J. Child Lang. 15 (1988), 395–410. Printed in Great Britain

Vocabulary simplification for children: a special case of 'motherese'?*

DONALD P. HAYES AND MARGARET G. AHRENS
Cornell University

(*Received* 2 *April* 1987)

ABSTRACT

A new corpus of spontaneous conversations between adults and children is examined for evidence that adults simplify their vocabulary choices when speaking with young children. If these simplifications are found to be age-dependent, then they would broaden the pattern of simplifications characteristic of 'motherese' to include lexical choice as well. For the age-range newborns to 12 years, the results are both consistent with and contrary to the attested set of grammatical simplifications. In this corpus, MLU and TTR are strongly age-dependent, but adults do not choose their words from the 10,000 most common word-types in English in an age-dependent manner. Rather, the additional types for school-aged children come from the same part of the vocabulary and share the same-shaped distributions as in adult speech with preschool children and infants. This absence of an age-dependent accommodation in word choice has implications for models of child lexical acquisition which assume adult language accommodation.

INTRODUCTION

The large body of research on 'motherese' has documented many age-dependent grammatical adjustments in adult speech to young children. These adjustments are among the many accommodations speakers make to each other during social interaction (Giles, Mulac & Johnson 1986, Hayes, Ahrens & Tsay 1987). In the case of 'motherese', they constitute accom-

[*] The authors wish to thank many who have contributed to these analyses: the parents, teachers, paramedics, nurses and children who generously allowed themselves to be recorded in the course of their normal activities; Tony Wootton, Gorden Wells, Catherine Snow and Harry Levin for loaning us samples of their similarly recorded texts; and finally, B. Hayes, H. Levin, F. Hayes, L. Meltzer, W. Lambert and D. Wilson for their comments on earlier drafts of this manuscript. Special thanks go to the anonymous Cornell alumnus who provided the mini-computer used to produce and analyse this corpus. Address for correspondence: Donald P. Hayes, Department of Sociology, 346 Uris Hall, Cornell University, Ithaca, NY 14853, USA.

modations made to maturing interactional partners. As such, many are systematic simplifications of the adult-to-adult speech standard. Among the most consistently noted of these simplifications are: reduction of preverb length and complexity; reduction in the number of verbal inflections; replacement of first- and second-person pronouns by third-person pronouns and other salient nouns or names; reduction in the number of embedded clauses and conjunctions; shortening of utterance lengths; reduction in the number of disfluencies and fragments; and slowing of speech rate (Cross 1977, Gleason 1977, Snow 1977, Clark 1979, Oksaar 1981, Gleitman, Newport & Gleitman 1984). The general effect of these systematic simplifications by adults is speech which is short, highly intelligible and grammatically very well-formed (Newport, Gleitman & Gleitman 1977).

The research reported here attempts to determine if word choice is also systematically simplified by speakers: specifically, do adults make age-dependent adjustments in their vocabulary choices for children? We find that the answer to this question depends on the statistical measures used to evaluate it. Previous research on adult-to-child speech has shown that expansion and repetition of words, phrases and sentences results in a restriction of adults' word choices (Snow 1977, Gleitman *et al.* 1984). Focusing on salient aspects of the immediate context has a similar restrictive effect (Ferguson 1964, 1977). The more direct and most often used means of measuring word choice, the type-token ratio (TTR), yields significant correlations between the number of different terms adults use and the age of the child being addressed. When analysed with this statistic, our data confirm this pattern. But, the TTR cannot show from which points in the lexicon the terms come when all the terms are arrayed according to their frequency of use. It especially does not indicate from which points the additional types used in speech to older children come.

A method of measuring the relative use of common and rare words against an independent frequency count was developed and used to analyse adult-to-child speech in order to determine what TTR alone cannot: whether adults skew their word choices towards the more common words for younger children, and draw more heavily upon less frequently encountered, uncommon words for older children.

METHOD

The corpus and samples

The primary data for these analyses are all natural conversations. Most were recorded by the participants in their own homes, without an investigator present. Others were recorded on a picnic, in a backyard, in an automobile travelling on vacation, in a Montessori classroom, at three nursery schools, in a hospital's 'pre-op' room and its nursery. Once the texts were transcribed

in machine-readable form, the analyses were completed by computer using programmes expressly developed for this word-choice analysis.[1]

The corpus of over 100,000 tokens consists of three sample sets. The first comprises 41 independent samples of adults speaking with children between the ages of three hours and 12 years. There are six additional independent samples from a longitudinal study of the same mother talking with her three preschool children at two points in time, one year apart. This corpus of 47 adult-to-child conversations consists of 62,500 tokens. Most of these samples are based on 1,000-word texts in full sentences. When a sample text contains many thousand tokens, the mean of its several 1,000-word constituents is used. The smallest sample text in this corpus contains 155 tokens (a published excerpt of a transcript from Schachter 1979). The second sample set, consisting of over 20,000 tokens, represents what four- to twelve-year-old children said to their parents. The third sample set, consisting of over 18,000 tokens, represents conversations between college-educated adults. These last samples serve as a standard against which adult–child conversations are compared.

