Brain and Language **68,** 277–283 (1999)
Article ID brln.1999.2108, available online at http://www.idealibrary.com on IDEAL®

Why Simple Verb Forms Can Be So Difficult to Spell: The Influence of Homophone Frequency and Distance in Dutch

Dominiek Sandra, Steven Frisson, and Frans Daems

University of Antwerp, Belgium

Two experiments are reported in which the determinants of spelling errors on homophonous verb forms in Dutch were studied. Both experiments indicated that errors were determined by the frequency relationship between the two homophonous forms and the distance between the verb and the word determining its spelling. We propose an interference model for spelling in which a phonologically driven retrieval process is the locus of the frequency effect and a morphosyntactically driven computational component can account for the distance effect. Alternative explanations are also explored. © 1999 Academic Press

Key Words: spelling; homophones; interference.

1. INTRODUCTION

In contrast with the large body of literature on morphology in the reading process (e.g., Feldman, 1995; Sandra & Taft, 1994), there has been very little attention to the role of morphology in the spelling process. Yet the study of errors on the spelling of morphologically complex words may shed light on the underlying production process. We will investigate the spelling of particular types of verb forms in Dutch.

Our starting point is what seems a paradox at first sight: even though the spelling of regularly inflected verb forms in Dutch is morphographic, i.e., it contains distinct orthographic sequences for the stem and suffix morphemes, it represents one of the most notorious problems in the written language,

The order of the first two authors is arbitrary. The second author is also affiliated with the Human Communication Research Centre, Department of Psychology, University of Glasgow. We thank Kate Nation and two anonymous reviewers for their helpful comments. We are most grateful to the staff and pupils of the different schools for their help and cooperation, and to the students who helped us in collecting data: Pamela Bruers, Ines Fels, Leen Maes, Priscilla Rombaut, Caroline Schoonejans, Ann T'Seyen, Sofie Tack, Helena Urgel, Roel Vankerckhove, and Jasmine Vervenne.

Address correspondence and reprint requests to Dominiek Sandra, University of Antwerp—UFSIA, Room D 109, Prinsstraat 13, B-2000 Antwerpen, Belgium. E-mail: sandra@uia.ua.ac.be.

0093-934X/99 $30.00

causing errors even among highly trained writers. For instance, errors are made on the third person singular present tense, even though its formation only requires the addition of a *t* suffix to the verb stem. This kind of error occurs despite intensive training on the verb spelling system from the fourth year of primary school onwards and the social pressure for correct spelling (of verbs in particular).

Not all Dutch verb forms present spelling problems. Even though forms like *speelt* (plays) and *treedt* (treads) instantiate the same morphographic spelling principle for the third-person singular of the present tense, i.e., stem plus suffix, no errors are made on the former whereas spelling errors to the latter form occur frequently. A possible reason for these errors is the presence of a homophone: *treed* (first-person singular, present tense) has the same pronunciation as *treedt* ([*tre:t*]), due to devoicing of stem-final *d*. Homophony may affect spelling in two ways: either the frequency relationship between the homophonic forms determines the likelihood of spelling errors on each form or the presence of a homophone creates spelling uncertainty for both forms equally. The homophone frequency relationship was the first factor manipulated in the experiments. A second (possibly additional) reason for the verb spelling errors might be the distance between the verb form and the word determining its spelling. More errors may be made when the distance between these two words is larger. Distance was the second factor manipulated in the experiments.

2. EXPERIMENT 1

2.1. Method

Materials. Verb forms were selected that were homophonous in the first and third person singular present tense (e.g.,[tre:t]/*treed–treedt*). The three possible frequency relationships (see Table 1) between the homophones were distinguished (token frequencies based on the

TABLE 1
Item Types and Frequencies for Experiments 1 and 2

			Frequency	
Homophonous type	*N*	Type	D	(D)T
Experiment 1				
TREED–TREEDT	8	D > DT	182	40
(*tread–treads*)	8	D < DT	38	334
	8	D = DT	42	34
Experiment 2				
VERSIERD–VERSIERT	10	D > T	228	17
(*decorated–decorates*)	3	D < T	246	1,754
	10	D = T	18	16

Note. ‘‘Type’’ stands for the frequency relationship between the homophones. For example, D > DT signifies that the D spelling of a verb is more frequent than its DT spelling.

