Rules and Exceptions in the Spelling of Loan Words in Dutch

Anneke Nunn and Anneke Neijt

0. Introduction

Knowledge of systematic and unsystematic properties of spelling is useful for educational purposes, for applications like speech-to-text conversion systems and spelling correctors, and for the compilation of spelling dictionaries. Furthermore, the most systematic and restrictive account for present day spelling may ultimately shed light on the question to what extent reading and writing actually involves spelling rules. To be able to test the psychological reality of spelling rules, we must first have an idea of what these rules might look like.

To gain insight into Dutch spelling we investigated to what extent the current spelling of Dutch words is predictable from the pronunciation. For that purpose, a set of rules which convert the sounds of words to their spelling is being compiled and applied to the pronunciation of words by means of a computer program. The pronunciation of words was taken from the Celex database (Center for lexical information). By comparing the spelling predicted by these rules with the actual spelling also listed in the database, it is possible to evaluate the accuracy of the conversion rules. The resulting rule set derives the current spelling of words, with a residue of a few types of exceptions.

We took the traditional spelling rules, such as the rules in the Dutch orthography dictionaries, cf. Woordenlijst (1954), Woordenlijst (1995), as the starting point of our conversion rules. In addition we took into account the four spelling principles which govern the application of these rules according to Te Winkel, who first published official Dutch spelling prescriptions. These principles are the following (cf. Te Winkel 1863:8-15):

**Principle of Received pronunciation**
Reprensent by means of letters all the parts that are heard in a word when it is pronounced correctly by civilised people.

**Principle of Uniformity**
Give the same shape to the same word and every part of which it is made, as far as the pronunciation allows this.

**Principle of Analogy**
Words of which the spelling is determined neither by the pronunciation, nor by uniformity, nor by etymology, are written in the same way as others, of which the spelling is known and which are apparently formed in the same way.
Etymological Principle
When choosing between two spellings for the same sound, the derivation or older form is decisive.

It is interesting to note that Te Winkel’s spelling principles have hardly changed previously accepted writing conventions, which suggests that they are not only prescriptions but can also be seen as descriptions of writing practice as it existed at that time (cf. Cohen and Kraak 1972:29). Furthermore, the current spelling is still essentially the same, so these principles may still be valid. We expected that potential shortcomings of rules and principles would show up as incorrect predictions when implementing them in a computer programme. In some cases, the rules indeed had to be adjusted.

In this article, we will not attempt to discuss the whole spelling system, but focus on the Etymological Principle and examine the set of words where this principle is particularly relevant, namely loan words. A strict interpretation of the Etymological Principle (as applied to loan words) implies that the spelling of loan words cannot be derived from their pronunciation by (Dutch) spelling rules. However, we will show that this implication is not correct: in some cases the spelling of loan words is not arbitrary, but can be described by sound-to-letter conversion rules.

This article is organised as follows. In section 1, we will argue that the spelling of loan words is not inherently irregular, contrary to what one may conclude from the treatment of loan words in the orthographic dictionary and overviews like chapter 9 of Booij (1995) where possible spellings for one sound are simply listed, and no attempt is made to discover regularity in their distribution. In the second section, an additional type of conversion rules is introduced. These rules do not refer to the pronunciation of words; they only adjust letter sequences. Section 3 investigates whether this additional type of rule also applies to loan words.

1. Is the spelling of loan words inherently irregular?

Te Winkel’s principles sometimes make conflicting predictions. In case of loan words the Principle of Received Pronunciation states that the pronunciation determines the spelling irrespective of the origin of the words in question. In terms of the computer programme we are developing, this principle leads to a programme which derives the spelling of all loan words by sound-to-letter conversion rules; we will call this option

A: The spelling of loan words is regular.
The Etymological Principle, on the other hand, states that the older spelling of a word is decisive (as we will only discuss underived words in this section, the other two principles can be ignored). The older form of loan words is the spelling in the donor language. This means that spelling may at the most be derived by spelling rules of the donor language, but not by Dutch spelling rules. The possibility of deriving the spelling of loan words by foreign rules is discussed in Neijt (1994). It was shown there that such rules do not accurately derive the spelling of Dutch loan words. Consequently, the Etymological Principle implies that the computer programme must contain a list in which all loan words can be looked up; let us call this option

B: The spelling of loan words is inherently irregular.

