The Nature and Acquisition of Lexical and Functional Categories

Karin Stromswold
Psychology Department & Center for Cognitive Science
Rutgers University

ABSTRACT

In this paper, I will argue (i) that the distinction between lexical and functional morphemes is universal; (ii) that the distinction between lexical and functional morphemes is psychologically and biologically real and not merely descriptive; and (iii) that lexical and functional morphemes are members of fundamentally different kinds of categories and that these differences are part of children's biological endowment.

This work was supported by a grant from the John Merck Foundation, a Ph.D. fellowship from the National Science Foundation, and an M.D./Ph.D. fellowship from the John D. and Catherine T. MacArthur Foundation. Correspondence may be sent to Karin Stromswold, Dept. of Psychology & Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08903 or to karin@ruccs.rutgers.edu.
The Nature and Acquisition of Lexical and Functional Categories

All fully-developed natural languages appear to have both lexical morphemes (e.g., nouns, verbs, adjectives) and functional morphemes (e.g., agreement and tense markers, determiners, etc.) (Greenberg, 1966; Comrie, 1981; Croft, 1990). In fact, one of the most salient differences between pidgin languages and creole languages is that pidgins generally lack functional morphemes whereas creoles have functional morphemes (for further discussion, see Bickerton, 1981; Romaine, 1988).

Certain cross-linguistic tendencies can be observed about lexical and functional categories. Across languages, members of functional categories are used to express certain semantic and syntactic distinctions and the number of such distinctions is finite and fixed both across languages and within a language. For example, cross-linguistically, members of functional categories often express information about gender, number, tense, aspect, evidential status, definiteness, affectedness, etc.. Cross-linguistically, functional morphemes can be free morphemes (e.g., the English auxiliaries) or they can be simple or complex affixes (e.g., the English past tense suffix -ed). Cross-linguistically, there is often a correspondence between certain lexical and functional elements. If members of functional categories are instantiated as affixes, the affixes that are semantically associated with nominal elements (e.g., number, gender, definiteness, etc.) tend to be bound to nouns and the affixes semantically associated with verbal elements (e.g., tense, aspect, durativeness, etc.) tend to be attached to verbs (Croft, 1990). If members of functional categories are instantiated as free morphemes, there is often a lexical category that corresponds to the functional categories (e.g., nouns and pronouns, lexical verbs and auxiliary verbs, adjectives and determiners/classifiers, etc.). Cross-linguistically and within a language, there are often restrictions placed on members of functional categories that are not placed on members of lexical categories. For example, pronouns require certain types of antecedents that lexical nouns do not require.

Evidence from language acquisition, language use, and language pathology suggests that the distinction between lexical and functional categories is cognitively and biologically real, and not merely a descriptive distinction. In the area of language acquisition, first language learners (Brown & Fraser, 1963; Brown, 1973; Gleitman & Wanner, 1982; Egido, 1983; Guilfoyle & Noonan, 1988; Lebeaux, 1988; Gerken, Landau, & Remez, 1990; Radford, 1990) and second language learners (Johnson & Newport, 1986; Morgan, Meier & Newport, 1989) appear to have more difficulty producing, processing, and imitating functional morphemes than lexical morphemes. In language production, semantic substitutions, sound exchanges, and word exchanges errors occur only between lexical elements, whereas functional elements participate in shift errors (Garrett, 1976; 1980; 1990). In language perception, functional morphemes have been argued to play a crucial role in syntactic parsing (Kimball, 1973; Wanner & Maratsos, 1978; Morgan & Newport, 1981). Lastly, preliminary data suggests that the event-related potentials (ERPs) associated with reading functional words are different than the ERPs associated with reading lexical words (Neville, 1991).

In addition to these results from normal adults and children, there is some evidence of a double dissociation between impairment with lexical morphemes and functional morphemes in acquired and developmental language disorders. Patients with Wernicke's
jargon aphasia reportedly understand and use function morphemes relatively well despite gross impairment with lexical words (Lecours, 1982; Buckingham, 1980), whereas agrammatic Broca's aphasics reportedly produce and comprehend lexical words significantly better than functional words (Goodglass, 1976; Saffan, Schwartz, & Marin, 1980). Furthermore, whereas normal subjects and Wernicke's aphasics are more likely to detect a target letter when it appears in an lexical word than when it appears in a functional word, Broca's aphasics are just as likely to detect the target letter in functors as in lexical words (Swinney, Zurif & Cutler, 1980; Friederici, 1983; Rosenberg et al., 1985).

