Anneke M. Nunn

Dutch Orthography

A Systematic Investigation of the Spelling of Dutch Words

TWO-STEP

bob nunn
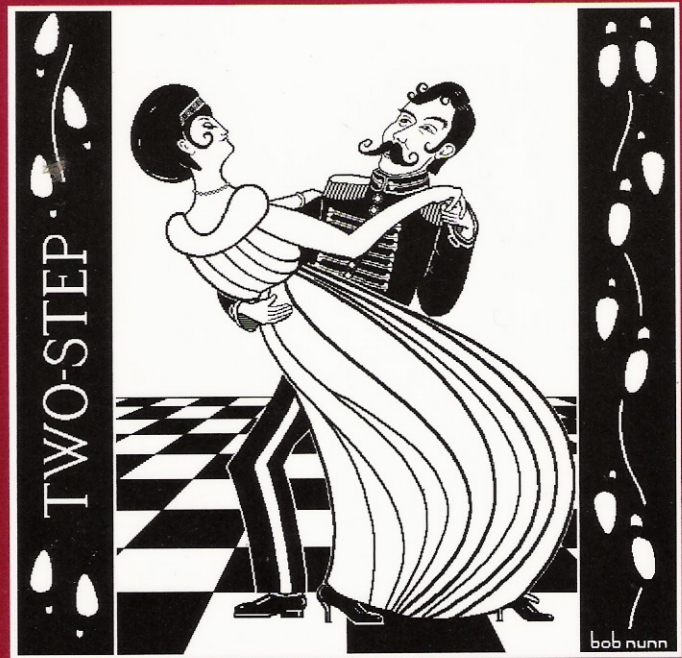
CENTER FOR LANGUAGE STUDIES

: LOT

Netherlands Graduate School of Linguistics

Landelijke Onderzoekschool Taalwetenschap

Dutch Orthography

Dutch Orthography

A Systematic Investigation of
the Spelling of Dutch Words

Een wetenschappelijke proeve op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op donderdag 25 juni 1998,
des namiddags om 3.30 uur precies

door

Anneke Marijke Nunn
geboren op 3 oktober 1965
te Billericay, Groot-Brittannië

Promotor: Prof. dr. A.H. Neijt-Kappen

Manuscriptcommissie: Prof. dr. G.E. Booij (Vrije Universiteit Amsterdam)
Prof. dr. G.R.W. Dibbets (voorzitter)
Dr. V.J.J.P. van Heuven (Rijksuniversiteit Leiden)

Printed in The Netherlands.

CONTENTS

Chapter 1

Introduction

> If writing has so much influence on language, if it can promote or delay the reception of thoughts, then the way one writes deserves to be the object of serious consideration.[1] Te Winkel (1863:8)

Speech is a wonderful means of communication, but it is not permanent. For this reason, visual ways to preserve speech have been developed for many languages and adopted and adapted by others. Some of these visual systems consist of icons of the concept referred to that are directly interpretable. Other systems use symbols which do not directly refer to meaning, but to the language that refers to that concept. Writing then consists of symbols which refer to linguistic entities like words, morphemes, syllables or sounds, see for instance Gelb (1963).

In alphabetic systems, spelling represents sounds, and it does so in a linear and sequential fashion. This is a very efficient way of encoding language. A small number of letters suffice to encode all words. Learning to read and to write a language with a perfect alphabetic spelling system equals learning the regular correspondences between letters and sounds. However, most spelling systems are not of this ideal type. Deviations from one-to-one correspondence between sounds and letters can be caused by the fact that there simply are not enough letters to uniquely represent each sound or by the fact that some spelling rules maintain a visible relation between an original word and its derived forms although they are pronounced differently. Other causes of sound-letter mismatches are changes in pronunciation in the course of time that are not reflected orthographically or the influx of loan words that are written according to the spelling of the language of origin. When the number of irregular words becomes too large, spelling can no longer be seen as a code for the pronunciation, but becomes an arbitrary code for words.

---

[1] "Indien aan het schrift een zoo groote invloed op de taal moet toegeschreven worden, indien het ook het opnemen der gedachten kan bevorderen of vertragen, dan verdient de wijze, hoe men schrijft, een voorwerp van ernstige overweging uit te maken."

However, the regularity or irregularity of a spelling system as a code for the pronunciation should not be established too hastily, since it is possible that the right generalizations have not been made yet, see Carney (1994:16) and Wester (1987).[2] Only thorough investigation of a spelling system will enable us to judge its regularity or irregularity.

