonology affects preferences.

6.1. Method

This experiment was run concurrently with Study 2 and used the same participants and procedure. The structure of the materials was the same as in Study 2, except that the experimental items were voicing-change plurals. These stimuli are listed in Appendix C.

6.2. Results

Mean acceptability ratings for voicing-change items and their regular controls are given in Fig. 4; these results can be compared to those from Study 2 (Fig. 2), which yielded a similar pattern. An ANOVA revealed a significant main effect of number: plural modifiers were rated worse than singular modifiers [$F_1(1,71) = 78.7$, $p < .001$; $F_2(1,6) = 97.9$, $p < .001$]. The main effect of type of morphology was significant by participants but not by items [$F_1(1,71) = 10.2$, $p < .01$; $F_2(1,6) = 2.4$, n.s.]. There was again a significant interaction between number and morphology [$F_1(1,71) = 6.8$, $p = .01$; $F_2(1,6) = 7.9$, $p = .03$]. Pairwise comparisons showed that for singular modifiers, regulars and voicing-change nouns were rated approximately the same [t's < 1], while for plural modifiers, regulars were rated worse than voicing-change plurals, though this effect was marginal by items [$t_1(71) = 4.0$, $p < .001$; $t_2(6) = 2.3$, $p = .06$]. Finally, though more acceptable than regular plurals, the voicing-change plurals were rated as significantly less acceptable than their corresponding singulars [$t_1(71) = 5.9$, $p < .001$; $t_2(6) = 7.7$, $p < .001$].

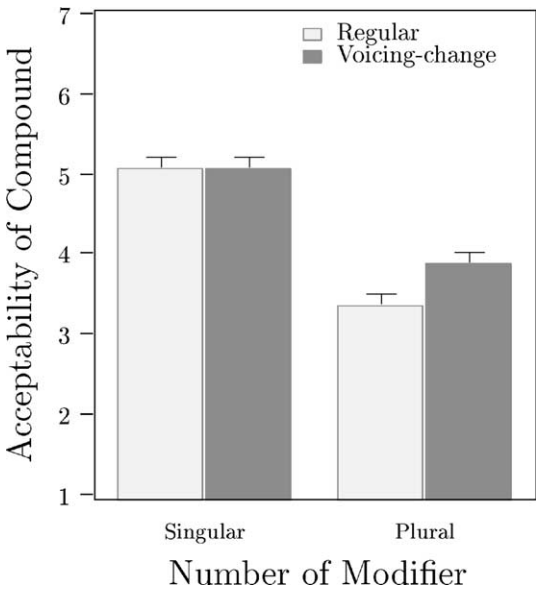

Fig. 4. Mean acceptability ratings for regular and voicing change nouns as modifiers in Study 4.

6.3. Discussion

This study provides additional evidence that singular nouns are the most acceptable modifiers and regular plural nouns the least, consistent with level-ordering and the results of Studies 2 and 3. Interestingly, the VC plurals patterned like the irregular plurals and pluralia tanta studied earlier: They are significantly less acceptable than singulars but more acceptable than regular plurals. Thus, VC plurals provide a second example of an intermediate case. Such cases are difficult to explain in the level-ordering framework, which assumes a regular/irregular dichotomy. The data are more consistent with the idea that modifier acceptability is a graded function of constraints that reflect semantic and phonological factors that differentiate among singular and plural forms.

However, the results also allow us to be more specific about the nature of the phonological constraint. We initially formulated it as "whether a potential modifier has the phonological structure typical of a regular plural." Studies 2 and 3 focused on perhaps the most prominent aspect of regular plural phonology, presence of the plural inflection. The VC plurals remind us that most regular plurals have another property: unlike the VC plurals, regular plurals preserve and incorporate the singular form. This way in which the VC plurals differ from fully regular forms such as *rats* can account for why the two types did not pattern together. Thus, "phonological similarity to the regular plural" is a non-unitary factor, involving at least two aspects of the structure of regular plurals.

7. Origins of the constraints

The evidence presented to this point suggests that it may be possible to explain the relative acceptability of different kinds of modifiers in terms of their semantic and phonological properties. The plausibility of the approach would be increased if it could be explained why these particular constraints are relevant and how children could learn them. As Gordon (1985), Pinker (1999) and others have noted, it is not clear how children learn the grammar of compound formation. All of the standard learnability arguments (e.g., concerning the lack of parental instruction or consistent feedback about grammaticality) apply. Moreover, although the corpus analysis (Study 1) was based on a sample of adult texts, it suggests that plural modifiers appear relatively rarely and that there is not a large difference between the frequencies of regular and irregular plural modifiers. Assuming that caregiver speech to children exhibits similar properties, if the child were attending only to how often these forms occur, there would be little basis for concluding that one is permitted and one is not. The question then is whether the semantic and phonological constraints we have discussed could be learned from information available to the child.

We think the answer to this question has the following form, in broad outline: The constraints that are relevant to compound formation are not specifically about the formation of compounds; rather they are inherited from more general properties

of pluralization and compounding that are exemplified many times over in the input to the child. Although a child is unlikely to have heard *mice-eater*, children are exposed to many noun-noun compounds such as *toy box*, *baby blanket*, and *fire truck*, as well as other constructions involving prenominal modification, such as adjective-noun phrases (*big box*, *yellow blanket*, and *squeaky truck*). Similarly, the child has a great deal of experience with plurals outside of the context of compounds, and thus has ample opportunity to learn how they behave. These utterances provide positive information about the characteristics of plurals and prenominal modifiers against which novel input can be assessed.

Thus, this account is related to the notion of "indirect negative evidence" (Chomsky, 1981), in the sense that it provides a means by which forms that have not been heard before (e.g., *rats-eater*) can nonetheless be judged ill-formed (see Allen & Seidenberg, 1999; Seidenberg & MacDonald, 1999, for discussion). This process can be modeled using connectionist networks (see Allen & Seidenberg, 1999, for an example) in which the weights on connections between units represent probabilistic constraints, the entire set of which is used in processing every expression. Well-formedness is a function of the extent to which an expression satisfies these constraints. Novel forms that are consistent with the constraints (because they exhibit properties that overlap with ones characteristic of items used in setting the weights) will be more acceptable than ones that are inconsistent with the constraints (e.g., because they exhibit characteristics that did not occur in previously trained items). Importantly, the same mechanism underlies both positive and negative generalizations: moving the weights toward values that reflect characteristics of the input simultaneously moves them away from characteristics that do not occur.

The next two sections consider how the two constraints introduced earlier in the paper might emerge out of more basic facts about pluralization and compounding, in the process moving from a descriptive account of the data toward an explanatory one. In this regard there is an important asymmetry in the status of the two constraints. Most researchers (e.g., Alegre & Gordon, 1996a, 1996b; Pinker, 1999) acknowledge that semantics affects the acceptability of plural modifiers. The points of contention involve the precise nature of the constraint and how children acquire it. Thus, our discussion of the semantic constraint will be relatively brief, and is geared toward addressing these questions. In contrast, the proposal that phonology affects the acceptability of modifiers in compounds is much more controversial. Accordingly, much of the rest of the paper is dedicated to establishing the plausibility of this constraint. In pursuing this goal, a diverse range of evidence is brought to bear, drawn from a corpus analysis, a behavioral experiment, and a computational model.

7.1. The semantic constraint

We hypothesize that the semantic constraint derives from what is likely to be a universal tendency of modification, that modifiers themselves do not have number semantics. For example, in an adjectival phrase such as *red table*, the notion of redness is neither singular nor plural; in general, adjectives can apply to one or many

entities but do not themselves have number meaning.[2] Some authors have noted that modifiers in compounds behave in much the same way (DiSciullo & Williams, 1987). That is, the *kitchen* in *kitchen table* does not refer to any specific kitchen or kitchens, but to the abstract concept of kitchen-ness, which is neither singular nor plural. What the child learns from such examples, then, is that there is a strong tendency for modifiers to lack number, including but not limited to modifiers in noun compounds. Thus our semantic constraint applies to all modifiers, for which abundant evidence is available to the child. This account is in contrast to the proposal made by Alegre and Gordon (1996b), in which semantic effects were restricted to regular plural modifiers in noun compounds.

