

SPELLING ERRORS IN THE WRITINGS OF UPPER INTERMEDIATE FOREIGN LEARNERS OF ENGLISH: THEORY AND A QUANTITATIVE ANALYSIS

ORTA-ÜSTÜ DÜZEYDEKİ İNGİLİZCE ÖĞRENCİLERİNİN YAZILARINDAKİ YAZIM HATALARI: KURAM VE NİCELİKSEL BİR ÇÖZÜMLEME

Yrd. Doç. Dr. Margaret J-M SÖNMEZ
Middle East Technical University
Department of Foreign Language Education

ABSTRACT

This paper presents a quantitative analysis of an un-elicited corpus of English spelling errors from Turkish learners. It focuses on the issues raised by an earlier corpus-based analysis of errors made by native speakers and on the theoretical significance of the findings of both these studies.

A review of the theoretical background to spelling error analysis is followed by a quantitative analysis of spelling mistakes made by upper intermediate level Turkish learners of English. The results indicate that native and foreign writers share a number of cognitive and perceptual processes in the production of English spellings, notably those concerning a degree of independence of orthography from phonology, sensitivity to word length, and the relative perceptual salience of segmental positions within the word.

Key words: *Spelling, error analysis, cognitive processes, foreign language learning.*

ÖZ

Bu çalışma Türk öğrencilerin yaptığı İngilizce yazım hatalarından oluşan bir bütüncenin niceliksel bir analizini yapmaktadır. Çalışma, anadili İngilizce olanların yaptığı yazım hatalarına dayalı bir analizin ortaya koyduğu sorunlar ve bu çalışma ile daha önceki bir çalışmanın bulgularının kuramsal önemi üzerine odaklanmaktadır.

Çalışmada, yazım hatalarına ilişkin kuramsal bir arkaplan incelemesini, İngilizce orta-üst düzeydeki öğrencilerin yaptığı yazım hatalarının niceliksel bir analizi izlemektedir. Elde edilen sonuçlar, yazımın sesbilimden bağımsızlık derecesi, sözcük uzunluğuna duyarlılık ve sözcük içindeki parçaların göreceli algılanabilirliği ile ilgili konular başta olmak üzere, bilişsel ve algılama süreçlerinde anadili İngilizce olanlar ile anadili Türkçe olanların İngilizce yazım üretiminde ortak yanları olduğunu göstermektedir.

Anahtar sözcükler: *ortografi, yanlış çözümlemesi, bilişsel süreçler, yabancı dil öğrenimi*

INTRODUCTION

Teachers of English as a Foreign Language have not published much about spelling. This may be because, as is the case for most native writers, spelling errors do not seem to cause their students practical problems, nor do they impede the progress of learning other aspects of the language. It is also largely ignored by researchers into the theory of language learning, probably because there is little current interest in spelling as a linguistic ability. While there has been some work done on the use of English spelling in teaching pronunciation (Dickerson 1985, 1987, 1990a, 1990b, 1992), there is none on English spelling per se. Neither, with the exception of pathological studies, is there much published research on the real spellings of present-day advanced native writers of English. By far the greatest amount of work on English

spelling is pathological in outlook, concentrating mostly on dyslexia, dysgraphia and aphasia. Interesting work has been done on typed errors, using spell-check programme design as a basis (Rogers and Willet, 1991). The only published research on handwritten spelling errors of normal adults is Wing and Baddely (1980) (W&B). W&B (1980: 255) say that an important reason for this paucity of research "is that error rates in normal people are very low". They calculated a rate of 1.5% misspellings in their corpus of spelling errors in the Cambridge University Entrance exam, and reported a similar rate of 1.1% from Chedru and Geschwind's 1972 study of mentally disturbed subjects (C&G). Methods of calculation may differ slightly (W&B included repeat errors in their rate of 1.5% but elsewhere in the same piece of research only allowed only one error per word), but not enough to alter the fact that these are, indeed, very low figures.