The topic of this study is the spelling system of Dutch. This language has an alphabetic spelling system, but a perfect one-to-one correspondence between sounds and letters does not exist in Dutch. Some letters represent different sounds. Both schwa and [ɛ], for instance, are represented by <e> in a word like *emmer*-[ɛmər]. Conversely, a sound can be represented by different letters or combinations of letters, depending on its specific context in the word. For this reason, spelling rules are needed to relate sounds in a given context to the appropriate spelling.

1.1 Aim of this study

In this study, I aim at providing a systematic, explicit and complete description of all generalizations about present-day Dutch spelling (of native words and loan words), and at providing a framework for integrating spelling rules in a transparent way (with respect to their domain, context (sounds or letters), ordering, etc.). A complete description of Dutch spelling is not yet available. It is necessary to fill this gap, since descriptions of parts of the system only could cause problems in other parts.[3] Such problems can be avoided only when all generalizations are combined.

Starting point
The basic spelling principles have already been given explicitly by Te Winkel, who was the first to publish official spelling regulations for the Netherlands and Belgium. It should be noted that Te Winkel's spelling is not an artificial construction, but more or less a description of the existing writing conventions. This suggests that these principles do not only *prescribe* how to write but also *describe* writing practice as it existed at that time, see Cohen & Kraak (1972:29). Te Winkel's principles, published in *Grondbeginselen der Nederlandsche spelling* (Principles of Dutch orthography, 1863), are the following:

Principle of Received Pronunciation

---

[2] There are more generalizations than is obvious at first sight. Venezky (1970), Klima (1972) and Carney (1994) even showed this for a notoriously complex orthography such as that of English.

[3] For instance, Wester (1985b), which deals with spelling alternations in related words, and Wester (1987), which describes the spelling of fricatives, make mutually contradictory assumptions about the abstractness of phonological representations which are encoded in Dutch spelling.

> Represent by means of letters all the constituent parts that are heard in a word when it is pronounced correctly by civilized people.

Principle of Uniformity

> Give the same word and every constituent part the same shape, as far as the pronunciation allows this.

Principle of Etymology

> The choice between competing spellings for one sound is determined by the derivation or by the older form that was used when pronunciations that are now identical could still be distinguished clearly.

Principle of Analogy

> Words whose spelling is determined neither by the pronunciation, nor by uniformity, nor by etymology, are written in the same way as others whose spelling is known and that are apparently formed in the same way.

The specific spelling rules based on these principles were refined and supplemented by a word list in *Woordenlijst voor de Spelling der Nederlandse Taal* in 1866. They have been reformed in several ways since, which was reflected in the publication of new versions of the *Woordenlijst* in 1954 and 1995. However, these reforms have never changed the basic properties of Dutch spelling. I will take Te Winkel's principles as a starting point for the description of present-day Dutch spelling.

Generalizations that are not yet captured by Te Winkel's Principles
A systematic inspection of the *Woordenlijst* reveals that there are generalizations about Dutch spelling that Te Winkel did not mention, and maybe did not even notice. The contrast between <ie> and <i> in *neurie ~ neuriën*, for instance is not explained by any of Te Winkel's principles. This kind of sound-independent spelling alternations is too prominent in Dutch spelling to be considered a mere inconsistency, cf. also examples *raam ~ ramen, stem ~ stemmen*. It seems that a higher-ranking principle induces violation of the principles, but Te Winkel does not propose such a principle. Note that these alternations lead to differences between words and their inflected forms and thus violate the Principle of Uniformity.

   In this study, I will formulate the Graphotactic Principle for such spelling patterns of Dutch, in addition to known spelling principles. I will also try to improve specific spelling rules, whose formulation in different editions of the Woordenlijst is sometimes insufficiently explicit or even incorrect, see for instance Kollewijn (1916), Wester (1987) and Neijt & Zuidema (1994a). For this purpose I will consult studies of parts of the spelling system such as Booij et al. (1979), Van Heuven (1980), Zonneveld (1980), Kerstens (1981), Dibbets (1983), Booij (1985, 1987, 1991, 1995), Wester (1985a, 1987, 1989), Neijt & Zuidema (1994a), Zuidema et al. (1994) as well as reform proposals that often

formalize current spelling patterns, see Geerts et al. (1988), Neijt & Zuidema (1994b).[4]

In the remainder of this introductory chapter, I will argue insight into the spelling system is important. I will discuss the relevance, method and exact scope of the present investigation. Finally, the organization of this thesis is outlined.