Ideally, then, the speaker would produce a number-neutral form for the modifier. Some languages, such as Greek, do use such a number-neutral "stem" form for modifiers in compounds. For example, the nominative singular form of the Greek word for spinach is *spanaki*; when it appears in a compound, the stem *spanak* is used e.g., *spanak-o-pita*, literally spinach pie (the *-o-* is a linking morpheme; see Joseph & Philippaki-Warburton, 1987). English, however, requires that modifiers in compounds be able to stand on their own as complete words (i.e., they cannot be bound morphemes). Thus, the speaker is confronted with a choice between singular and plural forms. For a variety of reasons, the singular form carries less number information than the plural form. For example, as noted earlier, singulars are generally higher in frequency than plurals, so from an information theoretic perspective, plurals provide more number information. Similarly, it is the phonological form of the singular, devoid of its number meaning, that is used in derivational word formation processes (e.g., a man who acts or looks young can be *boyish* but not *boysish*). Thus, the singular form is less reliably associated with singular meaning than the plural form is associated with plural meaning. Given these facts, the singular form is a better approximation to a number-neutral form, and is therefore preferred.

However, neither the singular nor the plural form fully satisfies the constraint. If the singular form fully satisfied the constraint, it would always be used, and the plural form would never be used. Instead, the constraint only gives the singular form a slight edge. As a result, if additional factors are present which promote the use of the plural, it may be acceptable or even preferred. Alegre and Gordon (1996a) discuss one such case, illustrated by the contrast between *week-long seminar* and *weeks-long seminar*. Here, the difference between singular and plural modifiers corresponds to a distinct difference in meaning. Thus, it would appear that plural modifiers may be acceptable if they convey a meaning which could not be conveyed by the singular.

The heterogeneity constraint of Alegre and Gordon (1996b) may actually arise through a similar mechanism. As mentioned earlier, Alegre and Gordon (1996b) argued that regular plural modifiers are acceptable if they are both abstract and heterogenous. A distinctive characteristic of heterogenous plurals is that plural marking is

[2] Note that in languages where adjectives exhibit number agreement, adjectives may have "singular" and "plural" *forms*, but the choice between the two forms is entirely governed by the number of the corresponding noun, and there is no discernible difference in *meaning* between these two forms.

used to indicate an idiosyncratic meaning (multiple types, rather than multiple tokens). In other words, the heterogeneity constraint may be another case in which the plural form is used to express a meaning which could not be conveyed by the singular.

Finally, an interesting consequence of this account concerns pluralia tantum modifiers. Because the constraint involves a pressure to avoid number marking on modifiers, it explains why pluralia tanta are sometimes reduced in compounds (e.g., *scissor kick*, *pant leg*; see also Gordon, 1985).

7.2. The phonological constraint

The phonological constraint arises from the fact that exposure to nouns and adjectives provides information regarding the phonological structure of modifiers. From a large sample of exemplars, children learn that modifiers have many and varied phonological properties but not just any: in particular, they tend to not have the phonological structure characteristic of regular plurals. This "negative" generalization can again be learned on the basis of exposure to positive examples, given the view of learning discussed above. The conjunction of semantic and phonological factors then gives rise to the graded effects seen in the above studies.

The idea that children might be picking up on this phonological information and that it plays a role in acceptability judgments is consistent with other recent research on language acquisition. Several studies have demonstrated that distinctions between grammatical categories are correlated with distinctions in the phonology and prosody of words in those categories, and that speakers possess knowledge of these correlations. For example, English disyllabic nouns tend to have first syllable stress, while disyllabic verbs tend to have second syllable stress. Kelly (1988) presented speakers with disyllabic nonce words having either first-syllable or second-syllable stress, and asked them to use the nonce word in a sentence. Speakers were more likely to use the nonce words as verbs when they had second-syllable stress. Furthermore, it has been suggested that such correlations might play an important role in the initial acquisition of grammatical categories (Kelly, 1992; Morgan, Shi, & Allopenna, 1996; Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999).

Given these links between phonology and grammatical category membership, it would not be surprising if phonological factors played a role in grammatical acceptability judgments. In fact, Gropen, Pinker, Hollander, Goldberg, and Wilson (1989) reported evidence for a relationship of this sort. English provides two ways to form dative constructions: the prepositional dative (*Naomi threw the ball to Ricardo*) and the double-object dative (*Naomi threw Ricardo the ball*). However, some verbs may only be used with the prepositional dative: *Naomi donated the book to the library* is acceptable, while *Naomi donated the library the book* is not. Monosyllabic verbs are more likely than polysyllabic verbs to be acceptable in the double-object construction. Gropen et al. (1989) taught children novel verbs, and found that they were more likely to use them in double-object constructions when the verbs were monosyllabic than when they were disyllabic. Thus, children are apparently aware of the phonological correlates of grammatical verb behavior and make use of them in their own production.

Extending this line of reasoning to the compounding case, it is necessary to establish three points. First, it must be shown that there are phonological cues to whether a particular word can be used as a modifier. Second, it must be shown that speakers are sensitive to these cues. Finally, it must be shown that this knowledge influences the acceptability of modifiers in compounds. The next three studies address each of these points.

8. Study 5

We have proposed that the dispreference for regular plurals as modifiers derives in part from the fact that they do not sound like a typical modifier. Conversely, irregular plurals are more acceptable because they do sound like typical modifiers. It needs to be determined, however, whether these generalizations are valid. Specifically, do singulars and irregular plurals differ systematically from regular plurals with regard to phonological structure? Phonology would not provide a valid cue if, for example, the singulars included a large proportion of words such as *lens*, which has the phonological structure of a regular plural. The goal of Study 5 was to assess this question empirically.

8.1. Method and results

All adjectives and singular common nouns occurring three or more times in the Treebank version of the Brown corpus (Francis, 1964; Marcus et al., 1993) were selected for analysis. This resulted in a list of 3151 adjectives and 6342 singular nouns. Next, phonological spellings for these words were extracted from the CMU Pronouncing Dictionary (Carnegie Mellon University, Pittsburgh, PA). Not all words on the list had phonological spellings in the CMU dictionary; thus, our final set of words consisted of 2598 adjectives and 5647 singular nouns. Although this list did not contain all adjectives and singular nouns in English, it is comprehensive enough that it should reflect the general characteristics of adjectives and singular nouns in the language.

Each of these words was examined to determine whether or not it was a possible regular plural in English. The English plural has three allomorphs, [s], [z], and [əz]. The choice between allomorphs is conditioned by the final phoneme of the stem. If the final phoneme is a voiceless consonant, the [s] allomorph is used (e.g., *cats*). If it is a voiced consonant or a vowel, the [z] allomorph is used (e.g., *dogs*). However, if the stem-final phoneme is a strident (voiced or unvoiced), the [əz] allomorph is used instead (e.g., *houses*). Thus, in order for a word to be a possible regular plural in English, it must not only end in one of the three possible allomorphs, but that allomorph must be the appropriate one for the phonological context in which it occurs. For example, although *blouse* ends in [s], it is not a possible regular plural in English, because this phonological context would require the use of the [z] allomorph. In contrast, *lens* is a possible regular plural in English.

Using these criteria, each adjective and singular noun in our list was coded for whether or not it was a possible English regular plural. The results are presented in Table 4. As can be seen on the left side of the table, a substantial portion of ad-

Table 4
Percentage of adjectives and singular nouns which are possible plurals in English

	Allomorph		Allomorph in correct context	
	Types	Tokens	Types	Tokens
Adjectives	8.2% (212)	4.4% (2840)	0.5% (12)	0.2% (149)
Singular nouns	9.5% (538)	11.2% (16,273)	1.6% (89)	1.4% (2051)

Note. The left side of the table presents data on words which end in a regular plural allomorph; the right side of the table presents data on words which end in a regular plural allomorph in the appropriate context. Raw counts are given in parentheses.

jectives and singular nouns end in one of the three plural allomorphs. However, only rarely do these allomorphs appear in the appropriate context, with the result that about 0.5% of all adjectives and 1.5% of all singular nouns are possible regular plurals of English. This figure is approximately the same whether one is considering the number of words (type frequency) or the number of instances (token frequency).

8.2. Discussion

The corpus analysis indicates that adjectives and singular nouns, which make up the vast majority of prenominal modifiers in English, are rarely possible regular plurals. Thus, regular plurals deviate from typical modifier phonology. The next step was to determine whether language users are sensitive to distributional information of this sort. This was the goal of Study 6.

9. Study 6

Study 5 demonstrated that regular plurals have phonological characteristics that differentiate them from the words (adjectives and singular nouns) most commonly used as prenominal modifiers. The goal of Study 6 was to evaluate whether speakers are sensitive to this distinction. We collected acceptability judgments for nonce words whose phonology was similar either to singular nouns or to regular plural nouns. The nonce words were placed in contexts that forced them to be interpreted as adjectives. If English speakers are sensitive to the typical phonological characteristics of modifiers (adjectives and singular nouns), they should rate nonce adjectives that sound like regular plurals as less acceptable than nonce adjectives that sound like singular nouns.

