J. Linguistics 26 (1990), 471-482. Printed in Great Britain

# NOTES AND DISCUSSION

# Elizabeth and John: sound patterns of men's and women's names<sup>1</sup>

# ANNE CUTLER, JAMES MCQUEEN AND KEN ROBINSON

# MRC Applied Psychology Unit, Cambridge, U.K.

(Received 11 December 1989; revised 20 February 1990)

He saw people...as careless and as indifferent about the name they imposed upon their child, or more so, than in the choice of *Ponto* or Cupid for their puppy dog.

# (Laurence Sterne, The life and opinions of Tristram Shandy, *Gentleman*)

### 1. INTRODUCTION

In one respect parents are never careless or indifferent about naming a child: if the child is a girl, they choose a girl's name, but if it is a boy, they choose a boy's name. Euphony, cultural coherence, fashion and family tradition all play a role in the choice of a name; but such factors only determine selection within the appropriate sex-marked set. With very few exceptions, English first names for men and for women do not overlap, and only custom enables us to know which is which. Although in many languages male and female names are distinguished by characteristic suffixes or other sound patterns, this is not in general true in English. There are semi-productive processes which derive female names from male names, but the majority of names are not derived, and there are no linguistic principles we can call upon to determine the sex of John versus Jane, Mary versus Gary, Kevin versus Karen.

There are, however, certain predictions one might make about the phonological form of English names in general. These predictions are derived from the structure of the English vocabulary as a whole. English consists predominantly of words beginning with strong syllables (Cutler & Carter, 1987), and the most common word pattern in English is a bisyllable with initial stress, such as common, pattern, English (Carlton, Elenius, Granstrom & Hunnicutt, 1985). Monosyllables are almost as frequent as bisyllables. Among polysyllables, 70% begin with a strong (primary or secondary

[1] The authors contributed equally to this work and the order of authorship is alphabetical. We are grateful to Jim Sholicar and Ted Briscoe for providing the noun statistics from Longmans Dictionary, and they acknowledge the Longman Group Inc. for allowing access to the machine-readable version of the dictionary. We further thank Ian Nimmo-Smith for statistical advice.

stressed) syllable, only 30% with a weak syllable (Cutler & Carter, 1987). Moreover, words with strong initial syllables are used more often than words with weak initial syllables; Cutler and Carter (1987) found that in a 190,000 word corpus of spontaneous conversation, 90% of the lexical words (nouns, verbs and adjectives) were monosyllables or polysyllables with strong first syllables. Thus although 27% of the English vocabulary consists of polysyllables with weak first syllables, only 10% of the words in this spoken sample belonged to this class.

In English, this bias towards strong initial syllables is exploited by listeners: there is a preference towards segmenting speech at the onset of strong syllables but not segmenting it at the onset of weak syllables. Misperceptions of word boundary location are more likely to result in erroneous insertion of a boundary before a strong syllable but erroneous deletion of a boundary before a weak syllable (Butterfield & Cutler, 1988). Detection of a word embedded in nonsense is slower if the word is followed by a strong vowel so that the final segment could form a potential strong syllable onset - so *mint* is harder to detect in *mintayf* than in *mintef*, in which the second syllable is weak (Cutler & Norris, 1988). (In this case, the listener's strategy of segmenting speech at the onset of each strong syllable interferes with performance of the word detection task.) Spoken utterances are continuous, and segmentation into individual words is far from a trivial matter;<sup>2</sup> but because of the bias in the vocabulary, a strategy of segmenting continuous speech at strong syllable onsets is extremely efficient at detecting the actual onsets of lexical words in typical English speech, and experimental evidence shows that listeners do indeed use such a strategy. Names, it is true, constitute in some ways a special subset of the lexicon. In certain languages they are exempt from some morphophonological rules (Comrie, 1979); language impairment as a result of brain damage can result in selective loss of the ability to produce proper names while the remaining lexical stock is unimpaired (Semenza & Zettin, 1988), or selective retention of names although the remainder of the lexicon is lost (Van Lancker & Klein, 1990); memory for proper names is less reliable than for comparable material (Cohen, 1989). But they function as nouns, and should presumably resemble the rest of the vocabulary in phonological form; and since they are often unpredictable in context, it is important that they should lend themselves well to segmentation from running speech. Thus we would expect that, like most of the rest of the English vocabulary, names should be highly likely to begin with a strong syllable. Furthermore, there is no reason to expect them