1.2 Relevance

As mentioned in the previous section, there is no satisfactory description of Dutch spelling. Apart from filling this gap, the present investigation is relevant in the following ways.

Firstly, a systematic description of Dutch spelling may contribute to the disentanglement of phonology and writing. Abstract phonological representations are sometimes postulated when differences between pronunciation and spelling of a word are interpreted as the effect of a sound rule of which spelling abstracts rather than of a writing convention. An example is formed by the analysis of English sound representations in Chomsky (1970), but see Derwing (1992:194) for arguments against Chomsky's approach. The study of Dutch orthography may help recognize and avoid the postulation of similar mistaken phonological analyses for Dutch, see for instance Appendix G.

Secondly, the investigation may ultimately shed light on the writing process. In order to test to what extent writers actually use spelling rules, we must first have an idea of what these rules might look like. The compilation of formal and explicit spelling rules can therefore be seen as a prerequisite for that type of research. The spelling rules proposed in this study were compiled with this possibility in mind. I therefore avoided formulating the rules in such a way that they immediately seem useless to the writer. The rules proposed exclusively refer to information available to all literate speakers of Dutch. They do not, for instance, refer to historical or dialectal pronunciations, or the etymological origin of words. However, since I looked systematically for generalizations, it is possible that some of the rules proposed here are not identical to rules that are used by most native speakers, see Mohanan (1986:58).[5] One reason why this

---

[4] Most research with respect to Dutch spelling has focused on topics related to spelling such as the teaching of the spelling of inflected verbs: Van der Velde (1956), Assink (1983), Zuidema (1988); spelling errors: Horbach-Kleijnen (1988), Van Luijn-Hindriks (1992); learning to read: Van Rijnsoever (1988), Reitsma (1992); general strategies of the writer: Verhoeven (1985); and the readability of spelling: Van Heuven (1978).

[5] "In the absence of clear evidence, we are forced to make guesses about which of the patterns have been internalized by a language user and which of the patterns are simply accidental correlations in the corpus. As soon as clear evidence becomes available from psycholinguistic experiments, we must be willing to revise our initial guesses on the basis of the evidence."

may be the case is the following. The spelling of some words (particularly loan words) is the result of extra-linguistic conventions introduced by spelling committees. In these cases we cannot be sure that the official spelling of words reflects writers' intuitions.

Finally, insight into the Dutch spelling system also serves more practical purposes:

- The explicit spelling rules proposed here can be used to facilitate the teaching of writing.
- Spelling rules may be used to improve the consistency of the spelling in dictionaries. By determining spelling by rules, similar words will be automatically treated the same way, see for instance Zuidema et al. (1994). In addition, the rules allow for a principled choice between spelling variants. The variant that follows from the rules is preferred.
- Rules that relate sounds and letters can be used in linguistic applications such as text-to-speech systems, see Heemskerk & Van Heuven (1993), Nunn & Van Heuven (1993), Rietveld & Van Heuven (1997).

1.3 Method and scope

Method

In order to gain insight into the Dutch spelling system I evaluated the spelling rules and principles in the literature. I implemented the rules which seemed most adequate in a computer programme in order to apply them systematically to a large set of words. For this purpose, I used a lexicon which consists of the set of monomorphemic words, derivations and some inflected forms from the *CELEX* database (*CEnter for lexical information*), see [*CELEX 1990*], Burnage (1990). This lexicon will be called the 'test lexicon' in this study. By inspecting differences between the spelling computed by these rules and the actual spelling, I could automatically trace flaws, omissions or ambiguities in the rule set. The rules were adapted accordingly in order to diminish discrepancies between actual and computed spelling of words.

Of course, it is possible to give a maximally accurate description by also describing exceptions by means of rules. However, in such a description, the spelling of (monomorphemic) words would be arbitrary, and we could not formulate restrictions on well-formed written words, or decide whether a new word has a regular spelling; exceptional and systematic spellings would be treated the same way. A description that provides insight must strike a balance between accuracy and generality.