A practical argument against option B is that loan words which happen to be written in the same way as native Dutch words, for instance *pasta* ‘paste’, are treated just like words with an idiosyncratic spelling like *cynisch* ‘cynical’. In option A, words like *pasta* are automatically accounted for, and we need only store words like *cynisch* as exceptions.

A more theoretical argument against option B is that if all loan words are considered to be inherently outside the Dutch spelling system, one may argue that the spelling of loan words is never adapted to the native pattern; regularisation of these words may never take place. However, such spelling adaptations have occurred both before and after the formulation of official spelling regulations. In (1a) we have listed some words cited by Te Winkel as examples of words that had already been (partially) adapted before he formulated his rules (Te Winkel 1865:178), and in (1b) examples of words of which the spelling was changed by the spelling dictionary of 1954:

(1) Spelling changes

<table>
<thead>
<tr>
<th>Old</th>
<th>New</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>‘old’</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>melon</em></td>
<td><em>meloen</em></td>
<td>‘melon’</td>
</tr>
<tr>
<td><em>cigarro</em></td>
<td><em>sigaar</em></td>
<td>‘cigar’</td>
</tr>
<tr>
<td><em>république</em></td>
<td><em>republiek</em></td>
<td>‘republic’</td>
</tr>
</tbody>
</table>

---

1 This is also the view of Te Winkel: ‘The spelling of foreign words and partly adapted words need not be derived by new and special rules. It is given: as far as the stem is concerned by the foreign spelling, as far as the endings are concerned by the normal Dutch spelling rules’ [De spelling van vreemde woorden en bastaardwoorden behoeft niet gezocht en eerst door nieuwe, afzonderlijke regels bepaald te worden. Zij is gegeven: wat het hoofddeel der woorden aangaat door de vreemde spelling; wat de uitgangen betreft, door de gewone Nederlandse spelregels.] (Te Winkel 1865:183).
However, the most important reason to reject option B is the existence of regularity within the spelling of loan words. Consider the sounds /i/ and /z/, which are written as <ie> and <z> in native words, and as <ie> or <i> and <z> or <s> in loan words. Interestingly, the distribution of <i>-<ie> and <s>-<z> in loan words is regular. The pattern in the spelling of /i/ in loan words is recognised in the orthographic dictionaries; for the pattern in the spelling of /z/ cf. (Nunn and Neijt 1996):

(2) Writing conventions for /i/ and /z/ in loan words

   a write /i/
      i as <ie> in the last syllable of a morpheme
      ii as <i> elsewhere
   b write /z/ after a long vowel or sonorant consonant
      i as <z> morpheme finally
      ii as <s> elsewhere

For example:

(3) a native words
   /i/ gieter, geniet 'watering-can, to enjoy'
   /z/ ezel, deze 'donkey', 'this'

   b loan words
   /i/ i. kilo, liter, vanille id., id., 'vanilla'
    ii. ventiel, manier, olie 'valve', 'manner', 'oil'
   /z/ i. mensa, basis id., id.
    ii. precieze, poreuze 'precise', 'porous'

(3b) illustrates that we find the same spelling (<ie>, <z>) as in native words in the final syllable of loan words, and 'foreign' spellings (<i>, <s>) elsewhere. Pairs like kilo-ventiel may suggest that <ie> is written in closed syllables only, in order to

---

2 These rules are the same as the rules in the previous edition of the orthographic dictionary, except for the fact that those rules explicitly refer to native and non-native suffixes. Here it is assumed that non-native endings like -ant and -isme do not count as morphemes for the spelling rules in (5), cf. Friezin-frisisme ('frisian woman, frisian term'), fabrieken-fabrikant ('factories', 'manufacturer').

3 The syllable final <s> in words like precies is derived by a spelling rule changing <z> to <s> (cf. Nunn 1992). This rule applies to the output of rules like (2). Other rules of this type (adaptation rules) will be discussed in section 2.
prevent an incorrect reading ([*vεntIl]), but pairs like vanille-olie contradict this suggestion: in va-nil-le we find <i> in a closed (orthographic) syllable, whereas <ie> in olie is not necessary to prevent an incorrect reading, since <i> may only correspond to a long vowel in word final position. Therefore it is not syllable structure, but the position in the morpheme that is relevant.