9.1. Method

9.1.1. Participants

Fifty native speakers of English were paid a nominal sum to participate.

9.1.2. Stimuli

Twenty nonce words that have the phonological characteristics typical of singular nouns (i.e., do not end with the plural inflection) were generated, as were the

corresponding plural forms (e.g., *fant*, *fants*). Twenty-four additional items were used as fillers; some of these were part of a separate study. Due to the small number of irregular plurals in English, it was not possible to create a similar set of "irregular" items. However, the phonology of irregular plurals is generally more similar to that of singular nouns than to that of regular plurals, e.g., *mice* and *price*, *men* and *pen*, *feet* and *sheet*. Each nonce word appeared in a context sentence. The nonce word always appeared at the end of the sentence, and the context always strongly constrained the nonce word to be an adjective. The prenominal modifier position was not suitable in this case because it would have allowed the nonce words to be interpreted as either nouns or adjectives, and therefore does not allow us to assess the impact of phonology separately from the impact of grammatical and morphological factors. Context sentences were randomly assigned to nonce items. The nonce words and contexts are presented in Appendix D.

The sentences (context + nonce word) were digitally recorded using Sound Edit Pro software (Macromedia, San Francisco, CA) and a Macintosh computer. The sentences were spoken by a female native speaker of English using a natural but lively prosody. For the experimental items, separate recordings of the entire sentence were made for the "singular" and "plural" versions. To avoid presenting the same context sentence more than once, two lists were generated; one version of each experimental item was assigned to each list in a pseudo-random fashion such that both lists had equal numbers of singular and plural versions. The order of items within each list was randomized once and this order was used for all participants presented with that list.

9.1.3. Procedure

Participants were tested individually in a small, quiet room. The experiment was controlled by a Macintosh computer running the PsyScope software package (Cohen, MacWhinney, Flatt, & Provost, 1993). On each trial, a warning signal was briefly presented on the computer screen, and then the recorded stimulus was played over the computer's speakers. At the conclusion of the sentence, a five point rating scale was presented on the screen. Participants were instructed to rate how acceptable the nonce word was as an adjective by using the mouse to select a point along the scale. Although the scale was numbered at regular intervals, participants were told that they could click anywhere along the scale, not just on the numbers. The location of the mouse click was recorded. After the rating was made, there was a brief pause followed by the next trial.

9.2. Results and discussion

Nonce words with regular plural phonology had a mean adjective acceptability rating of 1.39 ($SE = .02$). This value was significantly lower than for nonce words with singular noun phonology (1.50, $SE = .02$) [$F_1(1, 49) = 33.7$, $p < .001$; $F_2(1, 19) = 12.4$, $p < .01$]. Importantly, the only systematic difference between the two items in each pair was the presence vs. absence of the plural ending, suggesting that this was the factor that affected performance. Note that the ratings were quite

low on average; this was true for the fillers as well as the experimental items, and probably reflects a reluctance to rate nonce words as highly acceptable.

In summary, Studies 5 and 6 indicate that it is a distributional property of English that the phonological structure of regular plurals differs from that of adjectives, and that speakers are sensitive to this fact. Our final step was to develop an independent measure of how modifier-like a word's phonology is and determine if it predicted both the nonce adjective ratings collected in Study 6 and the modifier acceptability ratings collected in Studies 2 and 4.

10. Study 7

In the final study we attempted to obtain converging evidence concerning the role of phonological information in determining preferences for modifiers in compounds. If phonological information plays an important role in modifier acceptability, then it should have an effect above and beyond any influence of morphological structure. One way to address this question would be to look at words for which phonological and morphological cues conflict, e.g., *box*, which is a singular noun but sounds like a regular plural. The phonological constraint should cause such words to be somewhat less acceptable as modifiers than other singulars. An arguably better strategy would be to examine the acceptability of adjectives or singular nouns which do not have regular plural phonology, but whose phonological characteristics are nevertheless rare for words in modifier position. Again, the phonological constraint should cause such words to be somewhat less acceptable.[3] However, there is another way to disentangle morphological and phonological effects, by using a computational model. The model can be provided with information about the phonology of words, but not their morphological structure. One can then see whether the behavior of the model can be used to predict the human data. To the extent that the model is successful, it provides support for the claim that the regular/irregular distinction is not necessary to account for variations in modifier acceptability.

The logic of the study was as follows. We used an algorithmic procedure to derive a measure of how adjective-like a word is, based on phonological information alone. The procedure involved training a connectionist network to classify phonological word forms as adjectives or non-adjectives. This model was not intended as an account of how children acquire this knowledge, which involves other factors; it was only used as a procedure for generating a measure of relative adjective-likeness based on phonological structure. This use of a connectionist network to discover phonological properties that probabilistically differentiate between classes of words is similar to work by Cassidy, Kelly, and Sharoni (1999), who used a similar network to discover probabilistic phonological differences between male and female names.

The only way the model could learn to perform the categorization task was by picking up on any phonological differences between adjectives and non-adjectives

[3] We thank an anonymous reviewer for these two suggestions.

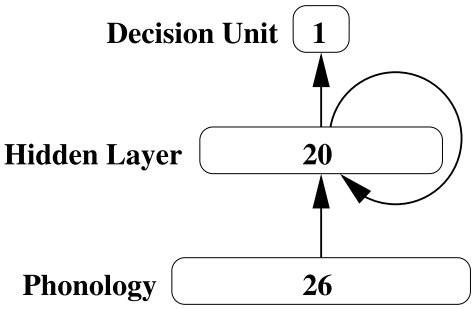

Fig. 5. Architecture of the network used in Study 7.

implicit in the corpus of training examples. Thus, the network was used as a discovery procedure. The network cannot learn to perform this task perfectly because the phonological differences between adjectives and non-adjectives are probabilistic rather than absolute. The model's performance on any given word can therefore be interpreted as providing a quantitative estimate of how adjective-like it is. The first goal, then, was to examine how well the model was able to learn the task, which would provide independent evidence regarding the extent to which adjectives and non-adjectives differ systematically in their phonology. The second goal was to determine whether the model's assessment of how adjective-like a word is could predict the behavioral data from Study 6, subjects' ratings of how adjective-like words are. Finally, we examined whether the model's performance predicted subjects' judgments of the acceptability of modifiers in compounds from Studies 2 and 4. This would suggest that acceptability judgments were affected by the phonological properties of words that the network detected in learning the task.

10.1. Model architecture

The architecture of the network used in this study is shown in Fig. 5. The phonological layer consisted of 25 units representing segmental features, and an additional unit representing stress. The representation of segmental features was drawn from Harm (1998). The phonological layer was connected to a hidden layer consisting of 20 units. In turn, this hidden layer had recurrent connections to itself, as well as connections to a "decision" layer consisting of a single unit. In addition, the network contained a "bias" unit (not shown in the figure) which was connected to the hidden layer only, and whose activation was always set to one.[4] The logistic sigmoid activation function was used for all units (see Rumelhart & McClelland, 1986, Chapter 10).

[4] The bias unit is a generalized version of the intercept parameter in linear regression, and serves the same purpose (see Bishop, 1995, Chapter 3).

10.2. Representation

Each word was presented to the model as a sequence of phonemes. At time $t = 0$, the units representing the segmental features of the first phoneme in the word were set to a value of 1, and all other units in the phonological layer were set to 0. At time $t = 1$, this process was repeated for the second phoneme, and so on, until all phonemes had been presented. For vowels receiving primary stress, the stress unit was set to 1 concurrently with the appropriate segmental units.

The target output for a given word was 1 if the word was an adjective, and 0 if the word was not an adjective. For purposes of computing error, this target was compared with the activation of the decision unit two time steps after the final phoneme of the word was presented to the model, as it took two time steps for the contribution of this phoneme to reach the decision unit.