Scope

It is not possible to discuss all aspects of Dutch spelling in this study, so some limitations of the subject matter were necessary:

- I will restrict the scope of the investigation to spelling phenomena which occur in prosodic words, that is monomorphemic words, and their derived or inflected forms. Consequently, spelling phenomena that only occur in compounds and word groups are not discussed here. Examples of these phenomena are capitalization, punctuation and the realization of the linking morpheme, e.g. ber*en*kuil (*beer + kuil*) versus ber*e*leuk (*beer + leuk*). This decision was motivated by the consideration that spelling phenomena that occur in compounds and word groups are of a different nature than those described in this thesis. They represent syntactic or semantic information or prosodic structure rather than sounds, see Booij (1985:51–54, 1987:219–223, 1995:187–188). For instance, capitals mark names or the beginning of sentences; spaces delimit grammatical words and punctuation indicates syntactic and prosodic constituents. The realization of the linking morpheme and, in some cases, placement of hyphens in compounds involves semantic information. A fruitful description of such phenomena calls for another approach than the one chosen here.
- I have primarily described the spelling system from the perspective of the writer, although the perspective of the reader is sometimes invoked to judge the readability of a spelling. For the latter perspective, the reader is referred to the literature about text-to-speech conversion, for instance Wester (1985a), Nunn & Van Heuven (1993), Rietveld & Van Heuven (1997).

1.4 Formalism and notational conventions

In this study, stress will be indicated by underlining the relevant syllable (e.g. *no<u>ta</u>ris*). Phonological syllable boundaries are indicated by '-' or [ ]$_\sigma$ (e.g. wa-tər, [wa]$_\sigma$[tər]$_\sigma$); orthographic syllable boundaries by '.' or [ ]$_S$ (e.g. *em.mer* or [em]$_S$[mer]$_S$). A word boundary is represented as '#'; a morpheme boundary as '+'.

To distinguish different levels of representation, I will assume the following notational conventions:

(1)   **abstract representation surface realization**
             Sounds   /brod/   [brot]
             Spelling \jamer\   <jammer> or *jammer*

Spelling rules are formalized as context-sensitive rewrite rules that translate phonemes into graphemes or modify grapheme sequences. Spelling rules are ordered in accordance with the Elsewhere Principle, see Kiparsky (1982), that is, a specific rule has priority over a general one. Further ordering statements are explicitly stated. The rules also refer to autosegmental spelling structure, see 2.4.

Since I am almost exclusively dealing with formal aspects of Dutch words (pronunciation or spelling) rather than their meaning, glosses will not be provided unless the meaning is crucially relevant, e.g. in the case of homographs.

1.5 Organization of this thesis

Chapter 2 focuses on the spelling of native words. To find the basic spelling rules of Dutch it is necessary to exclude loan words that often keep the spelling of the language they were derived from. It appears that even in native words the same phoneme can often be written in different ways. In most cases, the distribution of the different spellings for the same sound is predictable, but not always on the basis of the phonological context. Some spelling alternations are predictable on the basis of letter sequences rather than sounds. Hence, spelling is sound-dependent in some cases, but autonomous in others. On the basis of this observation, I introduced a model in which spelling is derived from the pronunciation in two stages: by phoneme-to-grapheme conversion rules and autonomous spelling rules. I will argue that the new model allows a better account of the distribution of different spellings for the same sound.

Chapter 3 focuses on the observation that written morphemes are often less variable than their spoken counterparts. Te Winkel therefore claims that spelling aims at a uniform representation of morphemes (as stated in the Principle of Uniformity). However, autonomous spelling rules introduce variation between related words. I will argue that a more adequate account for this phenomenon derives the uniformity of spelling from the fact that spelling encodes the abstract sound representation of morphemes. This account allows variation between related words as long as it is introduced by autonomous spelling rules. In some cases, however, spelling does represent the effect of sound rules across morpheme boundaries. For instance, the variable realization of the diminutive ending in [ram**pj**ə] from [ram] and [la**tj**ə] from [la] is visible in the written forms *raampje* and *laatje*. This can be accounted for by postulating that the different forms of the suffix count as separate morphemes, or by assuming that the Principle of Uniformity is overruled to prevent an incorrect reading: *raamtje* could be pronounced as *[ramtjə].

In chapter 4, the spelling of loan words is examined. I will show that, although the spelling of these words cannot be determined by the spelling rules for native words, they are not inherently exceptional. Instead they can be accounted for with a separate set of spelling rules. If there were no explicit criteria to distinguish indigenous words from loan words, distinct rule sets would not improve the description of the spelling system. However, we can recognize native words by certain phonological and morphological properties. All words

that do not share these properties are loan words. Although the spelling system is somewhat complicated by the postulation of distinct rule sets, the result is a much better description of the spelling of loan words with fewer exceptions.