THEORETICAL BACKGROUND

There are many and fundamental problems with the simple assumption of grapheme-phoneme correspondence upon which so much that is written about spelling is based. While it is still a useful, and perhaps the most used, part of a model upon which to base investigations into teaching and learning strategies at the elementary level, even here its shortcomings were long ago acknowledged by the proponents of the "look and say" method, and by those who developed the Initial Training Alphabet. Current research into English spelling indicates that learners relying largely upon phonological information spell worse, or learn spelling slower, than those of their peers who use other strategies such as developing a visually-based spelling lexicon (Peters 1992), although other research has indicated that the issue is more complex. Tenney 1980, for instance, shows how, in some cases, the visual level works together with other, linguistic levels in the production of correct spellings, and Barron shows that both visual and phonological strategies are used by the most successful learners of reading, suggesting that this must also be the case for learners' spelling, even in languages with high one-to-one phoneme-to-grapheme rates of correspondence (1980: 84). Other factors which have been shown to be important in both the learning and the production of spelling include lexical frequency (an irregularly spelt but frequent word may be correctly written where a regularly spelt but infrequent word is not) and analogy (research on these two factors is reported in Barry 1992).

It has been shown that even less than learners do advanced native spellers use a letter-to-sound correspondence for each letter of each word that they spell (see Ellis 1974:73, whose model may be elaborated with reference to Rumelhart and McClelland 1986:20-25, and Sterling 1992:287-289). In their contribution to Frith 1980, "Marsh, Friedman, Welch and Desberg conclude that 'visual' strategies take over from sound-to-letter decoding strategies from about age ten onwards. From then on new words are more and more spelled by analogy to old words" (Frith 1980:3). And the established core (high frequency) vocabulary is accessible to adult writers in the form of full words, that is, it is to a certain extent lexicalised (Lurie et al. 1970 - quoted in W&B 1980:252; Seymour 1992:54; Barry 1992:72). Some researchers even suggest that a high degree of spelling lexicalisation distinguishes the good speller from the merely average (Sloboda 1980:247).

Of course errors do creep in, and these have been di-

vided into two main categories (W&B:254-255). The first of these comprises "slips", which are found in common words and would be corrected by the writer if they were pointed out to him or her; they arise due to processing interference caused by, for instance, analogical thinking, mechanical omissions, additions, attention lapses, or thinking ahead. Errors which are postulated to be "invisible" to the writer (here the relevant words will be consistently misspelled, and the writers would not be able to correct them if they were pointed out to them) are known as convention errors, and, if we accept the existence of a spelling lexicon of core words, we would expect to find them in the less familiar words.

Most of the time, the experienced writer produces words, and maybe even whole phrases, without much recourse to segmental phonological analysis. A model that comprises a lexicon of core spellings seems to be consistent with the fact that convention errors in advanced spellers are not evenly spread across their vocabulary according to the types of error, but are centred on particular words. That is, an advanced writer who consistently spells *commissar* with only one *m* and *concomitant* with two is not found to make the same mistakes in other, perhaps more common, words such as - well - *common*, or *comic*. However, as W&B point out, it is in practice virtually impossible to draw a strict line between errors that are slips and those that are convention errors. Nor is it possible, without recourse to strictly controlled experimental tasks, to overcome this difficulty by accepting as errors of convention only those misspellings which are made more than once by any writer: since convention errors are expected to be found in infrequently written words, these words will by definition not usually be repeated and so the repeated error device will leave out of account most errors potentially classifiable as errors of convention. It seems, then, that although this distinction is undoubtedly useful in the theory of spelling, it becomes largely unusable in the case of a corpus-based study of real spellings.

W&B, working on a model of sentence writing which contains a short-term memory buffer, to account for the observed fact that the sentence is linguistically formulated faster than it can be written down, postulate that word- and sentence- position will affect the possibility of spelling error in any particular segment. They expected to find that longer words were more likely to be misspelled than shorter ones, and that more errors would be found in longer sentences than in shorter ones (W&B: 259). They found that there was indeed a correlation between word

length and rate of error, but the hypothesis that errors would be more frequently found at the ends of words and sentences had to be abandoned because this was not the case in their data. An adaptation of the same model to include insights from Baddeley's work on short-term memory, however, provided expectations which were supported by the research, and this was that the middle of words and sentences, being less perceptually salient than either the beginnings or the ends, would show higher proportions of errors (p.261). A particularly interesting feature of this hypothesis is that it is a purely graphic approach to spelling, since there is no way that we can so confidently claim lowered perceptual salience in the middle of spoken words.

Finally, the same paper draws our attention to a level of spelling which is all too often ignored, that of the motor level: spelling is not only a mental process, it must also involve the transmission of mentally-processed spellings to those muscular movements which result in writing. This part of the process is prone to error as much as are the other parts, and typically results in motoric errors ("based on similar movements", p. 272), stroke omissions (e.g. *n* for *m*, *l* for *f*) and stroke additions (e.g. *m* for *n*, *t* for *l*).

