METHODS & DESIGNS

English palindromes: A distributional model

GREGORY V. JONES

University of Bristol, Bristol BS8 1HH, England

A list of English palindromes that approaches comprehensiveness is given. A stochastic model for the distribution of palindromes within the language is proposed. The model relates frequencies of palindromes and of heteropalindromes (Jones, 1980) as a function of word length, and it is shown to predict accurately the actual numbers of palindromes that occur.

In a recent article (Jones, 1980), the term "heteropalindrome" was introduced to denote symbols that form different words when read forward and backward (e.g., "deliver" and "reviled"). A corpus of such words was provided in that publication in order to assist research into memory retrieval (Jones, 1978) and reading processes (Asso & Wyke, 1967; Shankweiler & Liberman, 1972; Wolfe, 1939). The present article complements the earlier publication (Jones, 1980) by providing a list of English palindromes (e.g., "deified"), which should be of use in the same areas. In addition, a simple stochastic theory of the distribution of palindrome frequency as a function of word length is presented and shown to predict accurately the observed frequencies.

METHOD

The method employed in compiling the corpus of palindromes was similar to that used for the heteropalindromes (Jones, 1980). Principal systematic aids included exhaustive mental permutating, consideration of English phonological and morphemic structure, and dictionary consultation (*Chambers Twentieth Century Dictionary*, 1962, was used). Proper names such as "Hannah" and "Malayalam" were not included.

The resulting corpus of 81 palindromes is listed in Table 1.

DISCUSSION

An interesting aspect of the corpus emerges when the distribution of the number of palindromes of different word lengths is considered. The bottom row of Table 2 gives this observed distribution, together with the corresponding one for heteropalindromes in the top row. The two distributions appear different in form: The palindromic distribution is sawtoothed, and the hetero-

Table 1
List of English Palindromes

Three-Letter Words			
aba	dad	gag	pip
aga	did	gig	pop
aha	dod	gog	pup
ala	dud	hah	tat
ana	eke	mum	tit
ava	ere	nun	tot
bib	eve	oho	tut
bob	ewe	pap	wow
bub	eye	pep	zuz
Four-Letter Words			
abba	deed	noon	poop
anna	esse	otto	sees
boob	kook	peep	toot
Five-Letter Words			
alula	kayak	refer	shahs
anana	level	rotor	siris
civic	madam	sagas	solos
deled	minim	seles	stets
dered	radar	seres	stots
dewed	reder	sexes	tenet
Six-Letter Words			
denned	marram	redder	terret
Seven-Letter Words			
deified	repaper	reviver	rotator
Nine-Letter Word			
redivider			

Table 2
Distributions of Palindromes and Heteropalindromes

Type	Word Length						Total
	2	3	4	5	6	7+	
OH	6	72	124	62	22	4	290
PP	2.45	30.59	11.14	28.39	4.69	7.21	84.47
OP	0	36	12	24	4	5	81

Note—Type refers to distribution type: OH = observed heteropalindromes; PP = predicted palindromes; OP = observed palindromes.

Requests for reprints should be sent to Gregory V. Jones, Department of Psychology, University of Bristol, 8-10 Berkeley Square, Bristol BS8 1HH, England.

palindromic is unimodal. A simple model accounts for this qualitative observation and predicts quantitatively the detailed relation between the two distributions, as follows.

STOCHASTIC MODEL

Notation

Notation used is as follows: n = length of word (two to seven letters); x_n = total number of words of length n ; A = effective alphabet size.

Distributions

It is assumed that the property of being a palindromic or a heteropalindromic word arises randomly among all words, as follows.

Even-length palindromes. The probability of a single word of length n having one of its halves as a particular alphabetic sequence (specified as the reverse of its other half) is $1/(A^{n/2})$. Hence, the expected total number of such palindromes is

$$P_n = x_n/A^{n/2}. \quad (1)$$

Odd-length palindromes. Here, the middle letter of the word is not constrained by symmetry, and hence,

$$P_n = x_n/A^{(n-1)/2}. \quad (2)$$

Heteropalindromes. The probability that one of $(x_n - 1)$ other words of length n will have a particular alphabetic sequence (specified as the reverse of a single given word) is $(x_n - 1)/A^n$. The expected total number of individual heteropalindromic words (i.e., counting the forward and backward forms as two separate words) will be

$$H_n = x_n(x_n - 1)/A^n, \quad (3)$$

which is well approximated by

$$H_n = x_n^2/A^n. \quad (4)$$

Combining Equations 1 and 4, it follows that

$$P_n = \sqrt{H_n}. \quad (5)$$

for n even. Combining Equations 2 and 4,

$$P_n = \sqrt{AH_n} \quad (6)$$

for n odd.