10.3. Materials and training procedure

The training set for the model consisted of adjectives and non-adjective open-class words drawn from the Treebank Brown corpus (Francis, 1964; Marcus et al., 1993). Non-adjectives were defined as those words which did not have an adjective sense; thus, words that have both an adjective and a non-adjective sense were treated as adjectives. In assessing the model's ability to classify words as adjective or non-adjective based on phonology, it was important that the decision of the model be based only on phonological information, and not on other cues available to the model. In other words, in the absence of any phonological information, the model's performance should not exceed chance. This necessitated certain constraints on the training set. The first was that the model was trained on approximately equal numbers of adjectives and non-adjectives. If the training set, like the English language, contained far more non-adjectives than adjectives, then the model could achieve above-chance performance simply by making a non-adjective response most or all of the time, without regard to phonology. Thus, instead of constructing a training set containing proportions of adjectives and non-adjectives comparable to those in the language, we began by selecting the highest frequency words of each type, then selecting words of progressively lower frequency until we had at least 1000 words of each type, samples that were large enough to allow the network to pick up on phonological differences between types if they exist. Phonological spellings for these words (including stress information) were obtained from the CMU Pronouncing Dictionary (Carnegie Mellon University, Pittsburgh, PA). The final lists contained 1010 adjectives and 1060 non-adjectives. 10% of the words in each list (101 adjectives and 106 non-adjectives) were then withheld from the training set to form the generalization set. The remaining words (909 adjectives and 954 non-adjectives) constituted the training set. The words in the training and test sets together accounted for 88% of all adjective tokens and 64% of all non-adjective tokens in the Brown corpus.

The frequency with which a given word was presented to the network during training was proportional to the natural log of its frequency (see Plaut et al., 1996; Seidenberg & McClelland, 1989, for discussion of the use of log frequencies

for training), such that more frequent words received more training than less frequent words. However, the thousand most frequent non-adjectives are much higher in frequency on average than the thousand most frequent adjectives, with the result that the median frequencies of the adjectives and non-adjectives were 25 and 110 respectively. Although this difference would not provide a basis for the model to improve its overall performance, it would (because of the nature of the model) cause the model have a bias toward a non-adjective response. This would result in better performance on non-adjectives than adjectives for reasons that have nothing to due with phonology. To compensate for this, the frequencies of the non-adjectives were scaled so that the median frequency was approximately the same for both types of words. This procedure preserved the frequency relationships of words within each word type, thus providing the model with information about the relative frequency of phonological patterns within the classes of adjectives and non-adjectives, but ensured that frequency could not be used as a basis for classification decisions.

Prior to training, all weights in the model were set to randomly chosen values between −0.1 and 0.1. The model was then trained for 500 cycles through the training set, which typically resulted in the best generalization performance; similar results were obtained with models trained for shorter or longer periods. Weights were updated after each cycle through the training set using the back-propagation-through-time algorithm (Elman, 1990). The learning rate was set to .005, and the momentum parameter was set to .5.[5] These values were found, by trial and error, to result in satisfactory learning of the training set.

10.4. Results

10.4.1. Performance on training and test items

The model was trained three times, each time with different initial weight values and a different subset of words set aside as the test set. The performance reported here is the average of the three runs. For the purpose of assessing the model's overall performance, it was considered to have made an "adjective" decision if the activation of the decision node was greater than .5, and a "non-adjective" decision otherwise. By this criterion, the model correctly classified 80% of the words in the training set (75% of the adjectives and 84% of the non-adjectives) at the conclusion of training. This level of performance is high given that the model had access only to phonological information about the to-be-classified words. This finding is consistent with observations by several researchers (e.g., Kelly, 1992; Morgan et al., 1996) that phonological information is potentially a good cue to grammatical category, although to the extent that it is language specific, the child may have to achieve a certain vocabulary size to fully exploit it.

The generalization test considered how accurately the model categorized words on which it had not been trained. The model correctly classified 75% of the words in the test set (70% of the adjectives and 79% of the non-adjectives). Thus, performance de-

[5] See Bishop (1995), Chapter 7, for discussion of learning rate and momentum.

clined only slightly compared to the trained words. This level of performance indicates that there is a strong relationship between the phonological properties of a word and whether or not it is an adjective, and that the model was able to learn about this relationship based on exposure to the words in the training set. We then turned to using the behavior of the model to try to predict the human data collected in Studies 2, 4, and 6.

10.4.2. Prediction of Study 6 ratings

In Study 6, we obtained ratings of adjective-likeness for a set of nonce words. The model can also be used to generate a measure of adjective-likeness based on the degree to which the decision unit was activated: the closer it is to 1, the more adjective-like the word is. To compare the human ratings with the model's behavior, the item means from Study 6 (both experimental items and fillers) were plotted against the model's output for the same words, averaged across the three runs; this is shown in Fig. 6. The two measures exhibited a significant positive correlation: Words that evoked a stronger adjective response from the model also tended to receive higher adjective ratings from human raters [$r = 0.48$, $t(73) = 4.61$, $p < .01$]. Thus, both model and human performance appear to have been influenced by the same phonological differences between adjectives and non-adjectives.

10.4.3. Prediction of ratings from Studies 2 and 4

In the previous section it was shown that the output of the model correlated significantly with human ratings of how adjective-like a word sounds. A major claim of

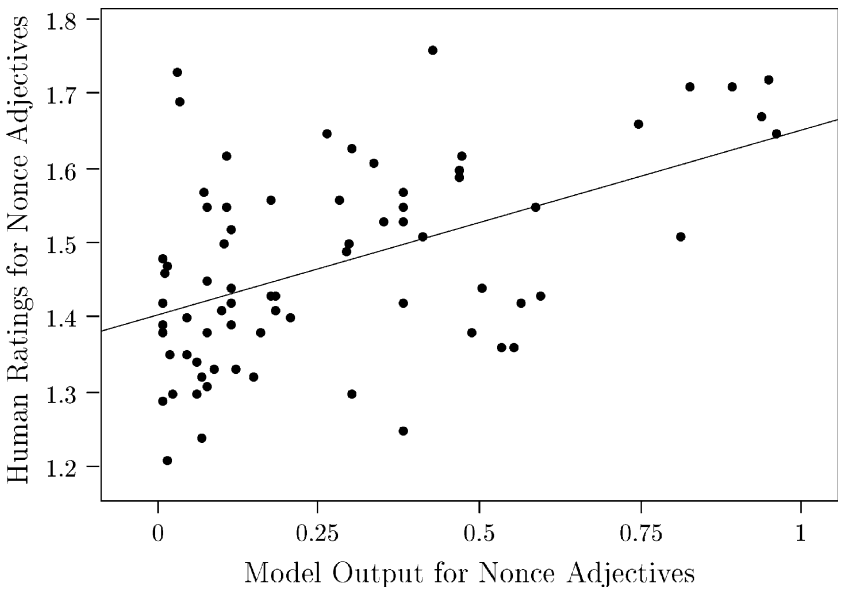

Fig. 6. Adjective acceptability ratings from Study 6 plotted against the raw output of the categorization model.

the probabilistic constraints approach is that language users acquire detailed knowledge regarding what prenominal modifiers (which are predominantly adjectives) sound like, and that this knowledge influences the acceptability of modifiers in compounds. In particular, it is claimed that the reason compounds like *rats-eater* are particularly dispreferred is because regular plural marking results in highly atypical modifier phonology. To establish this link, we examined the relationship between the model's adjective ratings (which are based purely on phonology), and the compound acceptability ratings collected in Studies 2 (regular and irregular modifiers) and 4 (regular and voicing-change modifiers). If modifier acceptability depends in part on having particular phonological properties, and the model has picked up on those properties, then the model's ratings should be correlated with human ratings of modifier acceptability.

For purposes of this analysis, the data from Studies 2 and 4 were combined. Fig. 7 plots the item means from the acceptability ratings against the model's output for the same words, again averaged across the three runs. A very robust positive correlation was observed: Words that produced higher adjective activation in the model also were deemed more acceptable as modifiers in compounds [$r = 0.56$, $t(54) = 5.01$, $p < .01$]. Because the model was presented only with phonological information about the modifiers, it follows that this phonological information is a strong cue to whether or not a given modifier will be acceptable.

10.4.4. Replications with different training sets

The design of the training sets used in the simulation describe above was determined by the purpose of the simulation, which was to determine the extent to which phonological information alone provides a basis for distinguishing adjectives from non-adjectives. The sample of words drawn from the Brown corpus is sufficient for this purpose. However, there are two potential concerns with the nature of these training sets.[6] The first concern is that training the model using scaled rather than raw frequencies means the information available to the model differs somewhat from the information available to a speaker of English. Although the model was intended as a "categorization machine" rather than a claim about how people actually learn and represent phonological information, it is reasonable to ask whether the behavior of the model is strongly dependent on on this aspect of the training procedure. Thus, we replicated the simulation using a different method of controlling for bias that does not involve frequency scaling: rather than using the 1000 most frequent adjectives and non-adjectives, 1000 adjectives and non-adjectives were randomly selected from the corpus. No scaling was applied, since the two sets thus generated were already similar in average frequency. Averaging over three runs of the model, categorization of the training items was 94% accurate (95% for adjectives and 92% for non-adjectives). Generalization performance was similarly high at 89% (92% for adjectives, 86% for non-adjectives). Crucially, the model's output still correlated significantly with the adjective acceptability ratings from Study 6 [$r = 0.43$, $t(73) = 4.11$,

[6] We thank an anonymous reviewer for bringing this to our attention.