The extent to which lexicalisation of spellings is the case with non-native writers is yet to be investigated, but experience with advanced mother-tongue spellers has already warned us not to rely on the same phoneme-grapheme models as those upon which traditional teaching and learning models of alphabetic writing systems have been based. There are, in addition, large numbers of inaccessible factors that should be taken into consideration if one wishes to say that a foreign learner's phonological information about an English word is of a particular form. Quite apart from the complex issue of interlanguage phonology, which is relevant to both sound perception and its mental representation but is usually only investigated from the mental representation-production angle, we have to address the complex issue of the accent(s) to which the learners have been exposed. Each student has an individual experience of English language teachers. Most, perhaps all, have been taught by more than two teachers; many of them have only been taught by non-native speakers, whose pronunciation may be far from that of native English speakers; those who have been taught by native English speakers will have had teachers speaking particular dialects which are different from the dialects spoken by the teachers of other students, and different again from the pronunciation

described in their text books. The American/British accent divide is only one of the many pronunciation difficulties foreign learners of English have to overcome.

For learners whose native languages use the same writing system (eg the Latin alphabet) as that of their target language, their foreign language spelling processes may involve analogies not only between an unknown word and other known words of the target language, but also between spelling patterns from their own language and that of any other foreign language they may be familiar with, and this level of complexity may be involved for all of the factors affecting the spelling process, for instance in perceived phoneme-grapheme and morpho-graphemic correspondences. These factors cannot be empirically tested through purely quantitative methods, and further research in this area should involve qualitative error analysis. In addition, a case-study approach with feedback from the subject(s) would be a useful addition to our understanding of foreign learners' spelling processes.

Finally, one should note that advanced vocabulary is more frequently encountered in written than in spoken form, and research for native speaking learners has shown that the relation between reading and spelling is complex, the process of the one being far from a simple reversal of the process of the other (Baron, Trieman, Wilf and Kellman 1980, Henderson and Chard, 1980: 112 & 116, Morton 1980:124): a point supported by the fact that there are children who are good readers but poor spellers, or poor spellers but good readers (Bryant and Bradley 1980).

This paper addresses the following specific questions:

1. Do the foreign learners show a significantly higher rate of spelling error than that of the native writers investigated in earlier papers?
2. Do the foreign learners' spelling errors show any tendencies to favour one type of error and, if so, is this different from native writers' error tendencies?
3. Are foreign learners' spelling errors found more frequently in longer words than in shorter ones?
4. Is W&B's "depression of error rate at the extremes" of words (1980:261) borne out in the writings of foreign learners?

Other questions arising from the above review of theory require qualitative analysis of spelling errors, and/or differently designed research, such as controlled tasks or case studies.

Research material

The sample comprises writings of 57 students in the second semester of their first undergraduate year at a Turkish university whose medium of instruction is English. All of the students were between 20 and 23 years of age, and native Turkish speakers. The spelling errors were extracted from their mid-term examination papers. In this exam the students had two hours in which to answer four essay-type questions. They were allowed dictionaries, but no other books or papers. The students had seen the correct written forms of all their misspelled words on the board in lesson time and in course notes given to them at the beginning of the semester. In some cases the correct spellings were in front of them (on the question paper) as they wrote. The extent to which they were familiar with the correctly spelled version of this vocabulary may be guessed at by the fact that all students included sentences and, sometimes, whole passages memorised from the notes.

METHOD OF ANALYSIS

1. The papers were coded so that it would be possible to identify repeat errors
2. The total number of spelling errors were extracted and counted.
3. In order to have results comparable to those of W&B the errors were then recounted according to their method in which only the first (left-most) error in each word was allowed and repeat errors were ignored.
4. The number of lines of script was counted and rate of spelling error calculated on the basis of 10 words per line and only one error per word but including repeat errors (W&B calculated their error rate in this way, presumably in order to be comparable with C&G's; elsewhere they ignore repeat errors).
5. The spelling errors were categorised according to W&B and C&G's categories, and the distribution of errors among the categories was calculated.
6. The length of the correct form of each misspelt word was noted and the distribution of errors according to word length was calculated.
7. The positions of errors within the words were identified using W&B's criteria for identifying positions. Here, as in W&B, errors were allocated to one of five positions within each word, according to the symmetrical distribution used in the earlier research. This means that there were always equal numbers of letters in the "start" and "end" positions, and equal numbers of letters in the "pre-mid" and "post-mid" positions. In this way words

with any number of letters over 5 could be segmented symmetrically. For the present research a separate set of results including words with less than 5 letters was included, 2 letter words being deemed to have only start and end positions, three letter words having start, end, and mid positions only, and four letter words having all positions except mid.

