Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective*

RUSHEN SHI,

JAMES L. MORGAN

AND

PAUL ALLOPENNA

Brown University

(Received 16 August 1996. Revised 17 August 1997)

ABSTRACT

Maternal infant-directed speech in Mandarin Chinese and Turkish (two mother–child dyads each; ages of children between 0;11 and 1;8) was examined to see if cues exist in input that might assist infants’ assignment of words to lexical and functional item categories. Distributional, phonological, and acoustic measures were analysed. In each language, lexical and functional items (i.e. syllabic morphemes) differed significantly on numerous measures. Despite differences in mean values between categories, distributions of values typically displayed substantial overlap. However, simulations with self-organizing neural networks supported the conclusion that although individual dimensions had low cue validity, in each language multidimensional constellations of presyntactic cues are sufficient to guide assignment of words to rudimentary grammatical categories.

INTRODUCTION

Learning how to assign words to grammatical categories, or form classes, is an important step in early language acquisition. At least partial success on

[*] Material in this article formed part of a doctoral dissertation submitted by the first author to the Department of Cognitive and Linguistic Sciences, Brown University. This work was supported in part by NIH grant HD 29426 to J. Morgan. We thank Katherine Demuth and Sheila Blumstein for helpful comments on earlier versions of this article; Robert Brandenberger, John Mertus, and Bin Tang for providing technical and programming support, and Huseyin Tek for Turkish transcriptions. Address for correspondence: Rushen Shi, Department of Psychology, University of British Columbia, 2509-2136 W. Mall, Vancouver, BC V6T 1Z4, Canada. e-mail: Rushen@cortex.psych.ubc.ca.
this task is necessary for syntactic analysis to proceed. This is because the syntactic patterns that govern how words may be combined to form sentences are specified not in terms of individual words but rather in terms of categories of words. In mature grammars, grammatical categories are defined in distributional terms: a given word is a noun, for example, if it shares privileges of substitution and equivalence with other nouns (Harris, 1954). Thus, *justice* is a noun not because it shares semantic or phonological characteristics with other nouns like *dog* or *metallurgy*, but because it can occur in the same sets of grammatical contexts as can other nouns. However, these grammatical contexts are defined in terms of other grammatical categories, such as *determiner* or *verb*. In turn, the grammatical contexts in which determiners and verbs can occur are defined in part with reference to nouns. Because categories are defined in terms of other categories in a perfectly circular system, infants require some independent means for breaking into this system. Some form of bootstrapping must provide infants with reasonably accurate initial guesses about the category membership of at least a few words.

Several bootstrapping schemes have been offered in the language acquisition literature. Maratsos & Chalkley (1980) argued that children could induce grammatical categories by constructing networks of correlations among words, morphemes, and meanings; a similar account has been proposed by Levy (1988). Pinker (1984) proposed a semantic bootstrapping procedure by which children could initially assign words referring to concrete objects to the noun class, words referring to actions to the verb class, and so forth; other semantically based procedures have been proposed by Braine (1976) and Schlesinger (1971). Without exception, all these proposals tacitly assume that children can extract and represent individual words (or morphemes) in the input speech stream. On all accounts, however, for naïve learners, segmenting words from continuous speech requires sophisticated perceptual and computational capacities (Christophe, Dupoux & Mehler, 1993; Jusczyk & Aslin, 1995; Morgan & Saffran, 1995). The existence of such capacities, poorly understood though they yet are, raises the possibility that perceptual analyses of input might yield information on aspects of linguistic structure beyond identification of words, such as grammatical category membership of words. This in turn depends upon the fashion and degree to which information pertaining to grammatical category membership is encoded in input.

The traditional view of structuralist linguistics and its descendants (Saussure, 1916/1959; Chomsky, 1957; Hockett, 1966) is that grammatical categories and phonology constitute independent levels of structure; certainly, grammatical category membership is not reducible to any set of phonological properties. Nevertheless, recent work by Kelly (1992) and Cassidy & Kelly (1991) has shown that form classes may indeed have
correlated, partly predictive phonological properties. Cassidy & Kelly (1991) found that in both adult- and child-directed English speech verbs contain fewer syllables than nouns. Kelly (1992) also noted that there are tendencies in English for disyllabic nouns to be stressed on the first syllable and disyllabic verbs on the second syllable, for nouns to have greater duration than verbs, for nouns to have more complex syllables than verbs, and for nouns to be more likely to have low vowels and nasal consonants than verbs. Behavioural studies reported by Kelly (1992) suggest that adults and children as young as four years old may make use of these properties in identifying novel words as nouns versus verbs.

In this article, we inquire whether information in infant-directed speech may support sorting of words (or syllabic morphemes) into sets corresponding to functional and lexical classes. In part, our goal is to provide explicit, cross-linguistic evidence in support of psycholinguistic intuitions that these two superordinate grammatical categories may be distinguished in language acquisition and processing by their phonological properties (Kean, 1979; Gleitman & Wanner, 1982; Cutler, 1993). Functional items include auxiliary verbs, case markers, complementizers, conjunctions, determiners, (at least some) prepositions, and pro-forms; lexical items include nouns, verbs, adjectives, and adverbs. Studies of theoretical linguistics, language acquisition, normal adult language processing, and language loss all indicate that these classes form a basic division of grammatical categories (Brown, 1973; Rosenberg, Zuriff, Brownwell, Garrett & Bradley, 1985; Abney, 1987; Radford, 1990). To note but a few distinctions, functional items are members of closed classes, whereas lexical items are members of open classes. Functional items tend to occur very frequently, but many lexical items occur infrequently. Functional and lexical items display clearly different acquisition patterns (though the reasons for this remain in dispute): children’s early productions were historically characterized as ‘telegraphic’ precisely because they tend to include (strong) syllables from lexical items, but to exclude functional items.

Recent analyses of functional and lexical items in English infant-directed speech (Shi, 1995) show that multiple partly predictive cues are available in input. These include syllable duration, syllable structure, vowel quality, relative syllable amplitude, and position in utterance. In English, functional items tend to have short durations, syllables with null or minimal onsets and codas, centralized vowels, low relative amplitude, and are likely to occur in utterance-initial position; lexical items tend to have long durations, syllables with complex onsets or codas, peripheral vowels, high relative amplitude, and are likely to occur in utterance-final position. None of these individually is a highly valid cue to category membership: on all measures, the distributions of values for functional and lexical items show substantial overlap. Nevertheless, taken as
a set, these cues are sufficient in principle to support assignment of words to functional and lexical classes with 85–90 per cent accuracy, as shown in modelling with self-organizing neural networks.

If these findings hold true only of English, they are of limited interest, but if they hold true across diverse languages, this would increase the plausibility of the hypothesis that phonology may indeed provide early bootstraps into certain aspects of syntax. Accordingly, in the research reported here, we inquired how the distinction between functional and lexical items is manifested in infant-directed speech in two languages quite different from English — Mandarin Chinese and Turkish. Morphologically, Mandarin is an isolating language, whereas Turkish is an agglutinative language; both are typologically distinct from English.

Our analyses of possible phonological predictors of basic grammatical category membership proceed from the hypothesis that functional items, unlike lexical items, universally tend to be productively and perceptually minimal. Several factors conspire to make this so. On one hand, functional items tend to have high frequencies and low morphological complexity so that speakers may be inclined to expend minimal effort in producing them (Zipf, 1949). On the other hand, functional items tend to carry low semantic load and to be highly predictable from syntactic and semantic context so that listeners’ comprehension of discourse will be unimpeded by their minimization.