42 million CELEX count, Baayen, Piepenbrock, & van Rijn, 1993). The distance factor was manipulated by presenting the same verb forms in main clauses and subclauses. As a result of Dutch SVO order in main clauses and SOV order in subclauses, the inflected verb, whose spelling is determined by the subject, is always immediately adjacent to the subject in main clauses, whereas it is often separated from it by other sentence material in subclauses (always four words in the experiment).

The same carrier sentences were used for the two homophonic forms, differing only in their subjects (first versus third person). For each form a main clause and a subclause were constructed. Two lists were set up for both main clauses and subclauses. Within each list, half of the verbs within each frequency condition were used in the first person, the other half in the third person. This assignment of items to grammatical person was counterbalanced across the two lists. Thus repetition of the same verb was avoided. Examples can be found in (1) and (2). Table 1 presents the item specifics.

(1) Main clause (adjacent): Ik (Hij) TREED(T) in mijn (zijn) vaders voetsporen;. . .
I (He) follow(s) in my (his) father's footsteps; . . .

(2) Subclause (distant): . . . omdat ik (zij) in het strenge klooster TREED(T).
. . . because I (she) enter(s) the strict monastic order.

Procedure. The experimenter read the sentences one by one at a relatively slow pace. Participants followed the sentences on their response sheet and filled in the blank spaces (two words per sentence; the target verb form and one other word type which served as distractor). The purpose was to give participants sufficient time to write down the spelling of the word without giving them much verification time. Sentences were not reread. After the final sentence was read, the response sheets were immediately collected. In order to have participants pay enough attention, they were told that the spelling test they were to do would be marked.

Participants. A total number of 198 pupils, age 18 (last year secondary school), were divided over the different lists.

2.2. Results and Discussion

The mean number of observations across conditions was 193. The number of intrusion errors (e.g., *treed* spelled as *treedt* or vice versa) were counted and chi-square analyses were calculated on these data. Because the total number of observations varied slightly across conditions (due to missing responses), the findings are expressed as error percentages in Table 2.

The analysis of the 2 (Homophone) $\times$ 3 (Frequency Relationship) matrix for the adjacent conditions (main clause) was significant ($\chi^2 = 7.50$, $p < .05$). The comparison between the numbers of D and DT intrusions in each frequency condition was significant for the D > DT type ($\chi^2 = 6.82$, $p < .01$) but not for the other two types (ps $> .10$), although the difference was in the predicted direction for the D < DT type. The analysis of the 2 $\times$ 3 matrix of the data for the distant conditions (subclause) was significant ($\chi^2 = 23.90$, $p < .001$). Comparisons between the two intrusion types were significant for the D > DT type ($\chi^2 = 6.64$, $p < .01$) and for the D < DT type ($\chi^2 = 17.51$, $p < .001$), but not for the D = DT type ($\chi^2 < 1$). The distance factor was assessed by comparing the error frequencies in main and subclauses for each frequency type separately. Significantly more errors were made in subclauses ($\chi^2 = 47.85$, $p < .001$ for D > DT; $\chi^2 = 25.51$, $p <$

TABLE 2
Percentages of Homophonic Intrusion Errors for Experiments 1 and 2

	Frequency type	Adjacent		Distant	
Experiment 1		Correct spelling		Correct spelling	
		D	DT	D	DT
	D > DT	4.46	11.88	23.20	37.63
	D < DT	7.43	4.46	28.35	9.79
	D = DT	8.42	8.42	19.30	19.08
Experiment 2		Correct spelling		Correct spelling	
		D	T	D	T
	D > T	5.04	11.02	7.57	45.75
	D < T	10.00	2.00	29.41	19.15
	D = T	7.44	15.66	12.75	32.92

.001 for D < DT; and $\chi^2 = 8.17$, $p < .01$ for D = DT). For each frequency type the relationship between D and DT intrusions was the same in main clauses and subclauses (all *p*s > .10).

These results suggest that both homophone frequency relationship and distance influence spelling accuracy. Target forms were more subject to intrusions when there is a homophone of higher frequency. More errors were made to target forms when the subject was not adjacent. Moreover, these two effects did not interact, suggesting a comparable distribution of errors in both the adjacent and the distant conditions.