The adult-to-child samples include seventeen families living in Scotland, England and the United States. Several investigators provided us with representative samples of their own similar field recordings: G. Wells; T. Wootton; C. Snow and H. Levin. The bulk of this corpus, however, came from recordings made in the Ithaca, New York, area. In producing this corpus, it was by design that there should be a wide range of parental income and educational attainments. Information about these seventeen families is provided in Table 1. The precise ages of the children were not available in many cases, so we have noted age to the nearest full year.

The use of a reference lexicon

Our procedures for measuring lexical use and for comparing texts are unique, so they will be described in more than the usual detail. To compare word choice in two texts requires that each be compared against a common standard – much as one uses a measuring tape as the standard to measure and

[1] The LEX programmes are written for IBM-PC machines, with DOS 2·0 or higher. The programmes were developed by the senior author and written by Peter Bond, Scott McAllister and David Post. They are not copy-protected and can be obtained by writing to the authors. The transcription rules in preparing a text for analysis are described in Hayes (1986*b*); the descriptive measures and comparative statistics for a half-million-word corpus are in Hayes (1986*a*); and a set of validation studies for these measures are reported in Hayes (1986*c*). These programmes lack lexical subtlety. For example, they cannot distinguish the several uses to which the same type can be put, nor can they distinguish modals from verbs when the two are interchangeable. Such refinements await computational developments which will permit their recognition.

TABLE 1. *The seventeen families*

Adult speakers and background	Children's sex and age	No. of samples
Mother, teenage, on welfare	Boys, infant and 2	4
Mother, working class	Girl, infant; boys 7 and 9	3
Mother, minority	Boy 2; girl 4	2*
Mother, minority, poor	Girls 3 and 7	2
Mother, father, working class	Boy 7	3
Mother, working class	Girl 3	1*
Mother, minority, middle class	Girl 3	1*
Mother, middle class	Girl 3	1*
Mother, middle class	Boy 5	4
Mother, middle class	Boy 2; girl 4	4*
Mother, middle class	Boy 5	1
Mother, father, middle class	Girl 3	2*
Mother, father, professional	Boy 4	5
Mother, professional	Boys, infant and 2; girl 4	11
Mother, professional	Boys, infant and 2	4
Mother, father, professional	Boys 4 and 5; girls 5 and 8	4*
Mother, father, professional	Girls 10 and 12	6

* Contains a sample(s) based on fewer than 1,000 tokens.

compare the heights of two children. We chose as our standard the American Heritage Dictionary list of English types (Carroll, Davies & Richman 1971). This is currently the best approximation of the frequency with which American English words are used. It is the largest, most modern, most comprehensive in topic coverage and most suitable list for analysing and comparing texts involving children. Other word-frequency lists are too small or were derived from too restricted an age or socio-economic status range to serve the purposes required here.

The Carroll *et al.* list is based on a corpus of over five million tokens sampled from over 1,000 different publications designed for children between the ages of nine and fifteen. Its word choice covers seventeen categories of school and non-school subjects. The list reports the relative frequency per million tokens for each of its 86,741 different types (adjusted for use across subject areas). The frequency per million and rank of each of the 10,000 most common words is incorporated into our programmes for analysing word choice. There is one difference in how the terms are used in the Carroll *et al.* list and in our reference lexicon. They separated the capitalized and uncapitalized versions of certain terms. We combined them. For example, 'beauty' (54·8/million) and 'Beauty' (5·6/million) became 'beauty' (60·4/million) for the purposes of our analysis. Consequently, the rank numbers associated with terms on the two lists do not coincide, though the rankings are not greatly changed.