Certain manifestations of minimality may appear consistently across languages, whereas others may vary from language to language. We hypothesize that functional item syllables consistently tend to have simple syllable structures (lacking onsets, codas, or both, and lacking consonant clusters in either position), simple, rather than diphthongized vowels, brief duration, low amplitude, and minimal F0 change. Functional item syllables also tend to include unmarked or underlyingly underspecified phonemes. How this last property is manifested, however, varies across languages: in English, functional item vowels tend to be reduced or centralized; in Mandarin, functional items tend to have neutral, rather than specified, tone; in Turkish, vowels of functional items harmonize with the final vowel of the stem to which they are affixed. In addition, functional items tend to have minimal numbers of syllables, and functional items and lexical items tend to occur at opposite ends of utterances (in English, functional items are more likely to occur utterance-initially, whereas lexical items are more likely to occur utterance-finally; the opposite is true in Turkish). It is our hypothesis that the perceptual basis for distinguishing functional items from lexical items does not reside in any of these individual properties, but rather, in any language, resides in the entire constellation of properties taken simultaneously.

Here, we provide evidence bearing on the plausibility of sorting words into
sets corresponding to rudimentary grammatical categories on the basis of complexes of phonological measures. In principle, learners might be able to perform such sorts at a point when little or no progress had yet been achieved in analysing the syntax or semantics of the language being learned. We do not provide evidence here on whether infants can use correlated phonological properties to assign words to rudimentary grammatical categories; indeed, whether they can do so is unknown at this time. We return to this point in the Discussion, where we will argue that the capacities demonstrably required for solving the word-level segmentation problem also suffice for solving the category assignment problem at the rudimentary level with which we are concerned.

STUDY 1

Mandarin is a Sino-Tibetan language, the official language of mainland China and Taiwan. Basic word order is SVO with fixed order of indirect object and time and place adverbials, although object preposing commonly produces SOV or OSV orders (Li & Thompson, 1981). The fact that modifiers mostly precede their heads shows that the language is left branching. Null subjects are allowed. Mandarin is a lexical tone language and is morphologically isolating (i.e. lacks grammatical inflections). Many, though by no means all words are monosyllabic, and most syllables are open (only certain nasals are permitted in syllable codas). Lexical item categories include nouns, verbs, adjectives (attributive and verb-like predicative), and adverbs. Functional item categories include demonstratives, copulas, modals, prepositions, postpositions, question words, noun classifiers, conjunctions, verb and noun particles, and sentence-final particles (which encode pragmatic intent). Some of these categories are not found in many other languages. For example, noun classifiers and sentence final particles do not exist in English. Overviews of Mandarin grammar and acquisition may be found in Erbaugh (1992).

METHOD

Subjects

Two mother–child dyads participated in this study. Both mothers came from Taiwan and were primary caretakers of their infants. According to parental reports, the language spoken at home was solely Taiwan Mandarin. The two dyads were recorded when the children (one boy and one girl) were 1;0 ± 29 and 1;0 ± 24 respectively. Neither child was yet producing words.

[1] Although we use the term ‘categories’ to refer to such sets, note that we do not claim that learners can induce either category labels or syntactic properties associated with grammatical categories from phonological analyses of input.
Recording and transcription
Recordings were made in a sound-proof laboratory playroom equipped with quiet toys and books. Each mother–child dyad participated in a 40-minute recording session. The mother was asked to speak to her child naturally while the dyads played in the playroom by themselves.

Audio recordings were made on a Panasonic 3700 digital audio tape recorder, using a Nady 101 VHF wireless receiver and transmitter, and a Lavaliere microphone. The microphone was clipped to the mother’s collar. The transmitter was small enough to be put into her pocket. These procedures and recording environment made possible high-quality recordings and allowed the subjects to move around in the room easily.

Recordings were transcribed by R.S., using standard Pinyin romanization (a broad phonemic coding). Each word was coded for its grammatical category and each syllable for its tonal category. Coding criteria were based on Cheng (1973), Barale, Harvey & Madden (1980), Li & Thompson (1981), and the most authoritative and widely used dictionary Xiandai Hanyu Cidian (Modern Chinese Dictionary).

Measures
Five per cent of the mother’s words from each transcript were selected at random for analysis. These samples included 98 words for Mother M1 (67 lexical items and 31 functional items), and 77 words for Mother M2 (49 lexical items and 28 functional items). A variety of distributional, phonological, and acoustic measures was computed on the words in these samples. Measures included TYPE FREQUENCY, ROUGH UTTERANCE POSITION (initial, medial, or final; that is, following a pause, nonadjacent to a pause, or preceding a pause), NUMBER OF SYLLABLES, PRESENCE OF COMPLEX SYLLABLE NUCLEUS (diphthongs), PRESENCE OF SYLLABLE CODA (i.e. syllable-final nasals), SYLLABLE DURATION, RELATIVE AMPLITUDE, and PITCH CHANGE NORMALIZED FOR DURATION. Two measures that are idiosyncratic to Mandarin were also included: SYLLABLE REDUPLICATION, and PRESENCE OF MARKED TONE. These language-specific measures are explained below.

First, syllable reduplication occurs in adult-directed Mandarin speech for certain grammatical operations such as creating delimitatives for certain verbs (e.g. /kan/ ‘to look’; /kankan/ ‘to look a little’) or adding vividness to adjectives or adverbs. Reduplication may also be used to create baby-talk forms, as in the case in other languages, such as French or English (Ferguson, 1964). The questions for this study were how frequently reduplication occurs in infant-directed speech, how strongly the process is subject to the grammatical constraints of the adult language, and whether reduplication was equally likely to apply to functional and lexical items.
Second, according to phonological descriptions of Mandarin (Chao, 1930; Erbaugh, 1992), Mandarin has four lexically specified tones: HH (1st tone), LH (2nd tone), LL (3rd tone), and HL (4th tone). The doubling of L and H indicates the length of the syllable. In addition, there is a postlexical tone called the ‘neutral tone’ (Yip, 1989), the surface realization of which is determined by its preceding tone:

\[
\begin{align*}
\text{L/}& \quad \text{HH} \\
\text{LH} & \quad \text{[Neutral tone]} \\
\text{H/} & \quad \text{LL}
\end{align*}
\]

In essence, neutral tones are realized in a fashion that produces tonal alternations in conformance with the putative phonological universal obligatory contour principle (Leben, 1978; McCarthy, 1986): low tones occur after tones with high components; high tones occur after pure low tones. We presume that such alternation is the unmarked case and that marked tonal sequences, in which a tone with a high component follows another tone with a high component or in which a pure low tone follows another pure low tone, may be salient to learners who do not yet have language-specific knowledge of Mandarin phonology. Our expectation was that marked tones (second tones in marked sequences) would occur more often in word-initial position in lexical than functional items.

RESULTS AND DISCUSSION

Results and analyses of all individual measures for each mother are shown in Table 1. For several of the continuous measures, the data exhibited extreme non-normality or significant nonhomogeneity of variance; in these instances, Mann–Whitney \( U \)s were calculated (z-score approximations to \( U \) are reported below). For some of the categorical measures, excessive numbers of cells in the contingency tables were sparsely populated; in these instances, Fisher’s Exact Tests were used in place of \( \chi^2 \). Yates’ correction for continuity was applied in all single-df \( \chi^2 \) analyses.