[2] Pauses at word boundaries, for instance, occur rarely, unless the boundary is also a syntactic break (Grosjean, Grosjean & Lane, 1979). Phonotactic sequencing constraints also provide little assistance, since most cross-boundary sequences could be word-internal (in fact, a strong syllable segmentation algorithm performs better with IMPERFECILY SPECIFIED phonetic input than a phonotactic constraint algorithm does with fully specified phonetic strings; Harrington, Watson & Cooper, 1989).

to differ systematically from the rest of the vocabulary in length (number of syllables). These predictions can be easily tested by comparing a dictionary of names with a dictionary of English words.

Nothing in the above discussion would lead us to expect differences between male and female names on either stress pattern or number of syllables. Since the two sets are virtually non-overlapping, however, it is simple to compare each set separately with the set of nouns. In the next two sections we report analyses of stress patterns and length in syllables for male names, female names and English nouns. As will become clear, the strongest finding of all is that there are in fact significant differences between male and female names on each parameter.

#### 2. STRESS PATTERNS

We categorized the names in *The Oxford minidictionary of first names* (1986), which contains 1667 entries (884 female names, 783 male names).<sup>3</sup> This was then compared with the nouns in the English vocabulary (specifically, 19,334 head nouns, having one to five syllables, in the *Longman's dictionary of contemporary English*). Because names properly belong to the class of nouns, and nouns tend to show a preference for strong initial syllables even more strongly than any other word class, it is appropriate to compare names with nouns rather than with the entire vocabulary, since significant differences from the vocabulary pattern might simply be due to conformity to the noun pattern. Table 1 shows the proportion of nouns, male names and female names with strong class). In all three sets, strong initial syllables predominate; but while in the noun set 85 % of initial syllables are strong, for male names this proportion rises to 95 % while for female names it drops to 75%.

Averaged across male and female names, 84 % of the names have strong initial syllables, which is virtually identical to the proportion among the nouns. But when the male and female subsets are compared with the nouns separately, the differences between the three sets are statistically significant. For male names, the tendency towards strong initial syllables is significantly

stronger than for nouns  $(X^2[1] = 59.3, p < 0.001)$ . For female names, the tendency towards strong initial syllables is significantly WEAKER than for

[3] Names in this dictionary are listed with IPA transcriptions. We counted all names which appeared as head words, and based all our analyses on the pronunciation and stress pattern given in the dictionary. Where a name was given alternative pronunciations or stress patterns (e.g. *Aileen*), the name was counted twice. The category 'strong' included all initial syllables with primary stress (e.g. *Anthony, Andrea*) and secondary stress (*Alexandra, Aloysius*). The category 'weak' included mainly initial syllables containing schwa (*Amanda, Adolphus*), but also unstressed initial syllables containing potentially full vowels (*Antonio, Augusta*).

	Nouns	Male names	Female names
Strong			
Initial syllables	0.847	0.945	0.747
Weak			
Initial syllables	0.153	0.055	0.253

## Table 1

Proportion of nouns, male names and female names with strong versus weak initial syllables

Male Female

Iviaic	I Cillaic
names	names

StrongInitial syllables0.9790.837Weak0.0210.163

### Table 2

Proportion of very frequent male and female names with strong versus weak initial syllables

nouns  $(X^2[I] = 64.4, p < 0.001)$ . In other words, male and female English names show systematic differences in sound pattern: female names are far more likely to have unstressed initial syllables.

This difference is preserved even when we consider only the most frequent names. *Everyman's dictionary of first names* (1987) provides the fifty most popular names for males and females born in 1925, 1950, 1965, 1975 and 1986 in England and Wales, and in 1925, 1950, 1970 and 1986 in the United States. Table 2 shows the proportion of strong and weak initial syllables for names appearing in these lists.<sup>4</sup> No male name popular in Britain in any year has a weak initial syllable, and only three male names in the American lists do. However, there are many consistently popular female names with weak initial syllables (*Elizabeth*, for instance, occurs on eight of the nine lists, *Patricia* and *Michelle* on five, and *Amanda, Rebecca, Teresa* and *Joanne* on four). The difference is highly significant (X<sup>2</sup>[I] = 18.39, P < 0.001).