Children with developmental dysphasia (also known as Specific Language Impairment) appear to have a relatively selective difficulty with functional elements (see, for example, Morehead & Ingram, 1973; Kershensteiner & Huber, 1976; Clausen, 1989, 1991; Gopnik, 1990, 1991; Leonard, 1989; Guilfoyle, Allen & Moss, 1991; Rice & Oetting, 1991), whereas some studies suggest that the speech of hydrocephalic children and Turner's syndrome females is relatively intact syntactically and markedly impaired semantically (see Yamada, 1990).

One fundamental distinction between lexical and functional categories is that lexical categories freely and regularly admit new members, whereas functor categories do not. Consider, for example, the enormous number of technology-related lexical nouns and verbs that have been introduced into mainstream English in the past century. During this same time, not a single preposition, auxiliary verb, pronoun, determiner, inflectional ending, etc. has been added to English.

Synchronically, the closed membership of functional categories seems to be maintained in acquired language disorders, in speech errors, and, perhaps, in language acquisition. The neologisms that jargon aphasics produce as a result of brain damage (Lecours, 1982) and that healthy adults produce as the result of production errors (Garrett, 1976; 1980; 1990) are confined to lexical categories. There are no reported examples of a normal or brain-injured adult inventing a new closed-class element. The same appears to be true for children. Although children are willing to invent novel words to describe novel actions, objects, and attributes (for examples from English, see Bowerman, 1974 and Pinker, 1989)\(^1\), I know of no examples in the acquisitional literature of children inventing a new member of an existing closed-class category (e.g., inventing a novel agreement marker for third person plural present tense or a novel masculine plural pronoun) or inventing a novel category (e.g., inventing a classifier system).

The lack of reported examples of functional category neologisms might reflect the fact that children do not invent members of functional categories. Alternatively, the lack of reported functional neologisms might just be the result of researchers not studying, recognizing, or recording functional neologisms. The spontaneous-speech transcripts of

---

\(^1\) However, as Pinker (1989) correctly observes, although children are willing to invent novel lexical words in spontaneous and elicited speech, the vast majority of the time, children use known words (see below). For the purposes of this paper, however, the important point is that children occasionally produce novel lexical words, but they never produce novel functional words.
six children were examined to determine the relative frequency of lexical and functional neologisms.

**Experiment**

**Corpora**

I analyzed the CHILDES spontaneous-speech transcripts (MacWhinney & Snow, 1985, 1990) of the six children listed in Table 1. The children ranged in age from 1;2 to 2;5 at the time when the first transcript was recorded and from 2;3 and 5;2 when the last transcript was recorded. These children's transcripts were analyzed because for each of these six children, the CHILDES transcripts were marked such that words that were invented by the child were tagged with character string "@c" (e.g., yesternight@c) and regularized inflectional forms were tagged with the character string "@n" (e.g., eated@n).

**Table 1: Transcripts Analysed**

<table>
<thead>
<tr>
<th>Child</th>
<th>Corpus collected by</th>
<th>Ages</th>
<th># of Words (Tokens)</th>
<th># of Words (Types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abe</td>
<td>Kuczaj (1976)</td>
<td>2;5-5;0</td>
<td>162,367</td>
<td>4,734</td>
</tr>
<tr>
<td>Adam</td>
<td>Brown (1973)</td>
<td>2;3-5;2</td>
<td>164,231</td>
<td>4,061</td>
</tr>
<tr>
<td>Eve</td>
<td>Brown (1973)</td>
<td>1;6-2;3</td>
<td>35,170</td>
<td>1,737</td>
</tr>
<tr>
<td>Naomi</td>
<td>Sachs (1983)</td>
<td>1;2-4;9</td>
<td>44,809</td>
<td>2,625</td>
</tr>
<tr>
<td>Peter</td>
<td>Bloom (1973)</td>
<td>1;10-3;2</td>
<td>91,831</td>
<td>2,338</td>
</tr>
<tr>
<td>Sarah</td>
<td>Brown (1973)</td>
<td>2;3-5;1</td>
<td>107,314</td>
<td>4,085</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td>605,722</td>
<td>19,580</td>
</tr>
</tbody>
</table>

**Method**

For each child, the CLAN word frequency program"freq" (MacWhinney & Snow, 1990) was used to create a list of all of the distinct words uttered by a child in his or her transcripts and to determine the total number of times each distinct word was said by the child. The number of different words (word types) said by the children ranged from 1,737 to 4,734 with a mean of 3,263.3 types. The number of words (tokens) said by the child ranged from 35,170 to 164,231, with a mean of 100,953.7 word tokens (see Table 1).