With only a few exceptions, we found highly significant differences between functional and lexical items on all measures for both subjects. As we discuss below, however, no one of these cues is a highly valid predictor of category membership. To assess the combined predictive power of the constellation of cues, we used the entire set of measures from portions of the word samples to train a self-organizing neural network. We then tested the network’s performance on the remainders of the word samples. These novel items were classified appropriately about 90\% of the time—a level of accuracy unquestionably sufficient for bootstrapping.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Mother M1</th>
<th></th>
<th>Mother M2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Type frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type frequency</td>
<td>Mdn = 300</td>
<td>Mdn = 21.50</td>
<td>z_u = 7.37**</td>
<td>Mdn = 1100</td>
<td>Mdn = 2000</td>
<td>z_u = 6.51***</td>
</tr>
<tr>
<td></td>
<td>Rng = 1–58</td>
<td>Rng = 4–92</td>
<td></td>
<td>Rng = 2–68</td>
<td>Rng = 8–88</td>
<td></td>
</tr>
<tr>
<td>Rough utterance position^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C:</td>
<td>7</td>
<td>0</td>
<td></td>
<td>C: 9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I:</td>
<td>8</td>
<td>7</td>
<td>χ²(2) =</td>
<td>I: 9</td>
<td>8</td>
<td>χ²(2) =</td>
</tr>
<tr>
<td>M:</td>
<td>36</td>
<td>19</td>
<td>5.62</td>
<td>M: 12</td>
<td>15</td>
<td>12.78**</td>
</tr>
<tr>
<td>F:</td>
<td>16</td>
<td>5</td>
<td></td>
<td>F: 19</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Syllables in word</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:</td>
<td>31</td>
<td>29</td>
<td></td>
<td>1: 35</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>2:</td>
<td>35</td>
<td>2</td>
<td>z_u = 3.76***</td>
<td>2: 14</td>
<td>4</td>
<td>z_u = 1.33</td>
</tr>
<tr>
<td>3:</td>
<td>1</td>
<td>0</td>
<td></td>
<td>3: 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Reduplication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−:</td>
<td>48</td>
<td>31</td>
<td>χ²(1) = 9.17*</td>
<td>−: 41</td>
<td>28</td>
<td>Fisher Exact</td>
</tr>
<tr>
<td>+:</td>
<td>19</td>
<td>0</td>
<td></td>
<td>+: 8</td>
<td>0</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Syllable coda:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ:</td>
<td>39</td>
<td>31</td>
<td>χ²(1) =</td>
<td>θ: 36</td>
<td>27</td>
<td>χ²(1) =</td>
</tr>
<tr>
<td>Non-θ:</td>
<td>28</td>
<td>0</td>
<td>16.15***</td>
<td>Non-θ: 13</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Diphth:</td>
<td>36</td>
<td>5</td>
<td>10.82**</td>
<td>Diphth: 37</td>
<td>6</td>
<td>19.00***</td>
</tr>
<tr>
<td>Syllable nucleus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple:</td>
<td>31</td>
<td>26</td>
<td>χ²(1) =</td>
<td>Simple: 12</td>
<td>22</td>
<td>χ²(1) =</td>
</tr>
<tr>
<td>Diphth:</td>
<td>36</td>
<td>5</td>
<td>10.82**</td>
<td>Diphth: 37</td>
<td>6</td>
<td>19.00***</td>
</tr>
<tr>
<td>Marked tone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−:</td>
<td>31</td>
<td>26</td>
<td>χ²(1) =</td>
<td>−: 22</td>
<td>24</td>
<td>χ²(1) =</td>
</tr>
<tr>
<td>+:</td>
<td>36</td>
<td>5</td>
<td>10.82**</td>
<td>+: 27</td>
<td>4</td>
<td>10.70**</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>M: 218.80</td>
<td>t(96) =</td>
<td></td>
<td>M: 211.66</td>
<td>t(96) =</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s.d.: 95.65</td>
<td>4.28***</td>
<td>t(96) =</td>
<td>s.d.: 81.46</td>
<td>27.33</td>
<td>6.84***</td>
</tr>
<tr>
<td>Relative amplitude</td>
<td>M: 0.937</td>
<td>0.857</td>
<td>z_u =</td>
<td>M: 0.954</td>
<td>0.877</td>
<td>z_u =</td>
</tr>
<tr>
<td></td>
<td>s.d.: 0.090</td>
<td>0.100</td>
<td>4.22***</td>
<td>s.d.: 0.071</td>
<td>0.116</td>
<td>3.48**</td>
</tr>
<tr>
<td>Pitch change (semitone/S)</td>
<td>M: 25.40</td>
<td>23.76</td>
<td>t(66) =</td>
<td>M: 31.11</td>
<td>26.26</td>
<td>t(75) =</td>
</tr>
<tr>
<td></td>
<td>s.d.: 14.79</td>
<td>7.56</td>
<td></td>
<td>s.d.: 17.49</td>
<td>16.87</td>
<td>1.28</td>
</tr>
</tbody>
</table>

^a C = citation (isolated) form, I = initial, M = medial, F = final.

*p < 0.05; **p < 0.01; ***p < 0.001.
Distributional measures of functional and lexical items
The first two measures concerned general distributional properties of the sampled words: how often they occurred and where they occurred with respect to pauses in the mothers’ speech.

_Type frequency_. The frequency of each distinct type in the random subsample of words was computed over the entire transcript for each mother. For Mother M₁, the sample included 46 lexical types and 12 functional types; for Mother M₂, the sample included 29 lexical types and 13 functional types. As expected, functional item types occurred more frequently than did lexical item types. Mann–Whitney analyses confirmed that frequency distributions differed across categories for both mothers at better than \( p < 0.001 \).

_Rough utterance position_. In the syntax of Mandarin, certain functional items occur in the beginning of a syntactic constituent (e.g. the demonstrative in an NP), others occur at the end of a constituent (e.g. verb particles), and there are also sentence-final particles (e.g. a, ne). Lexical items can also occur in the beginning of a syntactic constituent (e.g. the verb in a VP) or the end of a constituent (e.g. the noun in an NP). We quantitatively compared functional and lexical items with regard to how often they occurred in citation form (in isolation), or utterance-initial, -medial, or -final position.

Numbers of functional and lexical items appearing in different utterance positions are shown in Table 1. For both mothers, more lexical items than functional items appeared in citation form, as Morgan et al. (1996) had previously found for English. In the subsample of words tabulated in Table 1, no functional items occurred in citation form. In fact, over the complete transcripts, Mother M₁ never used any functional items in citation form, and Mother M₂ used only one on one occasion. Also for both mothers, more lexical items than functional items occurred in utterance-final position (similar to Morgan et al.’s finding for English), whereas equal numbers of lexical and functional items occurred in utterance-initial position. \( \chi^2 \) analyses (omitting citation forms) showed that frequency of occurrence at different utterance positions varied significantly across categories for Mother M₂, though, in this small sample, not for Mother M₁.

Phonological measures of functional and lexical items
The next set of measures assessed the phonological content of sampled lexical and functional items. Several recent theoretical accounts have provided cross-linguistic evidence for a universal **minimal phonological word** which must comprise at least one binary foot (McCarthy & Prince, 1995; Demuth, 1996). Phonological structures constituting binary feet include bisyllables with short vowels (CVCV), single closed syllables (CVC), and single syllables with long or diphthongized vowels (CVV). This minimal
word criterion has been taken to apply to lexical items, but not necessarily to functional items (in English, for example, several functional items, such as the or to, consist only of single syllables with short vowels, which do not constitute binary feet). Therefore, we expected Mandarin lexical items to be more likely than Mandarin functional items to be bisyllabic or multisyllabic, to have non-null codas, and to have diphthongs. Note, however, that the definition of ‘minimal word’ is disjunctive, so that there was no expectation that individual Mandarin lexical items would necessarily possess all three of these attributes.

Number of syllables. The number of syllables in each token in the subsamples of words was counted. Our expectation was that lexical items would be more likely to be bisyllabic or multisyllabic than would functional items. This expectation was borne out from Mother M1, but not for Mother M2. In this instance, the lack of a significant difference for Mother M2 is probably due to sampling error: only 29% of lexical items in the subsample for this speaker were multisyllabic, compared to 49% in the entire transcript; conversely, 14% of functional items in the subsample were multisyllabic, whereas only 9% were in the entire transcript. Comparable proportions hold for Mother M1 over the entire transcript: for this speaker, 53% of all lexical items were bi- or multisyllabic, compared to 8% of all functional items.