Measuring word frequencies in texts against a reference lexicon

In describing word choice in a text, we analysed the speaker's choice from the 10,000 most common types and from the rest of the lexicon as well. The programmes by which these analyses were carried out examine the terms in a sample text (usually 1,000 tokens) and determine how often each of the 10,000 most common types in English was used. Several statistical measures describe the complex pattern of word choice. The first is the sample text's median word rank derived by arraying all of the text's terms according to their frequency of occurrence and determining the rank (from the Carroll *et al.* list) of the term standing at the midpoint. For example, in a conversation between a mother and child, if the median word in the mother's text is ranked 420, then half of her text consisted of more common words, and half, less common.[2] The second measure is the proportion of the sample text's tokens coming from what we call the Basic Lexicon: the 5,000 most common types in English. The third measure describes how many rare terms were used in that text (per thousand tokens). A rare term is one that does not appear among the 10,000 most common types, is not a proper name or number, and is not an inflected form of some term included among the first 10,000.

The most comprehensive descriptor of word choice is a cumulative curve (Fig. 1). It shows how often the speaker used the 10,000 most common terms in a sample text. The horizontal axis arrays those terms from left to right according to the words' ranks on the Carroll *et al.* list. That dimension is expressed on a logarithmic scale making it possible to examine how the high-frequency closed-class terms were used, as well as to picture the full range of 10,000 types in a condensed form. *The*, the first ranked type, is at the left margin; the 10,000th term is at the right margin. The vertical axis represents the cumulative proportion of tokens in a sample text. The point at which the curve intersects with the left vertical axis corresponds to that proportion of a text's tokens which is accounted for by the single word 'the'. The point at which the curve intersects with the right vertical axis corresponds to the cumulative proportion accounted for by all of the 10,000 most common types. If the intercept at the right edge is lower than 100%, words which are ranked beyond the 10,000th type were used in that text. Conversations generally intercept near 100% (reflecting their heavy reliance on common words); popular television shows generally intercept somewhat lower; and printed matter generally intercepts at the lowest point (reflecting their greater use of uncommon words) (Hayes 1986*a*).

[2] In calculating the median word rank of a sample text, the most common 75 types (nearly all function words) are excluded. All other measures of lexical pitch are based on all tokens.

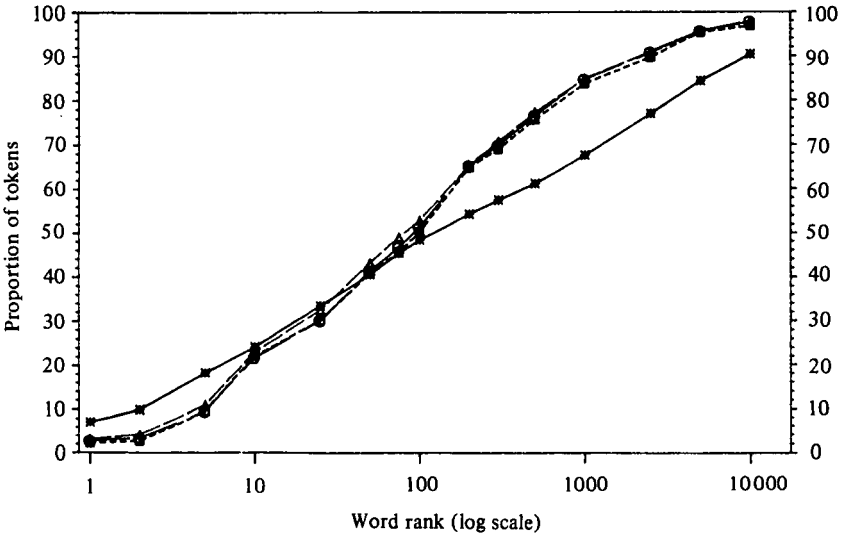

Fig. 1. Patterns of word choice: adults to children (cumulative proportions of tokens in texts). Square = adults to children (newborns to 1;11). Circle = adults to children (2;0–5;11). Triangle = adults to children (6;0–12;11). Star = newspapers.

Comparative data for interpreting these results

In order to interpret the statistics on adult–child speech, a set of comparative statistics covering a broad range of written and spoken texts is provided in Table 2. It is apparent that texts differ widely in their median word ranks, their use of the 5,000 most common terms and their number of truly rare words. This continuous variation in word choice is named LEXICAL PITCH, as in the expression, 'I had to pitch my remarks at a low (or high) level to match my audience's weak (or strong) knowledge of the subject (or language abilities).'