Where a student spelling was not clear due to handwriting or the identity word was doubtful due to context, it was not included in the data. British and US spellings, as well as alternative spellings allowed in *Collins English Dictionary* (e.g. *judgement*, *judgment*) were all considered correct. Grammatical errors (such as *he go*), and the plural of "pendulum" where spelt <pendulums> were not included.

RESULTS

1. Rate of spelling error

313 lexical items were misspelled 701 times. The estimated total number of words in the scripts from which the errors were collected was 63,770 and the rate of spelling error for these students is therefore 1.1%.

2. Classification of Errors

Using the classification of spelling errors used by both C&G and W&B (see pp. 254 and 263), they were found to fall into the following groups:

1. substitution: 235, or 40.8% of the 576 non-repeated errors.
 2. omission: 169, or 29.3%
 3. addition: 130, or 22.3%
 4. inversion: 41, or 7.1%
- one spelling, which did not fit into these categories, was given its own category:
5. other: 1, or 0.2%

3. Error: Word-length correlation

Using the same normative criteria as W&B, that is, taking average English word-length to be 4.7 characters, and thus defining "long words" as those whose correct spellings contain more than 5 letters, it was found that 519, or 90.1% of the errors, occurred in long words. It was decided to investigate correlation between word length and rate of spelling error in greater detail, and the number of errors for each of the 13 different lengths of words in which errors had occurred was calculated. The results, expressed as percentages of the total number of misspelled words, are shown below:

Table 1

Distribution of errors according to word-length

length	2	3	4	5	6	7	8
%	0.2	0.9	4	4.7	8.9	11.3	15.6
length	9	10	11	12	13	14	15
%	18.6	17.5	7.3	6.4	2.9	0.9	0.9

4. Position of error in word

The same sample analysed according to error position within word provided results which accorded fairly well with those reported in W&B. Total numbers of errors in each of the five positions were:

Table 2

Distribution of errors according to word position

Position	Start	Pre- mid	Mid	Post- mid	End
%	9.7	19.4	26.6	24	20.3
misspelled words (all words)					
%	10	20	28	23	19
misspelled words (5 letters or more only)					

Given the fact that only the left-most error was counted for each word, and the data in this research contained many words with multiple errors, it was decided not to analyse the word-position of errors any further because of possible misrepresentations of the data.

DISCUSSION

1. The spelling error rate of 1.1% is the same that that found by C&G and close to W&B's 1.5%.

2. Concerning the categories of spelling errors, C&G reported that, from most to least, their patients' errors were ordered as omissions, additions, substitutions, inversions, with this last category accounting for only a small proportion of errors found. Our data showed a very different ordering of error types. Here the results gave clear precedence to substitution, followed by omission, addition, and inversion. We cannot compare this categorisation and distribution of errors with those of W&B because their paper does not include that information, although a comment (p. 263) implies that, contrary to C&G's study but like this present one, W&B's corpus showed less addition (insertion) errors than omissions.

3. W&B's results of a positive correlation between number of spelling errors and length of word are confirmed for foreign learners. It was found that the three highest numbers of misspellings were found in words whose lengths were very much above the average word length with the peak, or highest proportion of spelling errors found in words of 9 letters.

In spite of some distortion of data inevitably entering the results due to the left-most-error error-counting method, the word-position analysis was useful in showing that the spelling errors of foreign learners showed the same patterns of distribution as those of native speakers, under the same analytic conditions. In fact, given the fact that the method dis-favours errors in the last position, the fact that there was a higher proportion of errors in the last position than in the first indicates that the so-called "depression of error at extremes" is more clearly a depression of error at start position.

CONCLUSION

The four specific questions set for this study have been answered with satisfactorily clear results:

1. Given the fact that pronunciation is the single area of language learning which is least likely to attain native performance in any adult learner, the close similarity of native and foreign learner error rates in spelling confirms the already asserted disjunction between phonology and orthography in practised writers.