Syllable reduplication. As noted earlier, reduplication is used in Mandarin to create baby-talk forms, as well as for other purposes. In our transcripts, for both subjects, about half of all bisyllabic lexical items were reduplications, almost all of which were distinctive baby-talk forms. Reduplication in functional items was very rare. Analyses of the subsamples showed that reduplication was significantly more likely to occur in lexical items than functional items for both subjects. The process of reduplication may tend to provide a canonical word form for lexical items (disyllabic) distinct from functional items (monosyllabic). This phenomenon has been noted in several other languages (Ferguson 1964), where disyllabification through the use of suffixes seems to play a similar role in infant-directed speech.

Syllabic structure. For both subjects, as expected, lexical items were more likely to have non-null syllable codas (syllable-final nasals) and complex syllable nuclei (diphthongs) than were functional items. On average, about 35% of the two speakers’ sampled lexical items possessed non-null syllable codas, whereas less than 5% of their sampled functional items did so. Across the two speakers, 65% of sampled lexical items had diphthongs, but only 10% of sampled functional items did so. In each instance, these differences were significant. Interestingly, in one-syllable lexical items, non-null codas and complex nuclei did not occur independently of one another (Mother M1: \( \chi^2(1) = 4.4, p < 0.05 \); Mother M2: Fisher’s Exact Test \( p < 0.001 \)). Rather, one-syllable lexical items tended to have non-null codas or complex nuclei, but not both.
Marked tone. We noted above that lexical and functional items in Mandarin differ in the frequency with which they have underlying underspecified or unspecified tone (neutral tone). Learners, of course, have no direct access to speakers’ underlying phonological representations, but they may well be able to distinguish different types of tones or tonal sequences. In essence, the Mandarin neutral tone rule operates to ensure that tonal sequences conform to the Obligatory Contour Principle (OCP). If this principle is a component of UG, as has been claimed (e.g. McCarthy, 1986), then learners with little or no Mandarin-specific knowledge of phonology might be expected to be able to distinguish tonal sequences conforming to OCP (unmarked sequences) from tonal sequences violating OCP (marked sequences). In these terms, our expectations were that lexical items would be more likely to complete marked sequences than would functional items.

Each token in the subsamples was coded for whether its initial tone completed a marked tonal sequence (i.e. manifested a [+H] tone – [H], [HH], [HL], [LH] – following a [+H] tone, or a [−H] tone – [L], [LL] – following a [−H] tone). By default, monosyllabic tokens in utterance-initial position (including those uttered in citation form) were coded as not completing marked tonal sequences. Note that under this scheme, words whose initial or only syllable carried neutral tone were never coded as completing marked tonal sequences. However, for words whose first or only syllable carried some specified tone, some tokens might be coded as completing marked tonal sequences whereas others would not, depending on the tonal context in which they appeared.

As shown in Table 1, for each mother only about 15% of functional items completed marked tonal sequences. Interestingly, for both mothers, a preponderance of the functional items used that had specified tones appeared in utterance-initial position and so did not contribute to violations of OCP. In contrast, half or more of the lexical items used by each mother completed marked tonal sequences. These differences across categories were significant for both mothers. (The identical frequencies shown for Mother M1 in Table 1 for syllable nucleus and marked tone are coincidental. These two properties were distributed independently of one another.)

Acoustic measures of functional and lexical items

The last set of measures compared lexical and functional items on three basic acoustic properties: duration, amplitude, and pitch. For items with 2 or 3 syllables, these measures were computed for each syllable. These syllabic measures were averaged to derive item measures for duration and pitch change. For relative amplitude, however, the maximum syllabic measure was used as the item measure.

Syllable duration and relative amplitude. For both subjects, on average, syllable durations for lexical items (Mother M1, 219 msec; Mother M2,
212 msec) were roughly twice as long as syllable durations for functional items (Mother M1, 133 msec; Mother M2, 103 msec) (mean syllable duration was computed for items with more than one syllable). These differences were very highly significant. Relative amplitude was determined by computing the ratio of the RMS energy of the selected syllable to the RMS energy of the loudest syllable in the utterance. For both mothers, lexical items were often the loci of peak amplitude, whereas functional items rarely were; overall, for both mothers, lexical items had significantly greater relative amplitude than did functional items.

Pitch change. Total amount of pitch change (including all fluctuations) was measured in semitones and normalized for syllable duration. This was the one measure that did not differ significantly across categories for either mother. For both subjects, the amount of pitch change was around 25–30 semitones/second for both lexical and functional items.

Individual measures as predictors of grammatical category membership

Although all the categorical measures we used produced significant differences between functional and lexical items for at least one (and in most cases both) mothers, none of these measures considered alone would be particularly useful for sorting the words in the subsamples into rudimentary grammatical categories. In some instances, one value of a categorical measure excludes almost all instances of one category or another. For example, no functional items for either mother were reduplicated, and only one for one mother had a non-null coda; on the other hand, very few lexical items (at least for Mother M2) began with tones that completed a tonal sequence conforming to OCP. However, in each case, the particular value of the categorical measure also excludes many instances of the opposite category. As we noted above, some of the categorical measures we included were disjunctively related. Thus, lexical items tended to have non-null syllable codas or diphthongized syllable nuclei, but not both. Disjunctively combining these measures would pick up most of the lexical items in the subsamples, but a sort based on this simple combination of measures would also include several functional items. Moreover, it is unclear how or why learners would choose this particular combination of measures for categorizing input words.

The distributions of individual continuous measures for the two categories of items all displayed some overlap, so that it is not clear that any of these individual variables had consistently high cue validity. The syllable duration data from Mothers M1 and M2 shown in Fig. 1 illustrates this general pattern. As noted, for both mothers, mean duration of lexical items was roughly twice that of functional items: this difference was highly significant. If, as observers knowing the true category membership of all items, we were to set cutoff values of duration that optimized category assessment, we would
find that duration had a maximal cue validity of $0.71$ for Mother M1 (with a cutoff of $125$ msec) and $0.90$ for Mother M2 (with a cutoff of $135$ msec).

However, even if language learners were to possess the memory capacity to retain accurate values of a given cue across substantial numbers of input tokens, they would have only single distributions to work with. Continuing our example, combined distributions of vowel durations from Mothers M1 and M2 are shown in Fig. 2. It may be readily seen that these distributions fail to afford any simple means for establishing optimal category cutoffs. Cue validity for vowel duration with a cutoff value guessed by the learner might well be no better than (and might even be worse than) chance. In principle, learners could adopt any of several unbiased procedures for estimating cutoff values. For example, the learner might adopt the mean (or median) of the
overall distribution of values as the cutoff, assigning items with values greater than the mean (or median) to one class and items with values smaller than the mean (or median) to the other. Alternatively, the learner might use a clustering algorithm, say, one in which between-cluster variance was maximized while within-cluster variance was minimized (i.e. \(k\)-means clustering).

For Mother M\(2\)'s data, all these unbiased procedures provide reasonable results. The clustering algorithm (applied after outliers more than 2 s.d. from the mean were removed) produced an overall correct classification rate of 66\% ; splitting the distribution at the mean was even better (81\% correct classification), and splitting the distribution at the median was better yet (84\% correct classification). However, for Mother M\(1\)'s data, none of the procedures performed impressively. Splitting the overall distribution at the mean produced the best result (60\% correct classification), clustering
produced the worst result (56% correct classification), and splitting the distribution at the median was intermediate. More specifically, for Mother M1’s data, 49% of lexical items and 19% of functional items would be misclassified if the mean of the overall distribution were adopted as the cutoff value. That virtually half the lexical items would be incorrectly classified is due to the fact that the distribution of durations was positively skewed. For this mother’s data, selecting the median rather than the mean would reduce the proportion of lexical item misclassifications but increase the proportion of functional item misclassifications; overall, none of the estimated cutoff values was close to optimal. For the other continuous measures, the distributional overlap between the two categories was even larger than was the case for duration. Thus, in general, no single measure, continuous or categorical, served uniformly across subjects as a highly valid predictor of grammatical category membership.