Texts expressed in the highest lexical pitch are all written. In this corpus, the highest-pitched text sample consists of eight abstracts of scientific papers (median word rank 4,389; 70% of the tokens from the Basic Lexicon; and 128 rare terms per 1,000 tokens). The lowest lexical pitch among written texts is found in books designed to be read to preschool children (median word rank 578; 94% from the Basic Lexicon; 16 rare words per 1,000 tokens). For speech texts, the highest pitch is in the testimony of expert witnesses (median word rank 1,008; 90% in the Basic Lexicon; 28 rare words per 1,000 tokens). Pitch in spontaneous conversations between adults is low (median word rank of 496; 94% in the Basic Lexicon; 17 rare words per 1,000 tokens).

Compared with lexical choice in virtually all forms of writing, spoken language, even that between college graduates, is expressed in low lexical

TABLE 2. *Selected statistics for major sources of spoken and written language* (Sample means)

	Number of 1,000-word texts	Rank of median word	Proportion of text from 5,000-word Basic Lexicon	Number rare words per 1,000 tokens
I. Printed texts				
a. Abstracts of scientific articles	1	4389	70·3	128·0
b. Newspapers	17	1690	84·3	68·3
c. Popular magazines	13	1399	85·0	65·7
d. Adult books	19	1058	88·4	52·7
e. Comic books	13	867	88·6	53·5
f. Children's books	21	627	92·3	30·9
g. Preschool books	21	578	94·1	16·3
II. Television texts				
a. Popular prime-time adult shows	17	490	94·0	22·7
b. Popular prime-time children's shows	15	543	93·3	20·2
c. Cartoon shows	14	598	91·7	30·8
d. Mr Rogers and Sesame Street	2	413	97·3	2·0
III. Adult speech				
a. Expert witness testimony	17*	1008	89·9	28·4
b. College graduates to friends, spouses	30*	496	93·9	17·3

* Some of these source texts contain fewer than 1,000 tokens.

pitch. Given this fact, it is expected that adult–child conversations will be even lower pitched. If the age-dependent simplification pattern with grammar can be generalized to lexical choice, then it is expected that: (*a*) a text's median word rank will be lower in speech with younger children than with older children; (*b*) the proportion of a text absorbed by the 5,000-type Basic Lexicon will be greater with younger children; (*c*) there will be fewer rare words used with younger children; and (*d*) adult use of the 10,000 most common types in English will be most skewed towards the most common types in speech with infants and least skewed with school-age children.

RESULTS

Replication of previous findings with these data: MLU and TTR

These analyses are designed to replicate well-known patterns of text simplification in adult speech to children to ensure that these samples are not atypical in some way relevant to language simplification. In the 47 cross-

sectional samples of adult-to-child conversations with newborns to twelve-year-olds, adult mean length of utterance (in words) increased with the child's age ($r = +0{\cdot}55$). In the longitudinal case, the mother's correlation is much smaller ($r = +0{\cdot}23$), but her children represented a much narrower age range – 45 days to five years. In child-to-adult texts, covering children ranging in age from 4 years to 12 years, the correlation of child's MLU and child-speaker's age is $r = +0{\cdot}47$. And, in independent samples having at least 1,000 tokens, the correlation of the type/token ratio (based on exactly 1,000 tokens) with child's age is $r = +0{\cdot}76$ in adult-to-child speech, and $r = +0{\cdot}83$ for children talking to adults.

By the standards of these two familiar indices of language simplification, these conversation samples between adults and children exhibit the expected age-dependent associations: fewer words per utterance and fewer different types used to and by the younger children. Our correlations are higher than generally reported, probably for two reasons: our child age range is broader than the restricted range most investigators use; and the mean number of tokens in our samples is larger than in most studies of children's language, making our language measures more stable.

Word choice in adult speech with children. All adult-to-child texts were divided according to the age of the child into three sub-sets representing three stages of language development: (*a*) infants – newborns up to two years; (*b*) preschoolers – two to six years; and, (*c*) school-age children – six to twelve years. The statistical results are shown in Table 3.

The first comparison of these three sub-sets is by the rank of their median word. The mean of the median word rank for the eight samples of adults talking to infants is the 434th-ranked word on the Carroll *et al.* list (S.D. = 156). The mean for the twenty conversations with preschool children is lower, word rank 390 (S.D. = 89), not higher as expected. For the thirteen texts of adults speaking with school-age children, the mean is word rank 418 (S.D. = 43). The correlation of median word rank with child's age across these forty-one conversations is $r = +0{\cdot}04$. Despite this negligible association, there is an interesting age-dependent pattern for this measure. The variability in adults' median word ranks, indicated by standard deviations, lessened considerably with children's increasing ages; for this measure, the adults in this corpus were more consistent when speaking with older children.