[4] We counted each name only once, irrespective of how many lists it appeared in

#### 3. NUMBER OF SYLLABLES

Female names are longer than male names. Figure 1 shows the proportion of nouns, male names and female names with one, two, three or four/five syllables (four and five syllables are presented as a single category because there were only four names in the *Minidictionary* - one male and three female - with more than four syllables). The distributions for male and female names are significantly different ( $X^{2}[3] = 123$ , p < 0.001). An analysis of the residuals from the independence model (Cox & Lauh, 1967) allows us to tell where two multi-category distributions differ; in this case the male and female distributions do not differ significantly on the proportion of bisyllables, but males have significantly more monosyllabic names than females and significantly fewer three- to five-syllable names. Both the male and female distributions also differ significantly from nouns: for males,  $X^{2}[3] = 203.87$ , p<0.001, and for females,  $X^{2}[3] = 110.85$ , p < 0.001. Analysis of the residuals shows that both male names and female names have significantly more bisyllabic instances than nouns, and significantly fewer four- to five-syllable instances. However, the two sets differ on monosyllabic and trisyllabic instances. Among male names there are significantly more monosyllables than among nouns, but among female names there are significantly FEWER monosyllables; among male names there are significantly fewer trisyllables than there are among nouns.

The tendency for female names to be longer is even more marked among the most popular names, as Figure 2 shows. The mean number of syllables for British male names across the five listed years is 1.9, 1.88, 1.92, 1.98 and 1.98, for British female names 2.2, 2.25, 2.28, 2.32, 2.31; for American names for the four years the means for males are 1.82, 1.88, 1.88 and 2.02, for females 2.24, 2.19, 2.38 and 2.44. Note that the male and female distributions of means do not overlap - the shortest mean for female names is still longer than the longest mean for male names. Again, the differences are significant  $(X^2[3] = 22.3, p < 0.001)$ .<sup>5</sup>

Our two analyses have, therefore, confirmed that English names are in general quite like English nouns - they are mostly mono- or bisyllabic, and they mostly begin with a strong rather than a weak syllable. But the analyses

have also revealed that, contrary to expectations, there exist systematic differences between male and female names in English. Female names

[5] *Everyman s Dictionary* in fact gives two US lists for 1986: one for white and one for nonwhite births. We have included only the white list. We note here that every tendency which we have observed for male-female differences is echoed in white-non-white differences. The mean number of syllables for non-white male names was 2.12, and for non-white female names 2.6. 42% of the non-white female names began with unstressed syllables, and of all the male lists, this one non-white list contained the largest number of names with unstressed initial syllable (there were four: *Antonio, Demetrius, De Andre and Maurice,* which in the US has WS stress).

Nouns



#### Number of syllables



Male names



Number of syllables



Female names



#### Number of syllables



Figure I

Proportions of nouns, male names and female names with one, two, three and four/five syllables.

are likely to be longer and are more likely to begin with an unstressed syllable.

What could explain this pattern? A partial explanation may seem to lie in the fact that, as mentioned above, there exist semi-productive processes for forming female names from male names by addition of a suffix. But derived names such as *Georgina, Thomasina, Charlene, Maxine* in fact form a very small subset of female names. Moreover, they hardly occur at all in the lists

Male names



#### Number of syllables



Female names





of most popular names. When we analyse only the underived popular female names, the pattern of significance on each analysis remains unchanged. Therefore dependence of female names on male names cannot explain the asymmetries we have observed.

Nor is it the case that the asymmetry in number of syllables and the asymmetry in stress pattern are simply two sides of the same coin. One might argue that if there are more long names, there must be more names with weak initial syllables simply because a certain proportion of polysyllabic names will always be weak-initial. But consider the comparison of the names vocabulary with nouns as a whole. Of the nouns, 81 % are polysyllabic and 15% weak-initial. This produces a ratio of approximately 5.5 to one for strong versus weak initial syllables among the polysyllables. If the two factors were causally related, one would expect a similar ratio in the two sets of names. But in fact we find 76% polysyllables and 5.5% weak-initial male names (fourteen to one), but 90 % polysyllables and 25 % weak-initial female names (only 3.5 to one). The ratios are clearly different, and the asymmetry in stress pattern therefore does not simply fall out of the asymmetry in number of syllables.