Sets of measures as predictors of grammatical category membership

To simulate how language learners might use the set of perceptually available cues that we measured, we examined the development of classificatory maps in Kohonen-style neural networks (Kohonen, Kangas & Laaksonen, 1992). Such networks are self-organizing (no external teaching or feedback) and project a multi-dimensional input space onto a 2-dimensional grid; learning occurs by adjusting the response of the network over time to better approximate the distribution of the training inputs.

In each of the 50 Monte Carlo simulations we ran for each subject, a network with a 20 x 30 grid of nodes was trained on a random 60% of subset of the analysed items from a subject and was then tested on the remaining 40% of the items from that subject. Values of each of the measures included in the input vectors (all those noted in the previous section except pitch change) were normalized so that each measure could potentially assume a similar weight in the resulting network (see Shi, 1995 for additional details). At the end of training, the node with the best fit to each training exemplar was labelled with the category of that exemplar. Test items were scored as follows. First, the single node with the best fit to the test item was identified; the classification neighbourhood included all nodes within a test radius r around this best fit node. If the neighbourhood contained nodes with the appropriate category label, but no nodes with the inappropriate category label, the test item was scored correct. If the neighbourhood contained nodes with the inappropriate label, but no nodes with the appropriate label, the test item was scored incorrect. If the neighbourhood contained some nodes with each label, the item was scored confused, and if the neigh-

[2] Results obtained with smaller grids are presented later, for measures from one of the Turkish mothers.
bourhood contained no nodes with either label, the item was scored unclassified. Two criteria were used to select an optimal neighbourhood size: maximal proportion of classified test items (correct plus incorrect) and most nearly equal proportions of confused and unclassified test items. These two criteria yielded the same neighbourhood size. Note that the labels were used only for external scoring of the test results and did not affect the internal operation of the network in any fashion. (In much the same way, language learners may know that words share similar grammatical properties without knowing that they are `nouns'.) Classification results for test items are presented in Table 2 and Fig. 3.

**Table 2. Mandarin network performance results (percentages)**

<table>
<thead>
<tr>
<th></th>
<th>Classified</th>
<th>Nonclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mother M1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical</td>
<td>77.5</td>
<td>22.5</td>
</tr>
<tr>
<td>Functional</td>
<td>74.4</td>
<td>25.6</td>
</tr>
<tr>
<td>Total</td>
<td>76.5</td>
<td>23.5</td>
</tr>
<tr>
<td><strong>Mother M2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical</td>
<td>79.4</td>
<td>20.6</td>
</tr>
<tr>
<td>Functional</td>
<td>79.4</td>
<td>20.6</td>
</tr>
<tr>
<td>Total</td>
<td>79.4</td>
<td>20.6</td>
</tr>
</tbody>
</table>

Across the two mothers, slightly less than 80% of all test items were classified. Of items that were classified, the overwhelming preponderance were correctly classified, as Fig. 3 shows. For Mother M1, 93% of all classified test items were classified correctly, including 90% of functional items and 94% of lexical items. Note that the overall level of accuracy achieved by a self-organizing learning system on the basis of multiple measures was substantially better than the optimal accuracy afforded an omniscient observer by the best single measure (vowel duration: 71%) for Mother M1. For Mother M2, 88% of all classified test items were classified correctly, including 90% of functional items and 87% of lexical items. For items in this mother’s speech, the level of accuracy achieved by the untaught Kohonen map was comparable to that afforded an omniscient observer by the best single measure. Moreover, note that the classification accuracies for functional and lexical categories were quite similar within mothers: high accuracy within one category was not obtained at the cost of sacrificing accuracy in the alternate.

Two tokens from Mother M1 (lexical *chi* ‘eat’ and functional *meiyou* ‘did not’) and five tokens from Mother M2 (lexical *ayi* ‘aunt,’ *li* ‘strength,’ *qilai* ‘get up’, and *zheme* ‘so,’ and functional *neng* ‘can’) were consistently misclassified over the sets of simulations. In every case, these word tokens
possessed multiple attributes that were typical of the alternate category. The grammatical classification of most of these words is uncontroversial; in these instances, the network simply produced incorrect predictions.

In the case of zheme ‘so’, however, the network results may indicate that our original grammatical classification was in error. We classified zheme as an adverb on the grounds that it occurs as an intensifying modifier before verbs, adjectives, and adverbs in the same manner as do adverbs in Mandarin, as shown in the sentences in (1). Unlike so in English, zheme does not appear as a pro-form in conjoined constructions, as shown in (2). On the other hand, in certain contexts, as in (3), zheme does behave like a pro-form. The grammatical classification of zheme is therefore somewhat ambiguous, and it is possible that the phonological evidence correctly indicates that zheme should be considered to be a functional item. Note that if we had originally

Fig. 3. Classification accuracy for Mandarin tokens.
classified *zheme* as a functional item, the network results for Mother M2 would have been better than those reported here.

1a. *zheme* xiaoxin de zoulu
   so carefully particle walk
   ‘walk so carefully’

1b. *zheme* piaoliang de haizi
   so beautiful particle child
   ‘such a beautiful child’

1c. Ta *zheme* xihuan Jingxi ya!
   he so like Peking Opera particle
   ‘He likes Peking Opera so much!’

2. *Wo* xihuan yuebing; *ni* ye *zheme*.
   I like mooncake you also so
   ‘I like mooncake, so do you.’

3. *zheme* chang
   in-this-way sing
   ‘sing like this’

In summary, although individual measures of Mandarin infant-directed speech tokens are at best moderately accurate predictors of membership in superordinate grammatical categories, sets of measures may be highly accurate predictors. Our results provide evidence that there exists a system capable of learning superordinate grammatical categories on the basis of phonological cues at a level of accuracy that is unquestionably sufficient for purposes of bootstrapping – a result comparable to that previously obtained for English (Morgan *et al.*, 1996).

We now have data that phonological properties of infant-directed speech may be sufficient to bootstrap rudimentary grammatical categories in both Mandarin and English. However, both these languages possess functional items that are predominantly free morphemes. Whether these results are generalizable to languages with other types of morphological systems is not clear. To address this question, in the next study we take up analyses of Turkish infant-directed speech.

**STUDY 2**

In contrast to Mandarin, Turkish, a Turkic language spoken in Turkey and a few areas in Greece, Bulgaria, and Macedonia, possesses a rich system of inflectional morphology. Syntactically, Turkish is a predominantly left-branching language. The language has postpositions but not prepositions, and noun modifiers including relative clauses precede their heads. Turkish has a neutral word order of SOV, although other word orders are used for
pragmatic purposes. For instance, the OVS order allows objects to be focused, and the OSV order allows subjects to be focused. Morphologically, Turkish is an agglutinating language. Affixes are almost all syllabic; when agglutinated, their phonological identity remains unchanged except for certain harmonic processes described later. Affixed functional item categories in Turkish include nominal case, number, and possession, and verbal voice, negation, modality, aspect, tense, person, and number. Nonaffixed functional item categories in Turkish include personal pronouns, demonstratives, conjunctions, complementizers, question particles, and an indefinite article. Overviews of Turkish grammar and acquisition may be found in Aksu-Koç & Slobin (1985).

METHOD

Subjects

Two mother–child dyads participated in this study. The mothers were both from Istanbul, Turkey. One mother stayed home with her baby. The other mother went to school part-time and shared child care with the baby’s grandmother, who was also a native speaker of Turkish. The predominant language for both families was Turkish. The dyads were recorded when the children, both girls, were 0;11:25 and 1;8:3 respectively. Due to difficulties in finding Turkish-speaking mothers in the New England area, only one one-year-old baby was located. The twenty-month-old was included in the study so that her mother’s speech could be compared with that of the other mother. The one-year-old was not yet producing words, and the twenty-month-old was producing 2–3 morpheme utterances.