These data are compatible with the simplification generalization in this one important respect: the average median word rank in all three sets of adult-to-child conversations is considerably lower than the 496th rank found in adult-to-adult conversations. Adults simplified vocabulary by skewing word choice towards the more common terms when speaking with children in general. The results are, however, incompatible with the expected age-dependent pattern of word choice simplifications.

TABLE 3. *Comparative statistics on conversations and newspapers : means for each set of samples*
(Standard deviations in parentheses)

	No. of samples	Rank of median word	Proportion of text from 5,000 word Basic Lexicon	Number rare words per 1,000 tokens
I. Children talking to adults				
a. Child 4;0–5;11	9	330 (56)	95·2 (1·9)	6·6 (3·0)
b. Child 6;0–12;11	9	351 (66)	95·3 (2·1)	10·7 (4·4)
c. All child speakers	18	341 (60)	95·3 (2·0)	8·6 (4·2)
II. Adults talking to				
a. Infants (newborns–1;11)	8	434 (156)	95·0 (2·7)	9·3 (4·5)
b. Preschool children (2;0–5;11)	20	390 (89)	95·6 (1·7)	9·0 (6·7)
c. School-age children (6;0–12;11)	13	418 (43)	95·7 (1·0)	11·7 (4·7)
d. A longitudinal case	11	426 (22)	96·2 (0·9)	9·9 (2·8)
e. All adults-to-child	47[a]	412 (88)	95·6 (1·7)	9·9 (5·4)
III. Adults to adults	30	496 (119)	93·9 (1·6)	17·3 (8·9)
IV. Large-circulation newspapers	17	1690 (489)	84·3 (3·0)	68·3 (14·9)

[a] The cross-sectional samples of adults talking with children (IIa, IIb and IIc) are separated, in these analyses, from the longitudinal case (IId). The number in the combined sample (47 texts), does not equal the total of IIa–d because the longitudinal analysis (IId) is based on all 11,000 tokens of what that mother said to her three children in both years. When this longitudinal case was added to IIa–c, only the means of her multiple samples with each child, each year were included.

The second comparison of adult speech to these three sets of children – the proportion of their word choices taken from the 5,000-type Basic Lexicon – follows the same pattern. The correlation between child's age and adult use of Basic Lexicon terms is $r = +0{\cdot}02$.

The expectation that adults will use more rare words with older children than with infants and preschoolers is also contradicted by the evidence. While adults used 17 rare words per thousand tokens when speaking with other adults, they used 9 with infants, 9 with preschool children and 12 rare words per thousand tokens with school children. Again, adults simplified lexical choice for children, but not in an age-dependent manner ($r = +0{\cdot}25$, $p > 0{\cdot}05$).

The full extent of the similarity of adult word choice for infants, preschoolers and school-age children is shown in Fig. 1. This describes three curves corresponding to adult speech to each of the three sub-sets of children. The curves are nearly identical. Adult word choice to infants,

preschoolers and school children was essentially the same at every point across the full 10,000 most common types in English. Given the differences in language development for infants, preschool and school-age children, the similarity in their curves is remarkable; it extends even to the points in the curves where there are abrupt changes.

The three S-shaped curves of adult-to-child speech in Fig. 1 are contrasted with the linear pattern of word choice in 17 British and American newspapers to show how far word choice in speech departs from that used in written texts expressed in formal style. The programmes which produce these analyses permit us to identify precisely which words are used differently in the S-curved adult-to-child texts from their use in linear texts. The first bend in the curves has been traced to the systematic under-use of the two most common articles (*the* and *a*), the most common prepositions (*of* and *in*), and the most common conjunction (*and*), and to the systematic over-use of *you*, *that* and *it*, compared with their use in linear texts. Together, this differential pattern of word use – a shortfall for words ranked 1 to 7, and greater use of words ranked 8 to 10 – produces the first bend of the S-curved pattern. The proportionately higher use of the major function words, beginning around the eighth-ranked term, continues through the first 100 terms (dominated by closed-class terms), and all of the most common content terms. Usage of the words between the 500th- and 2,500th-ranked terms falls off in conversation, dropping below their use in newspapers, producing the second bend in the S-curve. The point in the lexicon where conversations and printed texts differ most is near word rank 1,000: approximately 83% of the tokens in adult-to-child conversation are drawn from the 1,000 most common types, while 68% of the tokens in newspapers come from that same small portion of the lexicon.