Additional support for the rules under (2) is given by spelling changes. Loan words of which the spelling has been changed have not adopted the pattern of native words, but the spelling predicted by rules such as those in (2). The words which were originally written as (4a), for instance, are now written as (4c), and not as (4b):

(4)  

<table>
<thead>
<tr>
<th>a</th>
<th>viril</th>
<th>‘virile’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>désastreux</td>
<td>‘disastrous’</td>
</tr>
<tr>
<td>b</td>
<td>*vieriel</td>
<td>*dezastreuje</td>
</tr>
<tr>
<td>c</td>
<td>viriel</td>
<td>desastreueze</td>
</tr>
</tbody>
</table>

The rules for /i/ do have exceptions, e.g. receiver/*riciver, psycholoog/*psicholoog, gienje/*ginje, pi/*pie and taxi/*taxie (id., ‘psychologist’, ‘coin’, id., id.) and prefixes like anti-/*antie (id.). The rules for /z/ are not without exceptions either, e.g. confuse/*confuze, rozijn/*rosijn and azuur/*asuur (‘confused’, ‘raisin’, ‘azure’). Still, these rules give a more accurate and restrictive description of the spelling of loan words than the rules for native words.

Summarising, a spelling system in strict accordance with the Etymological Principle (option B) is undesirable since it implies that regular loans are treated as exceptions, since it is incompatible with the observation that the spelling of exceptional loan words is often adapted, and since it fails to account for the generalisations under (2) about the spelling of loan words. A set of sound-to-letter rules based on option A, with special rules for loan words is therefore to be preferred. Words of which the spelling may not be derived by these rules are listed as exceptions (e.g. cynisch instead of *sinies).

2. Two types of spelling rules

If there were just one type of conversion rules we could end this article here. However, a description of the Dutch spelling system with just one type of conversion rules leads to some incorrect spellings. Earlier findings showed that these incorrect predictions could be avoided by using two types of rules with different properties (cf. Nunn 1992). In this section we will briefly illustrate why it is necessary to postulate a second rule type, before examining the treatment of loan words by both rule types (section 3).
When morphemes are combined in words, this sometimes results in a sequence of two identical consonants:

\[(5)\]
\begin{align*}
a & \text{ haat + te } & \text{ haatte/*haate} & \text{‘hated’ (simple past)} \\
b & \text{ ge + haat + t} & \text{*gehaatt/gehaat} & \text{‘hated’ (past participle)}
\end{align*}

In (5), the sequence \(<tt>\) is sometimes simplified and sometimes remains unchanged. The simplification in *gehaat* may seem to be the effect of the sound rule degemination which is reflected in orthography (thus violating the principle of Uniformity):

\[(6)\]
\[\begin{array}{c}
/\gamma\hat{d}\ hat\ t/ \\
[\gamma\hat{d}hat]
\end{array}\]

However, deriving the spelling of the words from the surface pronunciation, which has been affected by degemination, also leads to the incorrect *haate* instead of *haatte*. The mismatch between sounds and spelling illustrated in (5) made some linguists conclude that Dutch orthography is simply inconsistent, (cf. Booij 1985).

However, the pronunciation of both words in (5) may be the same, but the spelling, or more specifically the orthographic syllable structure, is different: in (5a) the identical letters are heterosyllabic (*haat-te*), in (5b) tautosyllabic (*gehaatt*). It is possible to make a generalisation about the facts in (5) by referring to letters only:

\[(7)\]  
**New type of conversion rule:**
Delete one of two tautosyllabic consonant letters.

With this rule, the contrast between *haatte* and *gehaat* is not inconsistent, but regular. Additional support for (7) is formed by the fact that degemination also occurs after silent letters: adding a plural suffix \(<s>\) to the word *bourgeois* leads to *bourgeois*, not *bourgeoiss*.