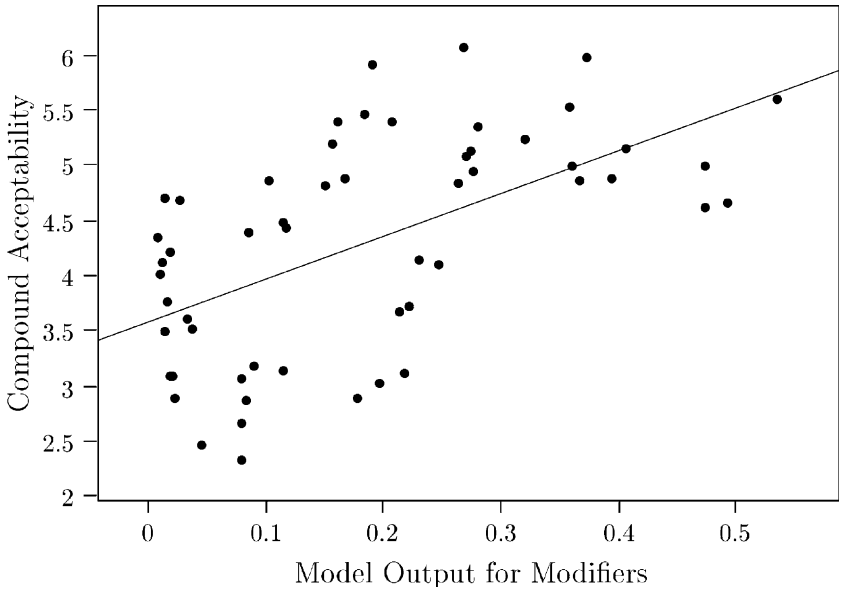

Fig. 7. Modifier acceptability ratings from Studies 2 and 4 plotted against the raw output of the categorization model.

$p < .01$] and the combined compound acceptability ratings from Studies 2 and 4 [$r = 0.45$, $t(54) = 3.71$, $p < .01$]. Thus, the model's performance does not appear to be highly dependent on the precise treatment of frequency in the training set.

The second potential concern with the original training sets was that they were based on a corpus of adult-directed rather than child-directed language. The use of an adult corpus makes sense for purposes of comparing the behavior of the model to that of the adult participants in Studies 2, 4, and 6. However, it is important to demonstrate that the phonological cues that allow the model to distinguish between adjectives and non-adjectives are also available in the subset of words that children are likely to hear. To address this concern, a sample of child-directed speech was obtained from the CHILDES database (MacWhinney, 1995). The sample consisted of all child-directed speech in the Bates (Bates, Bretherton, & Snyder, 1988; Carlson-Luden, 1979; Gleason, Perlmann, & Greif, 1984; Masur & Gleason, 1980) and New England (Ninio, Snow, Pan, & Rollins, 1994) corpora. This sample had been previously collected for a study on the occurrence of prenominal adjectives and noun-noun compounds in child-directed speech. New training sets were then generated using the same procedure as with the replication described above, with the additional constraint that all selected words must appear in the CHILDES sample. The child-directed speech in our sample involved a much more limited vocabulary than the adult-directed speech in the Brown corpus; thus, there were fewer adjectives and non-adjectives to choose from, and it was necessary to use a smaller set of items (400 of each type), rather than the larger set used in the previous simulations (1000 of each type). Note that, if anything, this smaller set should make the task of the model

more difficult. Frequency of presentation during training was based on each word's frequency in the sample of child-directed speech.

Averaging over three runs of the model, accuracy on the training items was 83% (84% for adjectives, 83% for non-adjectives). Performance on the generalization test was somewhat lower but still respectable at 69% (63% for adjectives, 75% for non-adjectives). Finally, the output of the model was still strongly correlated with the adjective acceptability ratings collected in Study 6 [$r = 0.43$, $t(73) = 4.01$, $p < .01$] and the combined compound acceptability ratings from Studies 2 and 4 [$r = 0.49$, $t(54) = 4.13$, $p < .01$]. These results demonstrate that, although adult- and child-directed speech differ in important ways, the model's performance does not depend on phonological information that is present only in adult-directed speech. In addition, the fact that performance deteriorated only slightly with a markedly smaller training set suggests that the child need not master the entire language before being able to make use of phonological cues to adjective status.

10.5. Discussion

An independent measure of phonological adjective-ness was devised and then used to predict both the nonce adjective ratings collected in Study 6 and the modifier acceptability ratings collected in Studies 2 and 4. Crucially, this measure was based on phonological information alone, and did not incorporate information about morphological properties, grammatical number, or meaning. The fact that this measure was able to account for a substantial portion of the variance in modifier acceptability ratings is strong support for the notion that phonological factors play an important role in determining the acceptability of modifiers in compounds.

11. General discussion

The apparent restrictions on the formation of compounds in English illustrated by the data in (1) raise challenging questions about the kinds of linguistic knowledge that give rise to these phenomena and how this knowledge could be acquired. The data originally surfaced in the context of Kiparsky's (1982) work on level-ordering but more recently have figured in the debate about connectionist vs. rule-based theories of language. The level-ordering account of the compounding examples is questionable because of the existence of numerous counterexamples to its basic descriptive generalizations; moreover, the theory ran into similar problems when applied to many other phenomena. The theory therefore provides little leverage in the rules versus connections debate.

Our studies provide additional data concerning these constructions. In general they yielded empirical support for the observation that regular plurals are dispreferred as modifiers compared to singulars; however, the other cases that we examined cast this fact in a somewhat different light. The data from the corpus analysis (Study 1) indicated that the bias is not limited to regular plurals but instead extends to irregulars: Both appear less often as modifiers than would be expected from their overall

frequencies in the language. Thus the bias appears to be against plural modifiers in general rather than only rule-governed plurals. Study 2 showed that irregular plural modifiers were less acceptable than singulars but more acceptable than regular plurals. This finding is important because some theories based on level-ordering (e.g., Pinker's) hold that singulars and irregular plurals should act alike, insofar as both are stored in the lexicon and therefore available to enter into compound formation.

We developed an alternative account in which the well-formedness of these constructions is a function of a constraint-satisfaction process that weighs multiple types of probabilistic information. The essential goals of this type of theorizing are to identify what the constraints are, how they are learned and represented, and how they give rise to the full range of cases. We began by identifying two primary constraints, one semantic and one phonological. Roughly speaking, on this account the acceptability of a modifier is a function of how semantically and phonologically similar it is to singular nouns. This proposal is consistent with the pattern of acceptability judgments in Study 2, and it provides a natural account of why irregular plurals have intermediate acceptability. It also extends without complication to two additional intermediate cases, the pluralia tanta (Study 3) and voicing-change plurals (Study 4).

Consideration of why these constraints are relevant and how they could be learned led us to propose that they are not specific to compound formation but rather derive from more general properties of language, in particular facts about modification, which includes noun-noun, adjective-noun and even predicate adjective constructions. We assume that in the course of acquiring a language, children pick up on phonological and semantic properties of words and facts about the distribution of words and other linguistic elements in sentences. That children engage in statistical learning is indicated by many recent studies (e.g., Jusczyk, 1997; Saffran, Newport, & Aslin, 1996). How such learning might contribute to the development of a lexicon, grammatical categories, verb subcategorization, and other aspects of language has also been the focus of considerable recent research (e.g., Allen & Seidenberg, 1999; Mintz, Newport, & Bever, 2002; Morgan & Demuth, 1996; Redington, Chater, & Finch, 1998; Seidenberg, 1997). The present studies do not provide direct evidence that children and adults encode these properties of modifiers, but the assumption is consistent with behavioral evidence concerning other, similar types of learning. Moreover, the resulting theory accounts for data concerning adult performance presented here and elsewhere.