Comparing child-to-adult speech with adult-to-adult speech

As would be expected, lexical pitch in speech by four- to six-year-olds was quite low. That of six- to 12-year-olds was somewhat higher, but also low by adult standards (median word ranks of 330 vs. 351, respectively). All measures of lexical pitch for children's speech are well below the levels found in parents', teachers' and nurses' speech to children.

When college graduates engaged in natural conversation, their pitch levels were higher than adult-to-child levels, but comparatively low when contrasted with the most widely read forms of writing. In adult speech with other adults, the median word rank is 496; the number of rare words, 17 per thousand tokens.

When children's and college-graduates' speech patterns are compared in Fig. 2, their S-curved word choice patterns are similar across the full 10,000 most common types of English, both standing in contrast to the linear pattern

of newspapers. The adults in these samples did make slightly greater use of the principal articles and *of* than the children, resulting in a slightly shallower first bend in the S-curve, and used the balance of function words and the major content words somewhat less, leaving a slightly less bowed second bend in their curve. Given the striking differences in the sizes of young children's and college-educated adults' lexical repertoires, the absence of a substantial difference in their productive word choice patterns during natural conversation, in a corpus this large, is surprising.

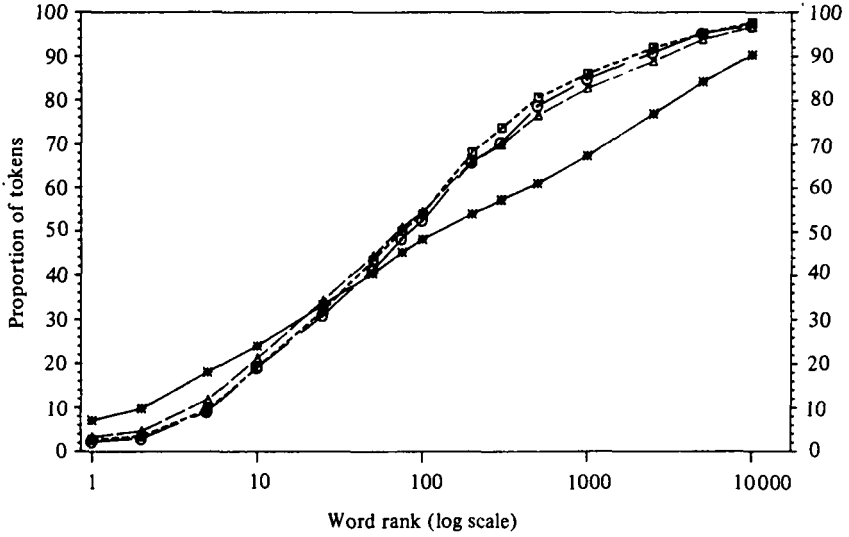

Fig. 2. Patterns of word choice: children to adults and adults to adults (cumulative proportions of tokens in texts). Square = children (4;0–5;11) to adults. Circle = children (6;0–12;11) to adults. Triangle = adults to adults. Star = newspapers.

A close examination of Figs 1 and 2 shows that all six curves (those for children to parents; adults to infants, preschool and school-age children; and adults-to-adults) closely resemble one another. Not only do they all have the general S-curved word choice pattern, but the curves differ little at any point across the 10,000 most common types. Their similarity contradicts the expectation that there is an age-dependent pattern of word choice in speech to children. At the same time, they confirm the robustness of the S-curved pattern and indicate the power of the unidentified processes causing word choice in conversations to differ from that in written texts.

Resolving the contradiction between these results and the TTR results

An apparent contradiction in these data can now be resolved. On the one hand, there is a strong age-dependent pattern in the type–token ratio. Adults

used an average of 246 types per thousand tokens for infants, 297 for preschool children and 331 for school-age children, suggesting that additional types are drawn increasingly from the pool of less common words as children grow older. Contradicting that suggestion, we find the distributions of adults' word choices from the 10,000 most common types were the same for all three age groups. For example, the number of rare words rose trivially for the older children, from nine per thousand tokens (for infants) to 12 (for school-age children). While adults increased the proportion of different types as children grew older, they did not draw more heavily upon uncommon words for the older children in spontaneous conversations in their natural contexts (Fig. 1).

These findings should make investigators cautious in interpreting TTR values. One cannot assume that a larger TTR implies that the additional words used are more uncommon, complex or demanding. In this large corpus of natural language, a higher TTR did not result in a higher median word rank, a lower proportion of Basic Lexicon words used or an increase in the proportion of rare words.