Predictions

In principle, Equations 1-3 could be used to predict absolute values for the distributions of palindromes and heteropalindromes. However, I am not aware of appro-

priate tabulated values of x_n , the total number of words of length n . Dictionary-based compilations omit many derivatives of listed words, including the very large numbers of noun plurals and verb third-person singulars with suffix "s." Text surveys (in particular, Table C1 of Kučera & Francis, 1967) include derivatives but, for the present purpose, overestimate due to the inclusion of items that are not strictly words (e.g., Kučera & Francis found 277 two-symbol strings) and underestimate due to necessarily limited sampling (e.g., Kučera & Francis found 50,406 strings altogether, a number approximately equal only to that of an ordinary English dictionary's main words).

The problem of estimating the distribution of x_n may be finessed by considering instead Equations 5 and 6, which do not involve x_n . Table 2 shows observed numbers of palindromes and heteropalindromes (the latter taken from Jones, 1980), together with the numbers of palindromes predicted on the basis of Equations 5 and 6. For words of odd lengths ($n = 3, 5, 7$), Equation 6 requires a value for the effective alphabet size, A . This would be 26 only if all letters were equally likely, which is clearly not the case. A better approximation is that the effective alphabet size at any point consists of either the five vowels or the 21 consonants (or in fact any other dichotomy), yielding an average effective alphabet size half that of the full alphabet, 13. Table 2 shows the complete theoretical palindromic distribution with the value of A set at 13 for the odd-length words. There is very good agreement with the observed distribution, as statistical comparison confirms [$\chi^2(6) = 4.93$].

The qualitative nature of the model's representations may also be considered. The unimodal distribution of heteropalindromes can be seen to arise as the product of two opposing trends. According to the results of Kučera and Francis (1977, Table C1), the frequency of words of different lengths, x_n , rises monotonically over the range of lengths considered here, peaking in fact at $n = 7$; the likelihood of a particular string of letters, on the other hand, decreases monotonically with its length. For palindromes, however, the pattern is different. There is, as before, a monotonic positive trend due to increasing values of x_n . But the negative trend due to the increasing constraints on letter sequences is nonmonotonic, the constraints being equal for n and $n + 1$, for each even n . This combination of monotonic positive trend and stepped negative trend produces the observed sawtooth function.

Finally, note that Equations 5 and 6 can be tested in two further ways. First, they should hold for the increased number of palindromes and heteropalindromes that would be evident if reference were made to *The Oxford English Dictionary* (Murray, Bradley, Craigie, & Onions, 1884-1928) rather than to *Chambers Twentieth Century Dictionary*. Second, it may be tested whether the equations hold for languages other than English.

REFERENCES

- ASSO, D., & WYKE, M. Experimental study of the effects of letter reversals on reading. *British Journal of Psychology*, 1967, **58**, 413-419.
- Chambers twentieth century dictionary (rev. ed.). Edinburgh, Scotland: Chambers, 1972.
- JONES, G. V. Recognition failure and dual mechanisms in recall. *Psychological Review*, 1978, **85**, 464-469.
- JONES, G. V. Heteropalindromes. *Behavior Research Methods & Instrumentation*, 1980, **12**, 393-394.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R.I: Brown University Press, 1967.
- MURRAY, J. A. H., BRADLEY, H., CRAIGIE, W. A., & ONIONS, C. T. *A new English dictionary on historical principles*. Oxford, England: Oxford University Press, 1884-1928.
- SHANKWEILER, D., & LIBERMAN, I. Y. Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, Mass: M.I.T. Press, 1972.
- WOLFE, L. S. An experimental study of reversals in reading. *American Journal of Psychology*, 1939, **52**, 533-561.

(Received for publication September 5, 1980;
accepted September 7, 1980.)