The phonological constraint reflects the fact that, although their phonological properties are highly varied, modifiers do not tend to have the phonological form of regular plurals, as indicated by the corpus analysis presented in Study 5. This is similar to the idea that regular plural modifiers are unacceptable because they sound strange, which was briefly considered by Pinker (1999) before being rejected. How this negative generalization could be learned is illustrated by connectionist models (e.g., Allen & Seidenberg, 1999; Munakata, McClelland, Johnson, & Siegler, 1997; Seidenberg & McClelland, 1989). The weights on connections between units are set on the basis of exposure to positive examples. The weights represent a set of simultaneous probabilistic constraints that are evaluated every time an example is processed. The well-formedness of a novel input depends on how well it fits the

constraints embodied by the weights. Positive generalization is the case in which a novel example conforms closely to these constraints (e.g., because it is similar to the items that were used in setting the weights); negative generalization is the case in which a novel example deviates from the constraints (e.g., because it differs from training items). Study 5 showed that few singular nouns or adjectives have the phonological shape of the regular plural; Study 6 showed that adult subjects dispreferred nonce modifiers that have the phonological structure of regular plurals; and Study 7 showed that a model-derived measure of similarity to the phonological structure of adjectives accounted for a significant amount of variance in the rated acceptability of modifiers. All of these considerations support the conclusion that acceptability is determined in part by phonological properties of modifiers (specifically, the extent to which they resemble other modifiers). Since few of the attested modifiers have regular plural phonology, words that exhibit this property make poor modifiers.

The semantic constraint arises in a similar manner. The child learns that although the semantics of modifiers are highly varied, they do not tend to include number. Given a choice between a singular and a plural form, the singular form is preferred because it is a closer approximation to a number-neutral form. Technically speaking, then, the constraint is not against semantically plural modifiers per se but rather against any number marking on modifiers. Given certain facts about English, the constraint results in a bias against plurals.

In summary, although more direct evidence that children encode these aspects of language is needed, the above proposal suggests how the child might acquire knowledge of these constraints on compounding from information that is available in the input. The account is tied to information (semantics and phonology) which the child must acquire in the course of learning a language; and the interactions among the constraints explain the full range of acceptability data, including the intermediate cases (irregular plurals, pluralia tanta, and voicing-change plurals).

11.1. Beyond two constraints: Other possible influences on modifier acceptability

Although the account presented in this paper focuses on two factors (one semantic, one phonological), we noted in the introduction that additional factors might play a role as well. Previous researchers have also proposed additional factors as a means for licensing the exception cases. However, the current account differs from previous proposals in a crucial way: All the constraints apply to all the cases, not just the exceptions. Furthermore, because it is claimed that the constraints are to a significant extent learned, rather than being innate, there must be a rational basis in the input to the language learner for any proposed constraint. Given these stringent requirements, there are nevertheless several additional factors which we believe may influence modifier acceptability; each of these is discussed in turn.

11.1.1. Atypical uses of the plural

One way of thinking about the semantic constraint we have proposed is in terms of the Gricean maxims (e.g., of quantity or quality) (Grice, 1975). From this perspective, number information regarding modifiers in compounds is spurious, because such

modifiers typically are neither singular nor plural in meaning. Using a number-marked form for a modifier amounts to providing extra information when it is inappropriate and unnecessary. From a Gricean point of view, then, if a speaker does use a number-marked modifier, there should be a reason for it. In this way plural modifiers can be used to cue the comprehender that a special or different meaning is intended.

The clearest example of this sort of usage involves cases where the singular form conveys the wrong meaning. Returning to an example discussed earlier, *week-long seminar* unambiguously refers to a seminar that is one week in duration. In this case, the plural may be used to indicate that the seminar lasts for multiple weeks, i.e., *weeks-long seminar*. This construction does not violate the semantic constraint, because this particular use of the plural is appropriate and necessary.

More generally, this analysis suggests that plural modifiers will increase in acceptability in cases where the use of plural helps to steer the comprehender toward the appropriate interpretation, or where the plural provides additional information that could not be adequately conveyed using the singular. Consider the compound *securities commission*. The noun *security* is polysemous: in one sense, it relates to safety, in the other, a financial instrument (stock). However, in the financial sense, the noun rarely appears in the singular form. Thus, the compound *security commission* is most naturally interpreted as involving safety, not investment; in this case, the use of the plural helps to ensure that the correct sense of *security* is conveyed. Additional investigation is needed to test this suggestion and to refine the nature of the constraint, that is whether the need to be informative is properly conceived as being part of the semantic constraint or is another constraint that interacts with the two we have identified.

We believe that a related mechanism is responsible for the heterogeneity effects observed by Alegre and Gordon (1996b). The heterogenous use of the plural is not limited to the context of compounds. For example, a seller of fine toiletries might advertise its *large collection of soaps and lotions*. The plural marking here is on two mass nouns (soap and lotion), which typically are not marked for number. This atypical marking is intended to indicate not that the store contains a great many bottles of the same lotion but rather that many different kinds of lotion are available. When the plural is used in this way, it involves a change in meaning beyond the simple contrast between "one" and "many." Because these plural forms convey this extra information about heterogeneity, they are more acceptable as modifiers. As with other examples we have reviewed, the child could come to learn this information not only from modifier examples but also from related phenomena, such as plural markings on mass nouns that are not modifiers, as in the soaps and lotions example. Specifically, children may learn from experience with mass nouns that certain semantic factors license the use of a plural where it would otherwise be unacceptable, and then generalize this observation to the case of prenominal modifiers.

11.1.2. Connectionist networks and frequency

Part of the argument for innateness in the case of the constraint against regular plural modifiers focuses on the fact that the constraint seems non-obvious and arbitrary. One of the central claims made here is that such non-obvious phenomena can emerge from the interaction of several straightforward principles (see Elman et al.,

1996, for extensive discussion of this issue). While such interactions can be a powerful explanatory tool, the potential complexity of such systems means that they cannot always be fully understood by cogitation alone. Fortunately, recent years have seen rapid advances in the development of computational models of human behavior. By implementing such a model, it is possible to simulate the sorts of complex interactions we envision. As mentioned at several points in this paper, connectionist networks in particular possess the kind of learning capabilities which are crucial to acquiring the proposed semantic and phonological constraints. Developing an explicit computational model of the compounding phenomena is beyond the scope of the current paper. However, by considering some of the principles that underlie connectionist networks, it is possible to get a feel for how such a model would behave, as well as to achieve deeper insight into the data presented in this paper.

For example, a salient characteristic of connectionist networks is their sensitivity to frequency. If our approach is on the right track, it would be surprising if frequency were irrelevant to modifier acceptability. Although this issue requires further investigation, we can speculate on the nature of such effects. For example, plural forms that are high in frequency, relative to the corresponding singular, may be more acceptable as modifiers in compounds than are modifiers for which the plural form is relatively rare. This sort of effect may contribute to cases such as *securities commission*, because within the genre of financial discussions *securities* is higher in frequency than *security*. In this sense the frequency is not a separate constraint, in that the distributional patterns would owe to other more basic factors such as the semantic effects discussed earlier. Given networks' sensitivity to frequency and its nonlinear combination of information, however, it is likely that the effects discussed here are modulated by frequency, such that the semantic and phonological constraints affect lower frequency and higher frequency items in somewhat different ways. Thus a better characterization of the frequency of alternative uses will likely improve our understanding of these phenomena.

11.1.3. Plural modifiers and possessives

For most nouns in English, the plural form and the possessive form of the singular sound the same (e.g., *dogs* and *dog's*). As a result, hearers are often exposed to forms that are phonologically indistinguishable from a plural modifier (e.g., *the dog's leash* versus *the dogs leash*). Thus, English speakers sometimes hear non-plural but plural-sounding forms in the prenominal position. Such experience, which can be thought of as another example of a frequency effect (the frequency of a phonological form), may modulate the phonological constraint by making plural-sounding prenominal modifiers somewhat less atypical, and thereby more acceptable. However, use of the possessive form is largely limited to animate nouns (especially those referring to humans). Thus, the homophony of the plural and possessive forms could result in the phonological constraint applying more strongly to inanimate than animate nouns, with the result that, on average, plural-modifier compounds will be more acceptable if the modifier is animate. This effect may contribute to the relative acceptability of forms such as *pilots union*, which otherwise does not seem to share many of the semantic features that promote acceptability of plurals in compounds. This

potential complex interaction of the semantic and phonological constraints is an important avenue for future research.