The meaning and categorical status (lexical or functional) of each word that was tagged with a "@c" or a "@n" was determined by checking the context in which the purported neologism was uttered. Each word tagged with a "@c" was classified as being

---

2 Rather than determining the frequency of words, per se, the CLAN frequency program determines the frequency of letter strings in the transcripts. A letter string is defined as a string of letters bounded on either side by either a space or a punctuation mark (MacWhinney, 1991). For example, freq counts the words *eat* and *eats* as two distinct words and the interjectives *ugh, uggh*, and *ughh* as three different words. For this reason, freq almost certainly inflates the token counts for the children. Token and type counts, therefore, should be considered to be approximate.
either a lexical neologism or a functional neologism. Words with regularized inflections (i.e. words tagged with "@n") were also classified as members of lexical or functional categories. The frequency with which each child used the inflections -s, -ed, -ing, ly, and -'s was determined by using the UNIX utility 'fgrep' to cull all of the the words in each child's CLAN-generated word-frequency list that ended in s, ed, ing, ly, or 's. The resulting list was searched by hand to eliminate any words that did not have an inflection (e.g., the words gas, bed, only, etc.).

The frequency with which each child used functor words was determined by using ffgrep to cull all of the lines in each child's word-frequency list that contained a functor. The resulting list of potential functors was searched by hand to eliminate any words that were not functors (e.g., standby, miss, etc.). In addition, the resulting list of functors was searched for examples of improperly inflected functors (e.g., many's) or functional neologisms that were not marked with "@c" or "@n" as being novel forms. Because each functor on the word list was examined by hand to determine whether the example represented a novel form, whereas only lexical words tagged with @n or @c were examined by hand, it is more likely that a novel lexical form was missed than a novel functor form was missed.

Results

As is shown in Table 2, the children said between 22,233 and 99,626 lexical words (mean = 58,487.8 lexical words), of which between 164 and 473 were lexical neologisms (mean = 233.7 lexical neologisms). Thus, between .20% and .85% (mean = .48%) of the children's lexical words were neologisms. The children said between 12,937 and 78,508

---

3 The following possible functors were searched for:

a, about, above, according, across, after, against, all, along, although, amid, among, an, and, another, any, anybody, anyone, anything, around, as, at, because, before, behind, below, beside, between, beyond, both, but, by, despite, down, during, each, either, even, ever, every, everybody, everyone, everything, except, few, for, from, he, her, here, hers, herself, him, himself, his, how, howabout, how, however, i, if, in, inside, into, it, its, itself, lest, many, me, mine, minus, my, myself, near, nearer, neither, next, no, no-one, nobody, none, noone, nor, nothing, of, off, on, once, oneself, onto, or, our, ours, out, out, over, ownself, rather, she, since, so, some, somebody, someone, something, than, that, the, their, theirs, them, then, there, these, they, this, those, though, through, throughout, thru, till, to, toward, under, underneath, unless, until, up, upon, us, versus, we, what, whatever, whatsoever, when, whenever, where, whereby, wherefore, wherein, whereof, whereon, wherever, whether, which, whichever, while, who, whoever, whom, whose, whosoever, whosoever, why, with, within, without, y'all, yet, you, your, yours, yourself.

4 Because it is possible that not all neologisms were tagged, that some purported neologisms were actually just variant pronunciations, or that some examples tagged as examples of "word play" were actually neologisms, the numbers cited should be considered rough estimates of the frequency with which children invent new words. This is particularly true for lexical neologisms.
functors (mean = 42,465.8 functors) and no child ever invented a novel bound or free functor.5