Recording and transcription

Recordings were made under the same circumstances described above for the Mandarin dyads. The recordings were transcribed by a native speaker of Turkish, using the standard Turkish alphabet (created by a committee of linguists in Turkey in the 1920s), an alphabetic system corresponding to the phonemic inventory of the language. The transcription was done quasi-orthographically, with some phonetic detail so as to give a clearer description of the actual speech output. Each syllabic morpheme was coded for its category, i.e. either lexical or functional. Noun and verb stems, adjectives, and adverbs were coded as lexical items. Demonstratives, postpositions, particles, conjunctions, grammatical morphemes\(^3\) (verb and noun suffixes), and so forth were coded as functional items. Coding criteria were based on Swift (1963) and Underhill (1976).

---

\(^3\) Note that the final consonants of grammatical morphemes were always included.
Measures
Two hundred of each mother's morphemes were selected at random for analysis, including 100 lexical items and 100 functional items each. A variety of distributional, phonological, and acoustic measures was computed on the items in these samples. Many of these were similar to those computed on the Mandarin samples: TYPE FREQUENCY, ROUGH UtTERANCE POSITION, NUMBER OF SYLLABLES, PRESENCE OF SYLLABLE CODA, VOWEL DURATION (rather than syllable duration), RELATIVE AMPLITUDE, and PITCH CHANGE NORMALIZED FOR DURATION. One measure idiosyncratic to Turkish, VOWEL HARMONY, was also included.

A pervasive and transparent morphophonological system of vowel harmony is a prominent characteristic of Turkish phonology. Eight vowels are found in stems, representing all combinations of the features \[ \pm \text{high}, \pm \text{front}, \pm \text{round} \]: /u/, /i/, /u/, /i/, /o/, /e/, /o/, /a/. Affix vowels are underlyingly specified for height only. High affix vowels harmonize with the preceding vowel in both frontness and roundness; hence [Im] ‘my’ may be manifest as /gönülüm/ ‘my heart’, /elim/ ‘my hand’, /kulum/ ‘my arm’, or /başım/ ‘my head’. Low affix vowels harmonize with the preceding vowel in frontness, but are always [−round]; hence [IEr] ‘plural’ may be manifest as /adamlar/ ‘men’ or /günler/ ‘days’. We are not aware of any phonological principle stipulating that harmony is the unmarked case. Thus, unlike tonal alternation, some linguistic experience may be required in order to apprehend the distinction between harmonizing and nonharmonizing vowels. Given the salience of vowel harmony in Turkish, however, the required amount of experience may be minimal. Our expectation was that harmonizing vowels would occur more often in morpheme-initial position in functional than lexical items.

RESULTS AND DISCUSSION
Results and analyses of all individual measures for each mother are shown in Table 3. Analytic strategies were identical to those described in Study 1. As in Study 1, with only a single exception, we found highly significant differences between functional and lexical items on all measures for both subjects. As was the case for Mandarin, no one of these cues is a highly valid predictor of category membership. To assess the combined predictive power of the constellation of cues, we used the entire set of measures from subsets of items to train a self-organizing neural network. We then tested the network’s performance on a novel set of items. These novel items were classified appropriately about 84% of the time—a level of accuracy slightly less than that obtained for Mandarin in Study 1 but unquestionably sufficient for bootstrapping.
**Table 3. Turkish infant-directed speech: individual measures of lexical and functional items**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mother M1 Lex.</th>
<th>Mother M1 Func.</th>
<th>Mother M2 Lex.</th>
<th>Mother M2 Func.</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type frequency</td>
<td>Mdn = 900</td>
<td>Mdn = 4400</td>
<td>Mdn = 800</td>
<td>Mdn = 2500</td>
<td>z_u = 7.42***</td>
</tr>
<tr>
<td></td>
<td>Reg = 1-161</td>
<td>Reg = 1-135</td>
<td>Reg = 1-104</td>
<td>Reg = 1-90</td>
<td></td>
</tr>
<tr>
<td>Rough utterance position*</td>
<td>C: 11</td>
<td>0</td>
<td>C: 9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I: 22</td>
<td>10</td>
<td>I: 37</td>
<td>5</td>
<td>χ²(2) =</td>
</tr>
<tr>
<td></td>
<td>M: 56</td>
<td>58</td>
<td>M: 41</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F: 11</td>
<td>32</td>
<td>F: 13</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td># Syllables in morpheme</td>
<td>1: 51</td>
<td>91</td>
<td>1: 50</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: 38</td>
<td>9</td>
<td>2: 44</td>
<td>3</td>
<td>z_u = 5.76***</td>
</tr>
<tr>
<td></td>
<td>3: 11</td>
<td>0</td>
<td>3: 6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Syllable coda</td>
<td>Ø: 14</td>
<td>52</td>
<td>Ø: 15</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-Ø: 86</td>
<td>48</td>
<td>Non-Ø: 85</td>
<td>54</td>
<td>χ²(1) =</td>
</tr>
<tr>
<td></td>
<td>+: 32</td>
<td>75</td>
<td>+: 35</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>−: 68</td>
<td>25</td>
<td>−: 65</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Vowel harmony</td>
<td>M: 165.78</td>
<td>98.10</td>
<td>M: 151.18</td>
<td>108.90</td>
<td>z_u =</td>
</tr>
<tr>
<td></td>
<td>s.d.: 70.42</td>
<td>50.75</td>
<td>s.d.: 41.12</td>
<td>61.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M: 0.945</td>
<td>0.674</td>
<td>M: 0.929</td>
<td>0.821</td>
<td>t(198) =</td>
</tr>
<tr>
<td></td>
<td>s.d.: 0.081</td>
<td>0.111</td>
<td>s.d.: 0.110</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>Relative amplitude</td>
<td>M: 23.31</td>
<td>24.25</td>
<td>M: 27.79</td>
<td>27.22</td>
<td>z_u =</td>
</tr>
<tr>
<td></td>
<td>s.d.: 16.72</td>
<td>21.63</td>
<td>s.d.: 18.63</td>
<td>26.09</td>
<td>1.53</td>
</tr>
</tbody>
</table>

* p < 0.05; **p < 0.01, ***p < 0.001.
* C = citation (isolated) form, I = initial, M = medial, F = final.
Distributional measures of functional and lexical items

The first two measures concerned general distributional properties of the sampled words: how often they occurred and where they occurred with respect to pauses in the mothers’ speech.

Type frequency. Morphophonemic alternates of morphemes were counted as separate types, assuming that naive learners have no knowledge of the morphophonemic rules which link these variants. Turkish vowel harmony applies to affixed functional items; coding the variable forms of these items as different types tends to reduce frequency differences between the two classes. Even adopting this conservative coding, however, functional item types occurred 3–5 times more often than did lexical item types. Mann–Whitney analyses confirmed that frequency distributions differed across categories for both mothers at better than $p < 0.001$.

Rough utterance position. Numbers of functional and lexical items appearing in different utterance positions are shown in Table 3. For both mothers, lexical items sometimes appeared in citation form, but functional items never did. Also for both mothers, more lexical items than functional items occurred in utterance-initial position, whereas the reverse was true for utterance-final position. Although asymmetrical distributions of lexical and functional items at utterance boundaries have previously been noted for Mandarin and English, the pattern occurring in Turkish is opposite that occurring in either of the other two languages. $\chi^2$ analyses (omitting citation forms) showed that frequency of occurrence at different utterance positions varied across categories at better than $p < 0.01$ for both mothers.

Phonological measures of functional and lexical items

The next three measures assessed the phonological content of sampled lexical and functional items.

Number of syllables. As expected, for both subjects, lexical items were more likely to be multisyllabic than were functional items. For both subjects, approximately 50% of lexical items were multisyllabic, whereas over 90% of functional items were monosyllabic. This difference was highly significant for both subjects, at better than $p < 0.001$.