The asymmetry in number of syllables is such that it even shows up in hypocoristics (pet-forms). There is a very strong pressure in English to form pet versions of names, and - not surprisingly, given the structure of the vocabulary - these are always monosyllables or SW bisyllables. Names with weak initial syllables always have such diminutives: *Elizabeth* becomes *Betty* or Liz, Sebastian becomes Seb, Michelle becomes Shelley. This process is obviously going to even out many male-female differences, as every native English speaker will have at least one name form which is a monosyllable or SW bisyllable. But where there are alternative diminutive forms for similar male and female names, asymmetries still exist. Some alternatives are perfectly symmetric: both Nicola or Nicholas can be either Nic(k) or Nicky. But while either Robert or Roberta can be Bobby, we feel that only Robert is likely to be Bob. Exactly the same asymmetry holds in the pairs Stevie/Steve, Ronnie/Ron, Micky/Mick, Jackie/Jack and Billie/Bill. The bisyllable can be either male or female, but the monosyllable is more likely to be male. Other pairs of diminutives show a mirror-image pattern, in which the monosyllable can be either male or female, but the bisyllable is more likely to be female: Chris can be either male or female, but Chrissie is more likely to be female. We have found no instances where the monosyllable is more likely to be a female and the bisyllable to be male. Thus even in diminutives there is a tendency for the female forms to be longer. The pattern among diminutives also indicates that the productive diminutive-forming -y suffix is even more closely associated with female than with male names. This suggests the possibility of a further asymmetry between male and female names. Across languages, there are some regularities in the relationships of vowels with certain semantic features; in English, for instance, there is a tendency for [i] sounds to be associated with the concepts 'small', 'sharp', 'bright' (in contrast with [a], which is said to be associated with the concepts 'large', 'dull'; see Taylor, 1976, for a review). If smallness is a concept associated with feminine characteristics rather than with masculine, then it may be that [i] sounds will occur more often in female than in male names not only in diminutive-forming suffixes, but in stressed syllable nuclei as well. Accordingly, we conducted a third analysis in which we examined the nuclear vowels in names and compared this analysis with the distribution of nuclear vowels in English nouns.

### 4. NUCLEAR VOWELS

Von Bismarck (1974) has argued that the perceptual dimension 'sharpness' or 'brightness' represents the largest component of the perception of the timbre of sounds. For vowels, a weighted average of the energy distribution at frequencies from the second formant (F2) upwards appears to be particularly relevant for this dimension: vowels such as [i] have a higher F2 than vowels such as [u]. Carlson, Fant and Granstrom (1975) have produced



Figure 3

Proportions of nouns, male names and female names as a function of nuclear vowels , male; O—O. female; + . . . +, nouns.

an ordering of vowels on the basis of this average; we used this ordering to place the English vowels along a continuum in ten categories from [i] to [u]. The proportions of nouns, male names and female names having primary stressed syllables containing each vowel are plotted in Figure 3. Diphthongs are assigned the value of their initial element; thus [e] and [ei] form a single category, as do [a] and [ai]. [c] and [A] are assigned to a single category on the basis of their similar value on the brightness dimension, as are [3] and [90]. It can be seen that the three sets do not differ markedly. However, it is noticeable that the female set contains a much higher proportion of [i] than the other two sets, and a lower proportion of the vowels towards the other end of the brightness continuum. On  $x^2$  tests there was a significant difference between the female names and the male names ( $X^2[9] = 68.06$ , p < 0.001) and between the female names and the nouns ( $X^2[9] = 144.34$ , p < 0.001). The male names do not differ significantly from the nouns. Analysis of the residuals shows that the source of the difference between the female and male

names is that the female names have significantly more stressed vowels with [i] while the male names have significantly fewer; the source of the difference between female names and nouns is that the female names are significantly more likely to contain stressed vowels with [i] but significantly less likely to contain stressed vowels with [C]/[A] and with [u].