A longitudinal case

One mother of three (an infant son, a two-year-old boy and a four-year-old daughter), having already recorded a four-hour period of interaction with her children, agreed a year later to a second recording. As in the corpus generally, this mother strongly raised her TTR ($r = +0{\cdot}88$) and slightly increased sentence length for her older children ($r = +0{\cdot}23$). Her lexical pitch rose as her children grew older in all three measures of word choice, but none rose significantly: (1) median word rank ($r = +0{\cdot}26$); (2) the per cent of tokens coming from the 5,000-type Basic Lexicon ($r = -0{\cdot}26$); and, (3) the number of rare words ($r = +0{\cdot}21$). This college-educated mother may have accommodated her word choice to her children's language development, but these accommodations were negligible.

In conclusion, in both the larger corpus of cross-sectional samples and in this single longitudinal case, word choice is weakly correlated with or unrelated to children's ages – a finding contrary to the well-established simplifications found with grammar.

DISCUSSION

The results show that adult word choice for children across the wide age-range – newborns to 12 years – was not age-dependent in three measures of lexical pitch. Adults in this corpus did not adjust their selection from among the common and uncommon terms to fit the children's ages. Rather, they lowered lexical pitch for all children by somewhat less than 20% from the

adult–adult speech standard. The question of why distinctive age-dependent patterns are not found in this corpus of adult-to-child speech follows from the larger question of why the adult–child speech patterns are not more markedly different from adult–adult speech patterns.

One explanatory factor stems from the observation that the content of adult conversations is predominantly mundane, focusing on household, school, work and interpersonal matters, and is expressed in informal style. The effect of such ordinary, everyday topics is to concentrate word choice on the most common words at the expense of the specialized, uncommon terms of the lexicon.

A second reason why adult–child and adult–adult word choice patterns differ so little relates to access to common and rare words. Research on lexical access (Just & Carpenter 1984) has demonstrated that recognition time for uncommon words is longer than for common words. By extension, retrieval of infrequently used, uncommon words from memory for use in spontaneous conversation should take longer than retrieval of frequently used, common words. Efforts to keep the speech stream flowing smoothly may force speakers to draw upon those terms which come readily to mind – and they are the most commonly used words of the lexicon.

The joint effects of the mundane nature of conversations and the faster retrieval time for common words leaves little room for further substantial lexical simplifications. Although it appears that the mechanisms by which adults make language adjustments in accommodating to their younger interactive partners (so effective in simplifying grammatical constructions), are either weak for word choice or different, it may be, instead, that because lexical pitch is already so low in casual conversations, extensive effort would be required to lower it any further – effort not conducive to the real-time, spontaneous nature of natural conversation.

Implications for children's lexical acquisition

Current thinking on child language development stresses the child's active role in shaping its own experience by selecting from among potential experiences and, by its own speech and behaviour, shaping others' behaviour towards the child (Bell & Harper 1977, Gleason 1977, Scarr & McCartney 1983). The absence of age-dependent lexical accommodations to children suggests that the lexical behaviour of adults is not as much under the child's control as the strong and consistent evidence of grammatical adjustments implies.

If parents and other adults are not lexical 'chameleons', systematically increasing the use of uncommon words as children grow older, where are children to encounter instances of the 600,000 rare types of English which lie beyond the 10,000 most common types? They cannot broaden their rec-

ognition and productive vocabularies without substantial experience with these concepts and with their word-names. Table 2 provides a comparison of the relative frequency with which rare words appear in the several major sources of a child's language experience. It shows that, like conversation, popular TV shows are a poor source from which children can differentiate meanings and uses of rare words. School books increase the use of these rare words tenfold between the first and sixth grades (Hayes 1987), but the child must become literate and devote time and effort to reading if it is to profit from those sources. A child's acquisition of knowledge about rare words from school books has been shown by Dreeben & Gamoran (1986) to depend strongly on two factors: the relative frequency of such words in basal readers and the number of minutes of actual reading instruction a child receives each day. By extension, the time a child spends reading books, including comic books, is time lexically well-spent.

In conclusion, a child whose language 'diet' is largely restricted to natural conversation and popular television would encounter few instances of the terms which make up the great majority of items used on standardized achievement tests and the most common intelligence tests to assess children's verbal progress and determine academic placement. Unless children have extensive experience with written language sources, by having been read to and by reading on their own, their knowledge of vocabulary beyond the 5,000-type Basic Lexicon will grow slowly, for there is no other way for this kind of knowledge to be acquired but through contextually rich lexical experiences.