11.2. Extension to other languages

We have stressed that our account of the compounding data emerges from some general learning mechanisms, some language-universal properties (e.g., that modification does not itself have number), and some particular features of English. It should then follow that analogous effects should emerge in other languages that permit noun-noun compounds, modulated by the particular distributional properties of these other languages. Like English, German permits noun-noun compounding. Clahsen and colleagues (Clahsen, Rothweiler, Woest, & Marcus, 1992; Clahsen, Marcus, Bartke, & Wiese, 1996) have presented data suggesting that, much like English speakers, German speakers regularly produce compounds with irregular plural modifiers, but avoid compounds with regular plural modifiers. Because the distributional facts regarding pluralization and compounding are quite different in German and English, parallel results about plural modifiers would challenge any account in which children deduce the constraint from the language input.

It is difficult to evaluate these claims because the currently available German data are sparse. For example, it is not known whether German has intermediate cases, analogous to the English irregular plurals and pluralia tanta, which as we have seen provide important evidence for evaluating alternative accounts. There is also controversy about whether German has a "regular" plural of the sort described by Marcus et al. (1995) (see e.g., Bybee, 1995; Kopcke, 1998). Additional distributional data are crucial both for a better understanding of the phenomena and a proper analysis of the extent to which the relevant constraints could be learned from the input. In this regard it is interesting to note that there is considerable homophony between the suffixes used to mark plurality on nouns (-, *-e*, *-er*, *-(e)n*, and *-s*), and the suffixes used to mark case, number and gender on adjectives (*-e*, *-er*, *-en*, *-em*, and *-es*). Thus children are routinely exposed to adjective-noun combinations in which the modifier has inflections that sound like plural markings. Interestingly, the *-s* suffix, which is argued to be the "regular" plural in German, is the only plural suffix which isn't homophonous with an adjective suffix, and it is also the only plural suffix that doesn't appear in compounds. This pattern suggests that a phonological constraint quite similar to the one in English may be relevant in German as well, and that children could learn this constraint from available positive evidence. Additional study of the role of homophony in inflections, both in these German cases and in the possessive/plural ambiguity in English, could prove extremely informative in developing a broader account of these phenomena.

11.3. Can other theories account for the data?

In closing we consider whether theories other than the one we have proposed can account for the data. The discussion of the problems with the basic level-ordering theory in the introduction, which was by no means exhaustive, suggests that it has

deep descriptive limitations. There have been attempts to modify the theory to accommodate some of the problematic cases, but these modifications severely compromise the theory. Exceptions to the level-ordering account have been explained in terms of additional mechanisms that vitiate the theory's original strong claims about levels and ordered application of rules. For example, Alegre and Gordon's (1996b) account of instances of regular plurals in phrasal modifiers (e.g., *red rats eater*) involves a feedback loop whereby a rule-generated plural can be cycled back through compounding. The effect is that the rule governing plural formation is sometimes assumed to apply before compounding (ruling out **rats-eater*) and sometimes after (to allow *red rats eater*). As Alegre and Gordon (1996b) noted, this type of account would have to be further constrained in order to retain any explanatory power.

Our arguments have focused on the specific claims of level-ordering, in large part because, to our knowledge, the only existing symbolic accounts are variants on level-ordering. Our data certainly do not rule out the possibility that some other type of dual-mechanism theory might be able to account for the observed patterns of acceptability. However, the additional data we have presented do offer new challenges in this regard. Any modified rule-based account would have to explain not only why regular plural modifiers are strongly dispreferred compared to singulars, but also why a variety of intermediate cases—irregular plurals, the pluralia tanta, and voicing change plural modifiers—are neither as acceptable as singulars nor as unacceptable as regular plurals.

We are not prepared to develop a second theory to account for these data, but we can address some likely directions that a modified dual-mechanism theory might take to capture these intermediate cases. One possibility would be to suggest that the intermediate cases are an artifact of averaging—half of the participants rated compounds like *mice-eater* as fully acceptable, while (perhaps due to performance constraints) the other half rated them as fully unacceptable. This possibility can be ruled out in the current case. For the irregular plurals in Study 2, the distribution of ratings across the scale (from 1 to 7) was 12, 18, 17, 15, 16, 14 and 8%—clearly there is no hint of a bimodal distribution.

However, similar arguments might be offered for other performance factors that could obscure the correspondence between the behavioral measures and the underlying competence. Clearly, such an account would require a principled explanation of exactly how performance factors result in the intermediate cases. To the extent that such an explanation starts to resemble the account we have proposed, it becomes questionable whether the original competence-based theory retains any explanatory power. More broadly, although it can always be argued that data concerning human performance do not accurately reflect an individual's underlying linguistic competence, this runs dangerously close to saying that the theory is not subject to behavioral test. There are no direct methods for observing linguistic competence; the available methods involve examining performance on tasks such as making grammaticality judgments. Grammaticality judgment data cannot be taken as primary evidence for competence on some occasions and discounted in other cases where the data do not conform to a particular theory (see Allen & Seidenberg, 1999; Schutze, 1996, for discussion).

A second general approach is to say that the intermediate cases reflect the joint effects of the two subsystems in the dual-mechanism theory, which yield conflicting output. It would then be necessary to explain why this conflict is produced more often for irregular forms than singulars, given that both are stored in the mental lexicon. Assuming that such an explanation could be devised, we could then ask whether a conflicting-routes approach is computationally feasible. Some hints in this regard come from several analogous claims that have been made for similar intermediate cases in the morphology of past-tense formation (Pinker, 1999; Prasada & Pinker, 1993) and in the processes in reading that map spelling to pronunciation (Coltheart et al., 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). In both of these cases, as in the ones considered here, the intermediate cases are words for which behavioral measures fall somewhere in between results for strictly rule-governed and exceptional items, and the challenge in each case is to explain how conflict or competition between the rule-governed and exceptional mechanisms could give rise to the intermediate cases. No dual-mechanism computational implementation has been offered for the past tense, but Coltheart and colleagues have made a serious attempt to implement a dual-mechanism account of reading. Their experience suggests that although it seems intuitively clear that conflicts between independent routes could produce intermediate cases, it is quite difficult to realize this approach in a computationally explicit way that maintains fidelity to the behavioral facts. In the most recent version of the dual-route theory of reading (Coltheart et al., 2001), the parameters of the model can be set to produce such effects, but doing so creates discrepancies with other aspects of the behavioral data (e.g., much higher error rates for irregular words than seen in people; longer latencies for words than nonwords).[7] Thus, it isn't sufficient to observe that a conflict between two subsystems could in principle underlie the intermediate effects; it is necessary to develop the idea to the point where it could be determined whether such a mechanism is compatible with all of the relevant behavioral data. Of course, the same is true of the account we have developed here; the basic theory can certainly be implemented in a way that fits the behavioral data, but it will be necessary to determine whether such a system is also consistent with other phenomena.

11.4. Conclusions

We have argued that the data concerning the occurrence of singular and plural noun modifiers, which have previously been taken to provide strong support for a rule-based account of morphological representation (Marcus et al., 1995; Pinker, 1991, 1999), instead are better captured by a theory in which well-formedness is a graded function of semantic and phonological constraints. These constraints reflect aspects of the structure of utterances to which the child is exposed in the course of

[7] Coltheart et al. (2001) attempt to account for the intermediate cases, called consistency effects by suggesting that they are an artifact of other properties of the stimuli. They are able to account for the results of one such study (Jared, 1997) but the same parameter settings yield poor simulations of many others (Seidenberg, Zevin, Harm, Plaut, & McClelland, 2002).

learning a language, extending well beyond the specific facts about compounding. This theory assumes that children bring powerful perceptual and cognitive capacities to the task of learning language, which involves encoding the statistical regularities present in the input. Which statistical regularities are encoded is determined by the nature of the input, properties of the learning procedure, and how knowledge is represented in memory (see Seidenberg, 1989; Seidenberg & MacDonald, 1999). These are general capacities rather than the language-specific types of knowledge proposed in theories such as level-ordering (see Saffran, 2002 for recent evidence concerning common mechanisms involved in linguistic and nonlinguistic statistical regularities). This analysis of the compounding data suggests an approach that could be applied to many other linguistic phenomena.

Acknowledgments

Authors are listed alphabetically. This research was supported by grants from the National Institutes of Mental Health (PO1 MH47566, K02 MH01188 and R01 MH58723) and the National Science Foundation (SBR95-11270). The authors thank Iris Berent, Carolyn Buck-Gengler, Peter Gordon, Lise Menn, Stephen Wilson and the attendees of the UCSD CRL Colloquium Series for useful comments and discussion, Melanie Blunschi and Julie Moffit for assistance collecting data, and two anonymous reviewers for helpful comments on the manuscript. Portions of this work were presented at the 40th Annual Meeting of the Psychonomic Society.