Table 2: Frequency of Lexical and Functor Neologisms

<table>
<thead>
<tr>
<th>Child</th>
<th># of Lexical</th>
<th># Lexical Neologism</th>
<th>% Lexical Neologisms</th>
<th># of Functors (Tokens)</th>
<th># Functor Neologisms</th>
<th>% Functor Neologisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abe</td>
<td>83,859</td>
<td>164</td>
<td>.20%</td>
<td>78,508</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Adam</td>
<td>99,626</td>
<td>473</td>
<td>.47%</td>
<td>64,605</td>
<td>?.1 (an't)</td>
<td>?.0015%</td>
</tr>
<tr>
<td>Eve</td>
<td>22,233</td>
<td>188</td>
<td>.85%</td>
<td>12,937</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Naomi</td>
<td>26,339</td>
<td>175</td>
<td>.66%</td>
<td>18,470</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Peter</td>
<td>54,093</td>
<td>238</td>
<td>.44%</td>
<td>37,738</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Sarah</td>
<td>64,777</td>
<td>164</td>
<td>.25%</td>
<td>42,537</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Overall</td>
<td>350,927</td>
<td>1,402</td>
<td>.40%</td>
<td>254,795</td>
<td>?1</td>
<td>?0004%</td>
</tr>
</tbody>
</table>

As is shown in Table 3, the children used the inflections -s, -ed, -ing, ly, and 's between 1,861 times and 13,762 times (mean = 3,965.0 uses) and between .63% and 2.68% (mean = 1.49%) of these were novel inflected lexical forms (e.g., *eated for ate). 1 Interestingly, even though the children regularized lexical be, do, and have (saying "bes", "doos", "haves", "beed", "doed", and "hadded"), the children never regularized homophonic auxiliary be, do, and have (see Stromswold, 1990, 1992).

The six children used a total of 119 possible novelly-inflected functors. Ninety-six of the 119 examples (81%) were instances of *mine's (Adam said 82 examples, Eve said 13 examples, and Sarah said one example.) However, most of the examples of *mine's were probably examples of mine + contracted copula is (e.g., mine's over here), rather than examples of mine + possessive 's. Thus most of the examples of *mine's probably were not examples of children regularizing the possessive inflection. Although it is notable that Adam and Eve may have occasionally said "mine's" for "mine," perhaps even more notable is the fact that none of the children ever used I's, my's, me's, you's, he's, him's, she's, hers's, we's, ours's, they's or them's as possessive forms as we might have expected if they had generalized what they know about forming possessive forms of common nouns and proper nouns to pronouns. Clearly the children distinguished between members pronouns and lexical nouns.

Of the remaining 23 examples, Eve said "I be manys" nine times in a row and "somes" three times over the course of 2 months ("let me have somes," "Mom'll go get somes," and "somes ... some carts 'a' and 'l'"). Notice that in all of these cases, many and some were used as plural nouns and not as quantifiers. Five of the remaining 11 examples were verb particles with verbal inflections (where froms huh?, he come ins ... he come in, cow stands upped ... but the horse not stand ups, I go downed). It is possible that the children thought that these verb particles were actually part of the verb (e.g., that in he come ins, come in = comein ). The last six examples included Peter saying "theres" 5

---

5 The one possible exception ("it's magic, an't I.") was said by Adam at 4;7. However, from context, "an't" appears to be a typographical error for ain't.
twice in a row ("what's on theres? can't go therers") and "no's" once ("no's juice), Sarah saying "thems" once ("I want buy two thems") and "outs" once ("that's outs"), and Abe saying "don'ts" once ("sometimes he's not friendly sometimes he barks sometimes he don'ts").

Table 3: Frequency of Lexical and Functional Words with Novel Inflections

<table>
<thead>
<tr>
<th>Child</th>
<th># inflections</th>
<th># Novelly Inflected Lexical Words [type]</th>
<th>% Novel Inflected Lexical</th>
<th># Functors with Novel Inflections. [type]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abe</td>
<td>13,522</td>
<td>363 [31]</td>
<td>2.68%</td>
<td>1 [1 don'ts]</td>
</tr>
<tr>
<td>Adam</td>
<td>13,762</td>
<td>109 [64]</td>
<td>.79%</td>
<td>83 [82 mine's, 1 froms]</td>
</tr>
<tr>
<td>Eve</td>
<td>1,861</td>
<td>27 [14]</td>
<td>1.45%</td>
<td>25 [13 mine's, 9 manys, 3 somes]</td>
</tr>
<tr>
<td>Naomi</td>
<td>5,089</td>
<td>94 [44]</td>
<td>1.85%</td>
<td>0</td>
</tr>
<tr>
<td>Peter</td>
<td>8,851</td>
<td>56 [30]</td>
<td>.63%</td>
<td>5 [2 theres, 1 ins, 1 ups, 1 upped, 1 no's]</td>
</tr>
<tr>
<td>Sarah</td>
<td>7,749</td>
<td>117 [61]</td>
<td>1.51%</td>
<td>4 [1 thems, 1 mine's, 1 outs, 1 downed]</td>
</tr>
<tr>
<td>Overall</td>
<td>50,834</td>
<td>766 [244]</td>
<td>1.51%</td>
<td>118</td>
</tr>
</tbody>
</table>