SUMMARY

A corpus of over 100,000 tokens from natural conversations was examined for evidence that lexical choice follows the 'motherese' pattern of age-dependent simplifications in grammar. We found that adult word choice in speech to children was, as expected, skewed towards the more common words, compared with their speech to adults. Second, however, contrary to the well-established pattern of age-dependent grammatical simplifications, lexical simplification was not age-dependent. The correlation between a child's age and the adult speaker's median word rank was $r = +0{\cdot}04$; between age and the percentage of a text's tokens drawn from the 5,000-type Basic Lexicon, $r = +0{\cdot}02$; and between age and the number of rare words (per thousand tokens), $r = +0{\cdot}25$. Third, the patterns of adult word choice for children closely matched the pattern in adult-to-adult speech. Differences between speech directed to children and to adults were trivial and dwarfed by the magnitude of the differences between those conversations and the text of the daily newspaper. Fourth, children chose words in a pattern closely resembling that of adults. Fifth, the differences between linear and S-curved

patterns of word choice reflect differences in the nature of written and spoken language. The many factors which contribute to the spontaneous nature of conversations between family members and friends work together to keep them lexically undemanding. These include the preponderance of mundane topics, the use of informal speech style and limitations on the time to retrieve rare words. Such factors are either overridden or less applicable in written language. Since lexical input from conversation is little geared to children's developing word knowledge (as shown here), and mundane conversation is so limited a source for learning about words outside the 5,000 most common terms, further development of lexical knowledge requires literacy and extensive reading across a broad range of subjects.

REFERENCES

Bell, R. Q. & Harper, L. V. (1977). *Child effects on adults*. Hillsdale NJ: Erlbaum.

Carroll, J. B., Davies, P. & Richman, B. (1971). *Word frequency book*. Boston: Houghton-Mifflin.

Clark, E. V. (1979). Building a vocabulary: words for objects, actions and relations. In P. Fletcher & M. Garman (eds), *Language acquisition : studies in first language development*. Cambridge: C.U.P.

Cross, T. G. (1977). Mothers' speech adjustment: the contribution of selected child listener variables. In C. Snow & C. Ferguson (eds), *Talking to children*. Cambridge: C.U.P.

Dreeben, R. & Gamoran, A. (1986). Race, instruction, and learning. *American Sociologial Review* **51**. 660–9.

Ferguson, C. (1964). Baby talk in six languages. *American Anthropologist* **66**. 103–14.

—— (1977). Baby talk as a simplified register. In C. Snow & C. Ferguson (eds), *Talking to children*. Cambridge: C.U.P.

Giles, H., Mulac, A. & Johnson, P. (1986). Speech accommodation theory: the first decade and beyond. In M. McClough (ed.), *Communication yearbook*. Beverly Hills: Sage.

Gleason, J. B. (1977). Talking to children: some notes on feedback. In C. Snow & C. Ferguson (eds), *Talking to children*. Cambridge: C.U.P.

Gleitman, L. A., Newport, E. L. & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language* **11**. 43–79.

Hayes, D. P. (1986*a*). The Cornell Corpus. Technical Report Series 86-1. Ithaca, NY: Department of Sociology, Cornell University.

—— (1986*b*). User's guide to LEX. Technical Report Series 86-2. Ithaca, NY: Department of Sociology, Cornell University.

—— (1986*c*). Validation analyses of lexical pitch measures. Technical Report Series 86-3. Ithaca, NY: Department of Sociology, Cornell University.

—— (1987). Speaking and writing: distinct patterns of word choice. Technical Report Series 87-3. Ithaca, NY: Department of Sociology, Cornell University.

Hayes, D. P., Ahrens, M. G. & Tsay, R. (1987). Tight and loose coupling: effects on accommodation in conversation. Technical Report Series 87-10. Ithaca, NY: Department of Sociology, Cornell University.

Just, M. A. & Carpenter, P. A. (1984). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.

Newport, E. L., Gleitman, H. & Gleitman, L. R. (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In C. Snow & C. Ferguson (eds), *Talking to children*. Cambridge: C.U.P.

Oksaar, E. (1981). *Language acquisition in the early years*. New York: St Martin's Press.

Scarr, S. & McCartney, K. (1983). How people make their own environments: a theory of genotype–environment effects. *Child Development* **54**. 424–35.

Schachter, F. F. (1979). *Everyday mother talk to toddlers: early intervention*. New York: Academic Press.
Snow, C. (1977). Mothers' speech research: from input to interaction. In C. Snow & C. Ferguson (eds), *Talking to children*. Cambridge: C.U.P.