Syllable structure. For both mothers, functional items were much more likely to have null codas than were lexical items. Approximately half of all function items had null syllable codas, whereas almost 90% of lexical items had one or more syllables with nonnull codas. This pattern is the mirror image of that found for number of syllables. Within lexical items, monosyllabic morphemes were more likely to possess nonnull codas than were multisyllabic morphemes (Mother T1, $\chi^2(2) = 8.87$, $p < 0.05$; Mother T2, $\chi^2(2) = 9.96$, $p < 0.001$).
Vowel harmony. We coded vowel harmony in much the same fashion as we had coded tone sequences in Mandarin: vowels were scored as harmonizing if they appeared to harmonize, regardless of whether this was due to application of a morphophonemic rule or due to the incidental juxtaposition of morphemes with harmonizing, but lexically specified, vowels. Specifically, an item was scored as harmonizing if its initial vowel was a high vowel agreeing in frontness and roundness or a low vowel agreeing in frontness only with the final vowel of the preceding morpheme. Note that low, round vowels do not harmonize in Turkish. However, because this may become apparent to learners only after they have discovered the patterns of morphophonemic alternations, we scored low, round vowels in the same manner as we did low, nonround vowels.

Data from the two mothers were quite similar. For each mother, about two-thirds of lexical items did not appear to harmonize, whereas about four-fifths of functional items did appear to harmonize. In our coding, however, utterance position was confounded with harmony – items occurring in initial position were coded as not harmonizing. Because these items were predominantly lexical, this strategy served to inflate the differences between lexical and functional categories with respect to vowel harmony. We therefore computed a second tabulation for each subject, from which utterance-initial and citation-form items were excluded. The results of these again showed highly significant differences between categories in vowel harmony, in the same direction as our initial tabulations: Mother T\textsubscript{1}, $\chi^2(1) = 20.78$, $p < 0.001$; Mother T\textsubscript{2}: $\chi^2(1) = 7.21$, $p < 0.01$.

Acoustic measures of functional and lexical items

Vowel duration and relative amplitude. For both subjects, on average, vowel durations for lexical items were about one and one-half times as long as vowel durations for functional items. These differences were again very highly significant (better than $p < 0.001$). For both subjects, relative amplitude of lexical items was significantly greater than that of functional items at better than $p < 0.001$. The strength of the vowel duration effects is somewhat surprising given that functional items in Turkish often appear in phrase- or utterance-final positions, locations that are associated with vowel lengthening in many languages, including English.

Pitch change. As in Mandarin, mean values of this measure did not differ significantly across categories for either subject. However, for both subjects, pitch change was more variable in functional items than in lexical items: many function items had very small pitch slopes, but a handful had quite large pitch slopes.
Sets of measures as predictors of grammatical category membership

As for Mandarin, although most measures showed highly significant mean differences, the distributions of all individual measures for the two categories displayed considerable overlap, so that none of the individual variables that we measured had high cue validity. Once again, to simulate how language learners might use the set of perceptually available cues that we measured, we conducted Monte Carlo simulations with Kohonen maps. The procedures followed were identical to those described earlier, except that with a larger number of training exemplars the optimal neighbourhood size was smaller than that used with the Mandarin data.

We were uncertain, however, how to treat vowel harmony. On the one hand, we know of no phonological universal that would predispose learners to expect or attend to harmony. On the other hand, vowel harmony is a salient characteristic of Turkish, one that might be apparent even to naïve learners. In the end, we decided to conduct two sets of simulations, one including the harmony measure, the other excluding it. Classification results for test items in these simulations are presented in Table 4 and Fig. 4 which presents results including harmony. The results excluding harmony are comparable, as is obvious from Table 4.

Both quantitatively and qualitatively, the Turkish modelling results were very similar to the Mandarin modelling results. Slightly less than 80% of test
items were classified; of these, about 80–85% were classified correctly. For Mother T1, when the vowel harmony measure was included in input, 86% of all classified test items were classified correctly, including 86% of functional items and 87% of lexical items. When the vowel harmony measure was excluded from input, essentially identical levels of accuracy were obtained on test items from Mother T1 (86% overall; 86% functional, 86% lexical). For Mother T2, when the vowel harmony measure was included in input, 82% of all classified test items were classified correctly, including 83% of functional items and 81% of lexical items. When the vowel harmony measure was excluded from input, slightly higher levels of accuracy were

Note that Figs 3 and 4 are the preferred way of representing the network results. It is not unreasonable to assume that the learner sets aside those ‘confusing’ and ‘unclassifiable’ items for the time being while focusing on the ‘classifiable’ items.
obtained on test items from Mother T2 (84% overall; 85% functional, 84% lexical). Including or excluding the vowel harmony measure had little appreciable impact on network performance, indicating that the information supplied by vowel harmony for grammatical classification was redundant with that supplied by other measures. As we discuss below, such redundancy may be important in facilitating discovery of cues to grammatical category membership.

The networks we used included large numbers of nodes relative to the numbers of training exemplars. To examine whether grids as large as 20 by 30 nodes were necessary to produce results comparable to those we obtained, we conducted several additional simulations using measures from Mother T1, across which grid size was varied (optimal neighbourhood size was selected for each grid size). Results of these simulations (based on five runs for each grid size) are shown in Fig. 5. As expected, the proportion of test items that were classified declined as grid size was reduced. Among test items that were classified, however, a consistent proportion (about 85%) was classified correctly, showing that large grids are not necessary for accurate classification.

**Discussion**

The results of the studies reported here indicate that, over typologically distinct languages, children’s input appears to include sets of cues sufficient
in principle to support sorting of words or morphemes into sets corresponding to rudimentary grammatical categories. In samples of infant-directed speech in Mandarin and Turkish, multiple distributional, phonological, and acoustic measures differed significantly for functional and lexical items. Taken individually, none of these measures is likely to serve as a highly valid predictor of grammatical category membership. Taken ensemble, however, the sets of measures enabled Kohonen maps to classify novel items as functional or lexical with accuracies ranging from 82 to 93%. Comparable findings have recently been reported by Allen & Christiansen (1996) with respect to the related task of word segmentation: multiple correlated cues, each of which is a low validity predictor, can together serve as the basis for identifying aspects of linguistic structure. These results raise the possibility that infants may be able to sort segmented words or morphemes into appropriate superordinate grammatical categories before they have learned the meanings of any of these words or morphemes and before they have embarked on any detailed syntactic distributional analysis of their input.

Such precocious knowledge of rudimentary grammatical categories could, in principle, assist in solving several basic problems in language learning, including discovery of word and phrase boundaries, learning of word-meaning mappings, and acquisition of refined sets of grammatical categories (see Morgan et al., 1996 for additional discussion). In none of these instances, to our knowledge, is knowledge of rudimentary grammatical categories necessary for learning. Rather, in each instance, such knowledge may supply additional constraint, particularly useful in the early stages of acquisition when children lack other types of knowledge that could be brought to bear. We illustrate this here in relation to the problem of acquiring a complete set of grammatical categories.

In general, the problem of grammatical category acquisition can be conceived as splitting superordinate categories into smaller ones or as joining subordinate categories into larger ones. In this article, we approached the problem in terms of splitting the total class of word tokens into two subclasses. In contrast, semantic and distributional approaches to grammatical category acquisition have typically contemplated the problem as one of joining individual word types into groups. These two approaches need not be mutually exclusive: joining procedures at least implicitly assume some superordinate category, inasmuch as it is words or morphemes to which these procedures are applied, rather than syllables or phrases. Early splitting of tokens into sets corresponding to functional and lexical categories can assist joining procedures both by providing information useful for joining subgroups and by providing the learner with a possible means of identifying optimal levels of joining.