In summary, approximately 1.5% of children's inflections were novel uses with lexical words, whereas less than.05% of children's inflections were possible novel uses with functors. Given that 42% of children's free morphemes were functors, this is quite striking. What is also striking is how similar the potential examples of inflected functors were to one another. The majority of the examples could be examples of a functor with a cliticized contracted copula *is* and not true examples of inflected functors. Thus, the analyses suggest that children formed novel inflected lexical words about 30 times more frequently than they formed novel inflected functors.

Discussion

The results of these transcript analyses indicate that children invent members of lexical categories but they never invent members of functional categories. This is quite striking because it is unclear how a child who has only begun to learn the members of both open and closed-categories would realize that the membership of some categories is potentially infinite whereas the membership of other categories is finite, if this knowledge were not part of the child's innate endowment. Merely knowing that natural languages contain categories with finite membership and categories with infinite membership does not insure that children will not invent functional neologisms. Children

---

6 Rather than being illicit plurals, *theres* may actually be an example of *there* + contracted copula *is*, *outs* might actually be *out* + contracted copula *is*, and *no's* might be a contracted form of *no is juice*. 

---
must also have innate knowledge of the universal properties of lexical categories and functional categories if they are to correctly identify categories that have infinite membership (i.e., N, V, Adj) categories that have finite membership (e.g., tense, agreement, determiners, etc.).

Perhaps the reason languages have lexical categories that have open-membership and functor categories that have closed-membership has to do with the function of lexical and functional morphemes. Lexical morphemes express bare predicates and arguments. As such, membership in lexical categories must be expandable in order that when new predicates and arguments appear in the world, they can easily be expressed in natural languages (e.g., technology-related objects and actions). There seem to be two types of closed-class morphemes in languages. The first type include functional morphemes that modulate and modify the meanings and relationships between arguments and predicates. Examples would include tense, modality, agreement, and case-markers. The second type of closed-class morphemes are pro-forms such as the pronouns, question words (who, what, when), pro-locations (here, there), pro-temporal terms (now, then), quantifiers (some, many, all), etc. Intuitively, closed-class pro-forms seem more like open-class categories than do closed-class modulators. Interestingly, the few functors that children inflected in novel ways (e.g., manys, somes, theres, etc.) were usually cases of pro-forms that were filling the role of an argument rather than acting as a modifier.

If functional categories are hard-wired and fixed, and they are not invented by children or adults, how do they come to be? One possible answer is that they come from members of lexical categories the meaning of which have become grammaticalized (for one treatment, see Lightfoot, 1981, 1991). For example, it has been argued that tense and modality markers tend to evolve from lexical verbs that have relatively little semantic content. Diachronically, such verbs increasingly act as closed-class pro-form and eventually become a -class modulators (see Harris & Ramat, 1987 and Heny & Richards, 1983).

The apparent double-dissociation between lexical and functional categories in acquired and developmental language disorders suggests that the biological substrate for the two types of categories are distinct. In addition, it seems plausible that the optimal architecture for lexical categories would be different from the optimal architecture for functional categories. What kind of cognitive architecture might best instantiate lexical and functional categories? What kind of architecture would allow for lexical neologisms but not functional neologisms? One possibility is that separate modules with fundamentally different architectures are used to represent members of functional and lexical categories. Members of functional categories might be represented via distinct nodes with few, if any, connections between members, whereas members of lexical categories might be represented by a network that has many more connections between nodes (Jacobs et al, 1991).

To conclude, children's willingness to invent lexical morphemes but not functional morphemes suggests that children have innate knowledge of the existence of finite categories and infinite categories of words. This innate knowledge lets them identify lexical categories as potentially having infinite membership and functional categories as having finite membership. They are willing to generalize what they know about one member of a lexical category to other members of that lexical category. The result is lexical neologisms. They approach the learning of functional categories conservatively
and do not generalize what they know about the behavior of one member of a functional
category to another member of the same functional category. The result is they do not
produce functional neologisms. The differences in children's approaches to lexical and
functional categories may result from or reflect differences in the kinds of the
architectures that subserve these two types of categories.
References


Stromswold (1994), Lexical and Functional Categories

Cambridge, MA: MIT Press.