To derive the correct spelling for words such as those in (5) we introduced a second rule type: first the abstract pronunciation is converted to an abstract spelling representation by rules to which we will refer as sound-to-letter conversion rules, and then the resulting letter sequences are modified by rules which will be denoted by the term adaptation rules:

\[
\begin{array}{c}
\text{(abstract) pronunciation} \\
\downarrow \\
\text{sound-to-letter conversion rules} \\
\downarrow \\
\text{underlying spelling} \rightarrow \text{adaptation rules} \rightarrow \text{surface forms}
\end{array}
\]
As already shown in (5), adaptation rules such as degemination must operate across morpheme boundaries. The same holds for other adaptation rules, such as consonant doubling (gram-grammen (gram+en), ‘gramme’-‘grammes’ (and the rule mentioned in footnote 3, which changes <z> into <s>). On the other hand, the context of sound-to-letter conversion rules is restricted to the domain of the morpheme. For example, consider the (simplified) writing convention for the sound /ʒu/ (cf. Woordenlijst 1995:21-22):

(8) write /ʒu/ as <ou> before a consonant and as <ouw> elsewhere

This rule accounts for the difference in spelling between boud ‘brazen’ and trouw ‘fidelity’. If the conversion rule were sensitive to information outside the domain of the morpheme, (8) would derive the spelling *getroud instead of getrouwd, ‘married’ when a suffix d is added to trouw.

In summary, for an adequate spelling description we need two rule types with different properties: sound-to-letter conversion rules with the morpheme as their domain and adaptation rules with the word as their domain.4

3. Adaptation rules and loan words

In section 1, we did not distinguish sound-to-letter conversion rules from adaptation rules. Now this distinction has been introduced we can try to find out what class the rules in (2) belong to. The fact that adaptation rules and sound-to-letter conversion rules have different domains of application may be used as a criterion for classifying rules. According to this criterion, the rules in (2) are conversion rules. /i/, for instance, is only written as <ie> in the last syllable, cf. rivier ‘river’, but the addition of a suffix such as the plural suffix -en does not change the spelling of the final /i/: the correct spelling is rivieren not *riviren (cf. monomorphematic visibel, ‘visible’). Since all rules discussed so far were sound-to-letter conversion rules, we do not yet know whether adaptation rules behave differently with respect to native words and loan words. In order to find this out, we will take a closer look at one rule, namely consonant doubling, (cf. Woordenlijst (1995:19):

4 There are some similar proposals in the literature: Wester (1984) also uses two rule sets for the reverse operation, i.e. text-to speech conversion: letter-to-sound conversion rules and phonological rules. Berendsen e.a. (1986), also use these two rule types as well as letter-to-letter rules which insert morphological boundaries.
(9)  **Consonant doubling**

double an intervocalic consonant after a short vowel

The distribution of consonant geminates and single consonants in loan words is illustrated under (10).\(^5\) (10a) contains words that seem to have undergone (9). Most words fall in this class. (10b) gives exceptions to (9) and (10c) lists words that have geminates in a context that does not satisfy the conditions of (9):

(10)  

a  regular according to (9)

additioneel, antenne, broccoli, cello, grutto, mammoet, officier, rabbì, terrein, zeppelin, etappe, gekko, massa ('additional', 'antenna', id., id., 'godwit', 'mammoth', 'officer', id., 'terrain', 'blimp', 'stage', 'mass')

b  irregular (exceptional single consonant letters)

accelereren, cabaret, nasi, impresario, spin(n)aker, comité; polaroid, image ('to accelerate', id., 'indonesian dish', id., id., 'comittee', id., id.)

c  irregular (exceptional geminates)

toss, jazz, yell; Lloyd (id., id., id., id.)
croissant, hausse, parallel, regisseer, concurreer, ('pastry', 'hype', id., 'to direct', 'to compete')
saffraan, abbrevieer, shuttle; manggis ('saffron', 'to abbreviate', id., 'mangis')

In line with the general approach followed in section 1, we may consider two options to describe double consonants in loan words. In the first, we follow the Etymological Principle and state that the spelling of loan words is given, so we do not apply rules such as consonant doubling to derive their spelling. In the second, we apply adaptation rules to all words irrespective of their origin. We now have option A' and B' for adaptation rules:

A': Loan words are subject to adaptation rules.
B': Loan words are not affected by adaptation rules.