Splitting words into functional and lexical sets can assist semantics-based joining procedures (e.g. Pinker, 1984). If the learner can identify which
morphemes are functional items and if the learner knows that possible meanings of functional items are highly constrained, then the number of possible mappings between morphemes and contextually available concepts can be significantly reduced. Simplifying discovery of word meanings will directly aid in joining new words to existing categories. Splitting words into functional and lexical sets can also assist distribution-based joining procedures (e.g. Maratsos & Chalkley, 1980). Under these procedures, the ability to pick out the most informative distributional patterns will be useful. Most typically for lexical categories (at least in the Maratsos & Chalkley model) key patterns involve functional item phrase-mates. Knowledge of which morphemes are functional and which lexical would thus allow the learner to focus on relevant co-occurrences.

In general, joining procedures yield multiple possible levels of category generalization but do not specify which is optimal. The hierarchical clustering analyses pursued by Mintz, Newport & Bever (1995) provide a recent example. Mintz et al. investigated whether words could be identified as nouns or verbs on the basis of distributional information available in child-directed speech. They tabulated the lexical contexts of sets of common words and calculated similarity measures over these contexts. The similarity measures were then used to cluster the words hierarchically into categories. Citing results from a middle level of clustering, Mintz et al. showed that limited distributional information is useful in inducing form classes. HOW useful depends on the particular level of clustering chosen; on this issue, the clusters themselves, which are all that would be available to the learner, are uninformative. However, by taking into account categories split out on the basis of constellations of acoustic and phonological cues, the learner might be able to converge on an optimal level of clustering by, for example, selecting that level at which clustering is maximized while the proportion of clusters containing tokens from more than one of the split-out categories is minimized.

The learning model presented here serves as proof of the existence of an inductive mechanism that can acquire approximately correct rudimentary grammatical categories on the basis of correlated properties of input speech. Whether infants can perform any comparable feat is unknown at present. As we note below, the resources available to infants are in some ways richer, but in others more impoverished, than those available to our learning model. Thus, we do not claim any point-by-point correspondence between infants and our model. Nevertheless, we believe it plausible to hypothesize that infants can exploit multiple partly predictive cues in input speech to form initial inductions concerning linguistic categories.

The basic computational strategy used by our learning model appears to be within the grasp of preverbal infants. In a series of studies, Younger (1990, 1992; Younger & Cohen, 1983, 1986) has shown that 9-month-old infants can form a variety of visual categories based on correlated features. Lalonde
& Werker (1995) demonstrated that the ability to form correlation-based visual categories is associated with decline in ability to discriminate foreign speech sound contrasts, which itself has been attributed to the advent of phonemic category representations. Morgan & Saffran (1995) argued that phonemic categories must be constructed from correlations of phonetic and contextual features and provided evidence showing that, by 10 months, infants can correlate diverse properties of speech stimuli in a word segmentation task.

One important way in which the information available to human learners is significantly richer than that used by our learning model is that the model neglected to exploit any lexical information (beyond the measure of type frequency, which we presume is registered automatically; cf. Hasher & Zacks, 1984). For adults, category memberships of speech tokens are retrieved from the appropriate lexical entries in most instances: knowing the categories a word has appeared in previously is a powerful predictor of what categories the word may appear in in the future. The point at which infants come to be able to exploit such information is unknown. Note, however, that the abilities to retrieve lexical meanings or to record distributional properties of words or morphemes both entail the capacity to recognize speech tokens as exemplars of lexical types, which is the key to exploiting lexical category information as well. Adding such information to the input of our model could only have improved its performance.

On the other hand, the particular representations of input properties that we used are unlikely to be shared by infants. To be certain, none of the measures that we included is obscure; in many cases, studies exist showing that infants can make discriminations along the relevant dimension. For example, Bijeljac-Babic, Bertoncini & Mehler (1993) demonstrated that newborns can detect changes in syllable numbers, Eilers et al. (1984) showed that 5-month-old infants can discriminate vowel duration, and so forth. In no case is the relevant psychophysical function yet known, but it does seem quite unlikely that infants can represent quantities like vowel duration or relative amplitude with anything approaching the precision used here. However, it may not be implausible to assume that infants can at least represent these dimensions ordinally. This is a sufficient basis for use of correlational procedures, albeit less powerful ones than those used here. As a consequence, infants will probably require substantially larger input samples. Given the small sample sizes we used here, this does not seem to be problematic.

The input we provided our model may appear to differ from that available to infants in a more profound manner. We hand-picked the variables to include in our simulations, based on our adult knowledge and intuitions concerning what constitute relevant properties signalling the functional–lexical distinction. By doing so, we screened our most noise present in input. But how do infants know what to correlate?
One answer might be that infants are pretuned to attend to particular properties of speech. If we examine distinctive properties of functional and lexical items across Mandarin, Turkish, and English, we may see that certain properties appear to be substantively universal. For example, functional items tend to have smaller vowel (or syllable) duration and relative amplitude, to have fewer syllables per word, and to occur more frequently than do lexical items. Other distinctive properties might be seen as manifestations of more abstract, formal universals. In all three languages, functional items tend to have simpler syllable structures than do lexical items. The particular structures that are manifest, however, vary from language to language. Mandarin, for instance, permits only single nasal consonants in syllable codas; functional items tend to lack any coda. In contrast, English allows clusters of up to four consonants in syllable codas; many functional items have single consonant codas, but few have cluster codas. Yet other distinctive properties may reflect even more abstract universals. In general, phonological properties of functional items tend to be governed by those of lexical items, rather than the other way around. In Mandarin, the surface manifestation of functional item (neutral) tone is governed by the (full) tone of the preceding lexical item. In Turkish, segments belonging to functional items are more likely to be subject to vowel harmony or consonant assimilation than are segments belonging to lexical items. Infants might be predisposed to attend to contrasts of these sorts: those that do not correspond to substantive universals can be identified simply by noting their correlation with those that do. Moreover, functional and lexical items might possess the properties they do precisely because infants expect them to do so and analyse words with the appropriate properties as exemplars of the corresponding categories.

These explanations, however, are otiose. There is no need to seek a grammatical explanation for the phonological or acoustic properties of functional items. For syntactic reasons, these items will occur with high frequency. Because of this, speakers will be inclined to minimize them in production, in accordance with the phonological patterns of their native language. Functional items tend to be syntactically predictable and semantically light, carrying little informational load. Thus, hearers will not find speakers’ minimizations to be disruptive. In effect, speakers and hearers will conspire for communicative reasons to ensure that functional and lexical items are phonologically distinctive.

If infants must observe correlations among speech properties in order to discover which language-particular properties cue the functional–lexical distinction, there is no reason why they cannot do the same for language-general properties. So long as infants are predisposed to entertain categories defined by correlated features, any dimensionality reducing procedure, whether adaptive map or principle components analysis, will serve to preserve predictive, correlated properties and to filter out as noise non-
predictive, uncorrelated properties. Given the statistical structure of their input, infants may discover correlation-based proto-linguistic categories of a variety of sorts. These initial inductions allow infants to form more sophisticated representations and to bring to bear more powerful forms of linguistics analysis; subsequently, some of these rudimentary categories will be seen to be spurious and will be discarded, whereas others will be incorporated in the developing linguistic system. As development proceeds, phonological and acoustic information about the physical forms of words will be supplemented and ultimately superseded by lexical, semantic, distributional, and syntactic sources of information as defining bases of category membership. This, however, in no way diminishes the potential importance of speech-based information in forming the initial inductive basis for superordinate grammatical categories.

In conclusion, this article has shown that sets of distributional, phonological, and acoustic cues distinguishing lexical and functional items are available in infant-directed speech across such typologically distinct languages as Mandarin and Turkish. We further showed through neural network modelling of naïve learning that constellations of these perceptual cues are sufficient in principle for categorization of words into two rudimentary classes that closely correspond to lexical items and functional items. Such an initial classification may assist children both in beginning to parse input utterances and in forming the basis for computationally limited distributional analyses that will yield more complete and mature sets of grammatical categories.

REFERENCES


SHI, MORGAN & ALLOPPENA


Bases for Early Grammatical Categories