The verb forms in the present experiment may have been particularly difficult because their suffix cannot be perceived in the phonetic sequence, being a zero morpheme in the first person and identical to the stem-final consonant (devoiced/*d*/) in the case of the third person. In Experiment 2 we used verb forms with a clearly distinct suffix in their spoken form.

3. EXPERIMENT 2

3.1. Method

Materials. In Experiment 2, prefixed verb forms were homophonous between the third person singular present tense and the past participle (e.g., *[vərsi:rt]/versiert -versierd*). In contrast to the items in Experiment 1, all verbs lacked a stem-final *d* in their spelling pattern. The same frequency relationships were used as before (see Table 1). The distance factor was again manipulated by using main clauses and subclauses. Whereas the spelling of the present tense is determined by the subject (see Experiment 1), the presence of a participle is signalled by the use of an auxiliary verb.[1] In main clauses this auxiliary was always separated from the (sentence-final) participle by intervening words. In subclauses the auxiliary and participle were

[1] The Dutch past participle has a typical prefix (ge-), which could be used by spellers as a cue for spelling the participle form. However, this prefix was absent in all participles in the present experiment since they contained a weak prefix (be- or ver-) preventing the use of the ge-prefix.

adjacent in the clause-final verb group. In other words, for participles the distance between the verb form and the word determining its spelling was largest in the main clause, for present tenses the opposite held. Examples can be found in (3)–(4).

(3a) Adjacent (present tense): De pygmee VERSIERT zijn hut met apenstaarten.
The pygmy decorates his hut with monkey tails.

(3b) Adjacent (past part.): . . . dat het huis voor een feestje werd VERSIERD.
. . . that the house was decorated for a party.

(4a) Distant (present tense): . . . dat ze haar huis met kerstballen VERSIERT?
. . . that she decorates her house with Christmas balls?

(4b) Distant (past part.): . . . wordt het kantoor door de collega's volledig VERSIERD.
. . . is the office completely decorated by the colleagues.

Since Experiment 2 was run together with Experiment 1, the same procedure was followed and the same participants were tested. Due to the scarcity of items in the D $<$ T type, we were able to find only three items for this type in the entire CELEX corpus.

3.2. Results and Discussion

The percentages of intrusion errors for all conditions are presented in Table 2. The mean number of observations across conditions was 214. The analysis of the 2 (Homophone) $\times$ 3 (Frequency Relationship) matrix on the data for the adjacent conditions (main clause for present tenses, subclause for participles) was significant ($\chi^2 = 17.53$, $p < .001$). The comparison between the numbers of D and T intrusions was significant for the D $>$ T and D $<$ T types ($\chi^2 = 4.91$, $p < .05$ and $\chi^2 = 8.00$, $p < .01$, respectively) but also for the D $=$ T type ($\chi^2 = 7.02$, $p < .01$). The analysis of the same 2 $\times$ 3 matrix for the distant conditions was significant ($\chi^2 = 51.24$, $p < .001$). A comparison between the two intrusion types was significant for the D $>$ T type ($\chi^2 = 69.83$, $p < .001$) and reached borderline significance for the D $<$ T type ($\chi^2 = 3.56$, $p < .10$) but was also significant for the D $=$ T type ($\chi^2 = 22.33$, $p < .001$). The distance factor was assessed by comparing error frequencies in the two distance conditions for each frequency relationship separately. All comparisons were significant ($\chi^2 = 47.10$, $p < .001$ for D $>$ T; $\chi^2 = 32.40$, $p < .001$ for D $<$ T; $\chi^2 = 17.26$, $p < .001$ for D $=$ T). The interaction between frequency relationship and distance was assessed for each type separately. The interaction was nonsignificant for the D $<$ T and D $=$ T types ($\chi^2 = 1.95$, $p > .10$ and $\chi^2 < 1$, respectively) but reached significance for the D $>$ T type ($\chi^2 = 4.06$, $p < .05$).