Female names, therefore, are significantly more likely to contain [i] as their stressed vowel; again, this group is by no means composed principally of names with [in] or [ine] as a suffix to a male-name stem, but of names like *Lisa, Beatrice, Celia, Tina, Vera* and the stereotypically feminine *Mimi* and *Fiji*.

#### 5. DISCUSSION

A Rose by any other name might be. for instance, Patricia, Christine, Elizabeth, or Michelle. A Ross by any other name, however, is most likely to be James, Mark, Donald or Michael (these names occupy the same ranks among the top ten on the frequency lists). Women's names are on the average longer than men's; they are more likely to begin with an unstressed syllable; and they are more likely to contain the vowel [i]. Thus there are systematic differences between the (virtually non-overlapping) sets of male and female names in English. A parent naming a child chooses from one phonological distribution for a boy and from another for a girl. How has this asymmetry arisen?

We cannot provide a definitive explanation. However, we can point to several aspects of the data which suggests that an eventual explanation will most probably involve semantic or sociological principles. Firstly, note that male names seem to form the unmarked case. Female names can be formed from male, but the reverse is not possible (removing [in] or [ine] from Doreen, Aileen, Bettina or Sabrina does not make Dor, Ail, Bet and Sabre male names); also, male names are more consistent than female names (the nine popular-name lists, of fifty male and fifty female entries each, contained 196 different female names but only 145 male names. Six male names-John, James, Robert, Richard, William, David - occurred on every list, and another three - Paul, Anthony, Thomas - on eight; but no female name occurred on every list). Secondly, semantic factors are also suggested by the apparent involvement of 'phonetic symbolism'. The statistically significant tendency for female names to contain more instances of [i] than male names do, or than English nouns do, and indeed the use of [in] or [ins] as standard suffixes for forming female names from male names, may well be related to the associations of this vowel with the concepts 'small', 'sharp' and 'bright'. Ohala (1983, 1984) has argued that there is an ethological reason for this sound-symbolic association: small vocal tracts, which produce high-pitched sounds, are typically possessed by smaller, weaker, less threatening beings. (It is hard to imagine a *Tina* more threatening than a *Hugh.*) Perhaps the frequency of [i] in female names has come about, therefore, because smallness and lack of threat are held to be desirable attributes of females. This suggestion could even be extended to a more general principle of phonological weight, which would then embrace the tendency for weak, that is to say phonologically lighter, syllables to occur more often in female names.

Such a principle may seem to conflict with the tendency for female names to be longer than male. Here, however, a possibility exists that semantic factors may exercise an effect via principles of ordering. The order of elements in conjoined expressions such as *dribs and drabs* is often fixed (see

for example, Abraham, 1950; Allan, 1987; Cooper & Ross, 1975; Malkiel, 1959), with the sequence governed by both semantic and phonological principles. Cooper and Ross (1975) state that one of the strongest phonological constraints is that monosyllables precede polysyllables (rough and ready, bread and butter), while one of the strongest semantic constraints is that male concepts precede female (husband and wife, kings and queens). Since, by the nature of our society, male and female names often occur in pairs, this suggests a possible conspiracy of the syllabic length of names with the preference for male-female ordering. Moreover, other differences between male and female names conform with this interpretation. Second elements in freezes tend to contain fewer final consonants than first elements (hem and haw), and it is true that female names are more likely to end in a vowel than male names are. Although nuclear vowels with high F2 tend to precede vowels with low F2 (dribs and drabs), there is also a tendency for short vowels to precede long (trick or treat); in accord with this is the fact that [i], which occurs more often than expected in female names, is a long vowel.

However, ordering principles do not provide a full account of the observed asymmetries. As Cooper and Ross (1975) point out, there are exceptions to the male-female ordering principle, and one of them is that 'mothers are special': Ma and Pa, mother and father. (Isn't the very first name we all learn Mama, with its strong initial syllable and low back vowel?) More importantly, ordering principles fail to account for the asymmetry in occurrence of weak initial syllables, since weak initial syllables often sound worse in second position than in first. Thus a complete explanation of the differences between male and female names has still to be established.