Appendix A. Experimental items for Study 2

	Experimental modifier	Control modifier	Head 1	Head 2
1	child	spouse	list	policy
2	foot	hand	check	wrapping
3	goose	duck	fund	pen
4	man	boy	patrol	players
5	mouse	rat	condos	observation
6	ox	horse	committee	folder
7	tooth	toe	examination	problems

Appendix B. Experimental items for Study 3

1. I last saw Jason in the trousers/jacket/jackets section
2. If you're going into the laundry room, make sure you don't mess up the stockings/ mitten/mittens pile
3. Put that back on the pliers/ hammer/hammers rack
4. I have no idea why this company sent me a binoculars/kaleidescope/kaleidescopes catalog

5. Don't forget to sign out whatever you take from the goggles/flipper/flippers locker
6. Let Dan give you some instruction on proper clippers/shovel/shovels use
7. Eva is taking classes on hairdressing, and they've been teaching her good shears/comb/combs technique.
8. The Adventure Society codes each of its suggested trips on a rigors/challenge/challenges index
9. It must be around here somewhere—Grandma always kept a tweezers/thimble/thimbles box
10. Everything around here has a place; put the tongs/trowel/trowels in the tongs/trowel/trowels place
11. My cousin works in a tights/sock/socks warehouse

Appendix C. Experimental items for Study 4

	Experimental modifier	Control modifier	Head 1	Head 2
1	knife	fork	auction	collection
2	loaf (of bread)	clove (of garlic)	bin	packer
3	shelf	drawer	division	manufacturer
4	thief	crook	buster	hangout
5	wharf	dock	area	authority
6	wife	mom	brigade	hotline
7	wolf	bear	association	ceremony

Appendix D. Experimental items for Study 6

1. The last paper I read by her was gunition/gunitions
2. I can't stand Nicole's pet hamster—it's totally bonner/bonners
3. Lauren had never been to a concert before, and she thought it was fant/fants
4. After reading the file, the judge concluded that the defendant was fethance/fethances
5. The cookie dough was sticky and glottity/glottities
6. Don't you think Miranda's new outfit is a bit kullery/kulleries
7. My dad finally bought a car that is hermect/hermects
8. Where did you get that cabinet from? It's so thoryun/thoryuns
9. Hans' coin collection contains some coins that are quite rare and very lond/londs
10. This creature's most striking feature is that it is mavitt/mavitts
11. After how the team played last year, I can't help being surprised that now they're so mippage/mippages
12. With the strange way she's always acting, I can't help but think she's nerdum/nerdums
13. I don't like tofu very much. I find it to be krare/krares
14. In a big city it is hard to find a neighborhood that is plivision/plivisions
15. These new toys look really pright/prights

16. With all the different food that was in the room, the air smelled zain/zains
17. The house was really old and looked tring/trings
18. They usually dance to music that is a bit more prore/prores
19. This fabric is so soft and smooth-It's completely naice/naices
20. The teacher should never have left the room unattended. When she returned it was brin/brins

References

Alegre, M. A., & Gordon, P. (1996a). Red rats eater exposes recursion in children's word formation. *Cognition, 60*, 65–82.

Alegre, M. A., Gordon, P. (1996b). *Why parks commissioners aren't rats eaters: Semantic constraints on regular plurals inside compounds*. Unpublished manuscript, Teacher's College, Columbia University.

Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.

Allen, M. R. (1978). *Morphological investigations*. Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.

Aronoff, M. H. (1974). *Word-structure*. Doctoral dissertation, Massachusetts Institute of Technology Cambridge, MA.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology, 43*, 83–128.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425–455.

Carlson-Luden, V. (1979). *Causal understanding in the 10-month old*. Unpublished doctoral dissertation, University of Colorado at Boulder.

Cassidy, K. W., Kelly, M. H., & Sharoni, L. J. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General, 128*, 362–381.

Chapman, C. (1996). Perceptual salience and affix order: Noun plurals as input to word formation. In G. Booij, & J. van Marle (Eds.), *Yearbook of morphology, 1995*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study. *Behavioral and Brain Sciences, 22*, 991–1060.

Clahsen, H., Marcus, G., Bartke, S., & Wiese, R. (1996). Compounding and inflection in German child language. In G. Booij, & J. vanMarle (Eds.), *Yearbook of morphology1995*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition, 45*, 225–255.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory & Cognition, 6*, 391–402.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 1355–1367.

DiSciullo, A.-M., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fabb, N. (1988). English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory, 6*, 527–539.

Francis, W. N. (1964). *A standard sample of present-day. English for use with digital computers. Report to the US Office of Education on Cooperative Research No. E–007*. Providence, RI: Brown University.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of english usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

Gleason, J. B., Perlmann, R. Y., & Greif, E. B. (1984). What's the magic word? Learning language through routines. *Discourse Processes, 6*, 493–502.

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

Grice, H. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3). New York: Academic Press.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in english. *Language, 65*, 203–257.

Halle, M. (1990). Phonology. In D. N. Osherson, & H. Lasnik (Eds.), *Language: An invitation to cognitive science* (Vol. 1). Cambridge, MA: MIT Press.

Harm, M. W. (1998). *Division of labor in a computational model of visual word recognition*. Doctoral Dissertation, University of Southern California, Los Angeles, CA.

Harm, M. W., Seidenberg, M. S. (in press). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36*, 505–529.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, 96*, 7592–7597.

Joseph, B. D., & Philippaki-Warburton, I. (1987). *Modern Greek*. London: Croom Helm.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27*, 343–358.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349–364.

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

Kopcke, K.-M. (1998). The acquisition of plural marking in English and German revisited: schemata versus rules. *Journal of Child Language, 25*, 293–319.

Lieber, R. (1988). Phrasal compounds in English and the morphology–syntax interface. In *Papers from the parasession on agreement in grammatical theory, 24th annual meeting of the Chicago Linguistics Society* (pp. 202–222). Chicago, IL: University of Chicago.

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Erlbaum.

Marchman, V. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science, 21*, 283–304.

Marcus, G. F., Brinkman, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29*, 189–256.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*, 313–330.

Masur, E., & Gleason, J. B. (1980). Parent–child interaction and the acquisition of lexical information during play. *Developmental Psychology, 16*, 404–409.

Mintz, T. H., Newport, E. L., Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science* 26, TBA.

Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686–713.

Ninio, A., Snow, C., Pan, B., & Rollins, P. (1994). Classifying communicative acts in children's interactions. *Journal of Communicative Disorders, 27*, 157–188.

Patterson, K., Lambon-Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia, 39*, 709–724.

Pinker, S. (1991). Rules of language. In P. Bloom (Ed.), *Language acquisition: Core readings*. New York: Harvester Wheatsheaf.

Pinker, S. (1994). *The language instinct: How the mind creates language?*. New York: William Morrow and Company.

Pinker, S. (1999). *Words and rules*. New York: Basic Books.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker, & J. Mehler (Eds.), *Connections and symbols*. Cambridge, MA: A Bradford Book: The MIT Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38*, 43–102.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*, 1–56.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*, 425–469.

Rumelhart, D. E., McClelland, J. L. (1986). On learning the past tenses of English verbs. In: McClelland, J. L., Rumelhart, D. E., the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 2. MIT Press, Cambridge, MA.

Rumelhart, D. E., McClelland, J. L., The PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1. MIT Press, Cambridge, MA.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Schutze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.

Seidenberg, M. S. (1989). Visual word recognition and pronunciation: A computational mode and its implications. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569–588.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of visual word recognition and naming. *Psychological Review, 96*, 523–568.

Seidenberg, M. S., Zevin, J. D., Harm, M. W., Plaut, D. C., McClelland, J. L. (2002). Modeling as data fitting: Comments on Coltheart et al. (2001). Manuscript in preparation.

Selkirk, E. O. (1982). *The syntax of words*. Cambridge, MA: MIT Press.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language, 25*, 169–201.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11–B21.

Siegel, D. C. (1974). *Topics in English morphology*. Doctoral Dissertation, Massachusetts Institute of Technology Cambridge, MA.

Spencer, A. (1991). *Morphological theory: An introduction to word structure in generative grammar*. Cambrige, MA: Blackwell.

Ullman, M. T., Corkin, S., Coppola, M., Hickock, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declaritive memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 266–276.