---

\(^5\) A problem for the choice between single consonant letters and geminates is the fact that it is often hard to determine the quality of vowels. Consider for instance [dlɪdɔnt-ɪlyzi], dilettant-illusie ('id.', 'illusion'), [kɾɔlɔsal-kɾɔlɛɣa], kolossaal-collega ('colossal'-'colleague'). This issue will not be addressed here.
Option B' implies that the distribution of single consonants and geminates is not derived by rules, and that the spelling of words like *nasi* and *jazz* is just as irregular as that of words like *cello*. Under option A' on the other hand, geminates like those in (10a) are the result of consonant gemination, and only words such as *nasi*, *jazz* must be stored as exceptions. Once again the description that violates the Etymological Principle is most attractive, since it has fewer exceptions.

Another reason to prefer option A' is the fact that there are indications that words like *nasi* are exceptional. As observed in section 1, we expect no regularisation when loan words are considered to be inherently outside the Dutch spelling system. However, spelling changes which can be interpreted as the removal of exceptions to adaptation rules have in fact occurred:

\[
\begin{align*}
\text{(11)} & \quad \text{mafia, rafia} \rightarrow \text{maffia, raffia} '\text{mafia}', \text{id.} \\
& \text{tafia, desa} \rightarrow \text{taf(f)ia, des(s)a} '\text{tafia}', '\text{dessa}'^6 \\
& \text{gramophone} \rightarrow \text{grammofoon} '\text{gramophone}' \\
& \text{albatross} \rightarrow \text{albatros} '\text{albatross}' \\
& \text{mandrill} \rightarrow \text{mandril} '\text{mandrill}'
\end{align*}
\]

The presence of 'spelling pronunciations' also supports option A'. For instance, consider words written with geminate consonants after an originally long vowel which often give rise to a pronunciation with a short vowel. On the other hand, words without geminates after an originally short vowel are often pronounced with long vowels:

\[
\begin{align*}
\text{(12)} & \quad \text{spelling} & \quad \text{original pronunciation} & \quad \text{changed pronunciation} \\
& \text{a} & \text{rotan} ('\text{rattan}') & [r\text{ɔt}\text{an}] & [\text{rot}\text{ɔn}] \\
& & \text{Donald (id.)} & [d\text{nəld}] & [\text{don}\text{ld}] \\
& \text{b} & \text{villa (id.)} & [\text{vila}] & [\text{vIla}]
\end{align*}
\]

This phenomenon suggests that native speakers assume that the spelling of loan words is not inherently irregular, but forms an accurate representation of the pronunciation.

We may conclude that Option A', a spelling system which is in strict accordance with the Etymological Principle, is to be preferred, since it gives the most restrictive description of the spelling facts, and since it is compatible with spelling changes. However, unlike what we saw in section 1, the choice for option A' is not supported by a separate non-native regularity here: there are words with regular geminates and exceptions, but there is no evidence that suggests that

---

6 The spelling variants *taffia* and *dessa* were abolished in 1995.
consonant degemination is applied differently in native words and loan words. Other adaptation rules, such as vowel degemination and diaeresis can also be derived by the same rules that are valid for native words. Again there is a residue of exceptions, for instance, vowel degemination has exceptions like bazooka (id.). However, there is no evidence for a separate non-native pattern. Thus, other adaptation rules do not distinguish between native words and loan words either. In this respect adaptation rules differ from sound-to-letter conversion rules.

In the approach chosen here, differences between the spelling of native words and loan words are exclusively accounted for by sound-to-letter conversion rules, and not by adaptation rules. In a way, the fact that only sound-to-letter conversion rules are sensitive to the distinction between native words and loan words supports the postulation of two distinct types of rules.

4. Conclusion

We started an extensive investigation of the Dutch spelling, using a set of spelling rules, and testing their effectiveness by means of a computer programme. In this article, we focused on one aspect of this investigation, namely the question to what extent the spelling of loan words in Dutch is regular. Dutch spelling rules for native words can be divided into two types: sound-to-letter conversion rules and subsequent adaptation rules. We found that both rule types also apply to loan words, even to the extent that they induce modifications of irregular spellings. In addition, we found that native words and loan words are described by different sound-to-letter rules, but that they are subject to identical adaptation rules.

References