2. Like native writers, foreign learners' spelling errors show a clear tendency to favour one type of error over the others. In this work, substitution errors were significantly more numerous than the other 4 types identified in the initial categorisation of spelling errors. Pre-dominance of substitution errors was not the case with C&G's study of disturbed native writers; its relative importance in the W&B corpus has not been reported. So no further conclusions can be drawn from these results at present.

3. That longer words are more likely to be misspelled than shorter words, by foreign learners as well as by native writers indicates that spelling difficulty is, in some respects at least, inherent in perceptual levels of cognition, rather than in more linguistically specific, language processing activities.

4. Unlike W&B's findings, the data analysed here showed a relatively insignificant depression of error at word ends. A possible flaw in W&B's methodology was noted in connection with this. It was postulated that the results of a study using a different counting technique

could indicate a greater proportion depression of error at word start position.

One aim of this paper was to see if research into foreign learners' spelling errors could usefully add to the theory of spelling processes, and the findings here indicate that they can. This analysis has shown that, like native writers, fairly advanced foreign learners of English use spelling strategies other than phoneme-grapheme correspondence. It showed that for them, as for native writers, longer words cause more spelling problems than shorter words. It was further seen that there is a close similarity between the segmental distribution of spelling errors in the data from native and foreign writers, to the extent that word boundaries have greater salience than central portions, with the beginnings of words having the greatest salience.

Broadening the research base of spelling theory tests the extent to which our knowledge is limited by having been previously confined to native writers. A quantitative analysis of spelling errors tends to test the cognitive aspects of spelling rather than its purely linguistic components, and may therefore lend itself to the formulation of hypotheses concerning universals of language processing. The results of this preliminary study have been encouraging in this respect, in that they show certain spelling processes to be shared by writers with very different linguistic backgrounds. It is nevertheless too early to claim universality for these processes, since we have here looked only at spelling errors in English, and our writers were all native users of the Latin alphabet.

REFERENCES

- Baron, Jonathan, Rebecca Treiman, Jennifer F. Wilf, and Philip Kellman. (1980) "Spelling and Reading by Rules." in Frith 1980 : 159-194.
- Barron, Roderick W. (1980) "Visual and Phonological Strategies in Reading and Spelling." in Frith 1980: 195-213.
- Barry, Christopher. (1992) "Interaction between Lexical and Assembled Spelling." in Sterling and Robson 1992 : 86.
- Bryant, Peter E. and Lynette Bradley. (1980) "Why Children sometimes Write Words which they do not Read." in Frith 1980: 355-370.
- Dickerson, Wayne B. (1985) "The Visible Y: A case of spelling in pronunciation learning." *TESOL Quarterly*, vol. 19, no: 2: 303-316.
- (1987) "Orthography as a pronunciation resource." *World Englishes*, vol. 6, no: 1: 11-20.
- (1990a) "English <s>: Cracking a symbol-sound code." *Issues and Developments in English and Applied Linguistics* 5 : 39-64.
- (1990b) "Morphology via Orthography: A Visual Approach to Oral Decisions." *Applied Linguistics*, vol. 11, No. 3: 238-252.
- (1992) "Orthography: A Window on the World of Sound." In Brown, Adam (ed) *Approaches to Pronunciation Teaching*. London: Macmillan (*Review of English Language Teaching*, vol.2, no:2).
- Ellis, Andrew W. (1984) *Reading, Writing and Dyslexia: A Cognitive Analysis*. London: Lawrence Erlbaum Associates.
- Frith, Uta. (1980) (ed.) *Cognitive Processes in Spelling*. London: Academic Press.
- (1980a) "Introduction." in Frith 1980: 1-5.
- Henderson, Leslie and Jackie Chard. (1980) "The Reader's Implicit Knowledge of Orthographic Structure." in Frith 1980: 85-116.
- Krashen, Stephen. (1981) *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Marsh, George, Morton Friedman, Veronica Welch, and Peter Desberg. (1980) "The Development of Strategies in Spelling." In Frith 1980; 339-354.
- Morton, John. (1980) "The Logogen Model and Orthographic Structure." in Frith 1980: 117-134.
- Peters, Margaret L. (1992) "Towards Spelling Autonomy." in Sterling and Robinson 1992 : 220-223.
- Richards, Jack C. (1974). (Ed.) *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman.
- Rogers, H. J. and P. Willet. (1991) "Searching for Historical Word Forms in Text Databases using Spelling-Correction Methods: Reverse Error and Phonetic Coding Methods". *The Journal of Documentation*, vol. 47, no: 4