Homophone frequency and distance again determined spelling errors. However, two findings deviated from this general pattern: the different error frequencies for the two homophones in the D $=$ T type (in both distance conditions) and the significant interaction between homophone frequency and distance in the D $>$ T condition. Both results indicate that the participants strongly preferred the spelling with a D ending. This resulted in considerably more errors on present tenses than on participles, even when homophone frequencies were matched (cf. results for D $=$ T) and even more so when these frequencies also favored the D spelling (cf. the interaction for D > T). This bias toward a D spelling was only eliminated when the whole-word frequencies considerably favored the T form (cf. findings for D < T). This preference for the D ending can be explained by the different co-occurrence frequencies in Dutch homophonous verb forms between the be- and ver- prefixes (the two prefixes used in the experiment) and the D and T endings: 76,452 for verb-final D and 47,019 for verb-final T in the CELEX corpus. Hence, verbs with be- and ver- prefixes are in general more likely to occur in the past participle form than in the third-person singular present form.

4. GENERAL DISCUSSION

In the present experiment, experienced writers of Dutch made errors on verb forms whose spelling can be derived by the application of very straightforward morphographic rules. Note that this occurred despite the participants' heightened awareness of the spelling of these forms given the context of a dictation task. The relationship between the homophone frequencies, together with distance, influenced the participants' spelling behavior. More spelling errors were made when a homophonous verb form of higher frequency existed alongside the target form and when the target form and the word determining its suffix were separated by a number of words. These effects were consistent across two experiments, using different types of homophonic verb forms. In addition to these factors, the co-occurrence frequency between the prefixes used in Experiment 2 and the D suffix also determined spelling accuracy.

In order to explain this data pattern, we start from the simplest model possible: Writers spell verb forms by relying on the relevant morphosyntactic input and determining the appropriate orthographic output pattern. Let us call this the computational procedure for verb endings. (We are not concerned with the details of this procedure, only with its existence.) In effect, this is the only way to spell a homophonic verb correctly all of the time. However, many errors do occur and, as we have shown, they occur in predictable patterns. In addition, when a prefix is highly associated with a particular spelling of a suffix, this spelling is often the one which is produced. These findings show that spelling patterns are not only determined by computational procedures but also by retrieval processes which operate on the basis of learned associations between phonological input and orthographic output patterns. Moreover, the impact of these retrieval processes on subjects' spelling is stronger when the verb form is removed from the word controlling its spelling.

There are two ways to interpret this. First, the computational process may fail to come up with the spelling of the inflectional ending before the retrieval processes have produced an output. This retrieval output is subject to (co-occurrence) frequencies. When this is the incorrect, higher frequent form, subjects may erroneously rely on this output and not wait for the computational process to terminate. According to such an *interference* account of verb spelling errors, the fact that a response announces itself, brings subjects off the right track. This will happen more often in cases where the computational procedure has to rely on information that is at some distance from the verb and takes some time to be identified. It might even be that the computational procedure is never started because subjects are distracted or are not able to identify the morphosyntactic information in time. In this case, they will rely on the first response becoming available, which will be the one delivered by a retrieval process.

However, it might be not so much the relative speed of these two processes that determines the probability of accurate spelling but rather a kind of uncertainty that the speller experiences when presented with *two* orthographic patterns (when one is computed and the other retrieved, or when both are retrieved as stored orthographic forms). This uncertainty will be greater when the spelling predicted by the relevant morphosyntactic information clashes with the information coming from the verb, i.e., the fact that one form is spelled more often than the other, or that the suffix spelling is more often associated with the prefix. When the word containing the morphosyntactic information is more distant, the verb information will be more predominant, leading to more mistakes.

Although we have proposed two versions that make reference to distinct processes, the results can also be explained by an interactive connectionist framework in which representations are distributed. Connectionist networks that map spelling to sound (e.g., Plaut et al., 1996; Harm & Seidenberg, in press) have shown to be particularly sensitive to word frequency and spelling-sound consistency. Given the fairly close relationship between phonemes and graphemes in a language like Dutch (cf. Van Orden et al., 1990), it might be straightforward to set up a network that handles sound to spelling consistencies. However, to account for the results of our experiments, such a model will have, a.o., to represent long-distance dependencies across different syllables in order to explain co-occurrence frequency effects. At the moment, no such model exists.

REFERENCES

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Feldman, L. B. (Ed.) (1995). *Morphological aspects of language processing.* Hillsdale, NJ: Erlbaum.

Harm, M. W., & Seidenberg, M. S. (*in press*). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review.*

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review,* 103(1), 56–115.

Sandra, D., & Taft, M. (Eds.) (1994). *Morphological structure, lexical representation and lexical access.* East Sussex: Erlbaum.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review,* 97, 488–522.