Finally, note that cross-linguistic comparisons are necessary to establish whether the asymmetries which we have observed are true only of the name vocabulary of the English language. In many languages asymmetries between male and female names may not exist at all, or may be confined to sexmarked suffixation. However, it is a curious fact<sup>6</sup> that in French, which is prosodically quite different from English, the small class of monosyllabic names is very largely male (Jean, Paul, Yves, Marc, Luc, Jacques), and that male names tend to begin with a (phonologically heavier) closed syllable (Alphonse, Pascal, Marcel) more often than female names do (with the proviso that the formation of female names from male is more productive in French than in English).

Walter Shandy has strong views that given names can influence one's character: 'how many...might have done exceeding well in the world, had not their characters and spirits been totally depress'd and Nicodemus'd into nothing' (Sterne, 1759). To what extent may our behaviour be shaped by

[6] Brought to our attention by Bob Ladd, to whom we express our thanks.

possession of a name which by its very sound pattern proclaims us as typically male or typically female?

Authors' address: MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.

#### REFERENCES

- Abraham, R. M. (1950). Fixed order of coordinates: a study in lexicography. *Modern Language* Journal 34. 276-287.
- Allan, K. (1987). Hierarchies and the choice of left conjuncts (with particular attention to English). JL 23. 51-77.
- von Bismarck, G. (1974). Timbre of steady sounds: a factorial investigation of its verbal attributes. Acustica 30. 146-159.
- Butterfield, S. & Cutler, A. (1988). Segmentation errors by human listeners; evidence for a prosodic segmentation strategy. Proceedings of SPEECH 8S (Seventh Symposium of the Federation of Acoustic Societies of Europe), Edinburgh. 827-833.
- Carlson, R., Elenius, K., Granstrom, B. & Hunnicutt. S. (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. Speech Transmission Laboratory, Quarterly Progress and Status Report, 1985/I. 63-94.
- Carlson, R., Fant, G. & Granstrom, B. (1975). Two-formant models, pitch and vowel perception. In Fant. G. & Tatham, M. A. A. (eds). Auditory analysis and perception of speech. London: Academic Press. 55-82.
- Cohen, G. (1989). Remembering proper names. Paper presented to the Sixth Annual Conference, British Psychological Society Cognitive Psychology Section. Cambridge.

Comrie, B. (1979). Morphophonemic exceptions and phonetic distance. *Linguistics* 17. 51-60.

- Cooper, W. E. & Ross, J. R. (1975). World order. Papers from the Parasession on Functionalism, Chicago Linguistic Society. 63-111.
- Cox, D. R. & Lauh, E. (1967). A note on the graphical analysis of multidimensional contingency tables. *Technometrics* 9. 481-488.
- Cutler. A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. Computer Speech and Language 2. 133-142.
- Cutler. A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance 14. 113-121.
- Dunkling, L. & Gosling, W. (eds) (1987). Everyman's dictionary of first names. 2nd edn. London: Dent.
- Grosjean, F., Grosjean, L. & Lane, H. (1979). The patterns of silence: performance structures in sentence production. Cognitive Psychology 11. 58-81.
- Hanks, P. & Hodges, F. (eds) (1986). The Oxford minidictionary of first names. Oxford: Oxford University Press.
- Harrington, J., Watson, G. & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. Computer Speech & Language 3. 367-382.

Malkiel, Y. (1959). Studies in irreversible binomials. Lingua 8. 113-160.

Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica* 40. 1-18.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of Fo of voice. Phone tica 41. 1-16.

Procter, P. (ed.) (1975). Longman's dictionary of contemporary English. London: Longman. Semenza, C. & Zettin, M. (1988). Generating proper names: a case of selective inability. Cognitive Neuropsychology 5. 711 721.

Sterne, L. (1759). The life and opinions of Tristram Shandy, Gentleman. York: Ward. Taylor, I. (1976). Introduction to Psycholinguistics. New York: Holt, Rinehart & Winston. Van Lancker, D. & Klein, K. (1990). Preserved comprehension of personal names in global aphasia. Paper presented to Eighteenth Annual Meeting, International Neuropsychological Society, Orlando, Florida.