Chapter 5 examines spelling alternations that do not correspond to variation in the pronunciation. These alternations constitute additional support for the introduction of autonomous spelling rules. It turns out that the spelling model proposed in chapter 2 has simpler, more general rules than provided by previous accounts. In addition, the model is more restrictive, since phoneme-to-grapheme conversion rules and autonomous spelling rules have restricted (and mutually exclusive) properties. We will find that, although some autonomous spelling alternations serve to increase readability, others are applied to satisfy graphotactic conditions, which underlines the fact that spelling is more than a code for the pronunciation. In this chapter I will also discuss the computation of orthographical syllables. In contrast to chapters 2 and 3, this chapter also deals with loan words, in anticipation of the conclusion that autonomous spelling rules apply to native words and loan words alike.

Chapter 6 summarizes conclusions of the present investigation and makes some suggestions for further research.

This study ends with some appendices. In Appendix A the computer programme for testing spelling rules is discussed. Appendix B gives the criteria used here to classify words as native or non-native. Appendices C, D and E give an overview of phoneme-to-grapheme conversion rules for native and non-native words, and autonomous spelling rules, respectively. Appendix F examines some complications regarding the choice between single or double consonants. Appendix G argues that the uniform spelling of morphemes in related words is sometimes caused by spelling conventions. Finally, Appendix H gives an overview of spelling reforms of the Dutch spelling system from 1804 onwards.

Chapter 2

The spelling of native words

2.1 Introduction

This chapter focuses on the spelling of native Dutch words. It is organized as follows. In 2.2, I will examine Dutch phonemes and their orthographical counterparts (graphemes), and take a first look at the correspondence between sounds and letters. It appears that most sounds can be written in different ways and that the distribution of different spellings is sometimes predictable on the basis of phonological context. Section 2.3 summarizes the literature on the relation between sounds and spelling. We will see that there are two extreme views on the nature of the spelling system: spelling is seen as an autonomous rule system unrelated to phonology, or as a code for the pronunciation. There also is one proposal that combines properties of both extreme views. In 2.4 I will argue that a description in terms of both phoneme-to-grapheme rules and autonomous spelling rules, which are not conditioned by the pronunciation, allows a more adequate and insightful account of the alternation of single and geminate letters. In 2.5 I will show how other types of spelling variation can also be accounted for in a spelling model with two rule types. The conclusions of this chapter are formulated in 2.6.

2.2 Dutch phonemes, graphemes and the correspondence between them

2.2.1 Dutch phonemes

To be able to relate sounds and spelling, it is necessary to abstract from several aspects of speech. Firstly, the continuous sound signal has to be segmented into

a sequence of discrete segments which can be represented by letters.[1] Secondly, Dutch orthography does not provide a phonetic transcription of speech. A phonetic transcription (even one which abstracts from some types of variation, see Vieregge 1985:34–35) is more detailed than spelling, since it represents predictable contextual differences of phonemes of which speakers are not even aware, such as the variation in length of the vowel in *maal* and *maak* and different realizations of /r/, see also Cohen & Kraak (1972:20). Spelling encodes phonemes and abstracts from these allophonic variations (we will see below that actually spelling does not provide a phonemic, but a morpho-phonemic representation).

Dutch has the following set of native phonemes: consonants, short and long vowels, diphthongs and schwa. For a discussion of Dutch phonemes, see for instance Zwaardemaker & Eijkman (1928:214), Van Wijk (1939:39, 44), Cohen et al. (1978:43), Booij (1995:4–20). For an alternative account of the length distinction see Oostendorp (1995).

The following enumeration of phonemes does not contain palatalized consonants such as /ʃ/ /ɲ/ and /c/ since I will consider the underlying representation of these consonants to be /sj/ /nj/ and /tj/, see Booij (1995:7):

(1)   **Native phonemes of Dutch**
       Consonants
       plosives   p b t d k
       fricatives   f v s z x ɣ h
       nasals   n m ŋ
       liquids   l r

       Vowels
       short vowels ɪ ɛ ɔ ʏ ɑ
       long vowels   i y u e ∅ o a
       diphthongs   ɛi œy ɑu
       schwa[2]   ə

This phoneme inventory hinges on some decisions which will turn out to be

---

[1] The division of the continuous speech stream into discrete segments such as phonemes is an abstraction; it is not always possible to point out discrete entities. Spelling represents non-discrete speech by discrete entities, i.e. letters, which accounts for the fact that once literate, speakers can more easily distinguish phonemes, see Morais et al. (1986). Linguists like Aronoff (1982) and Kraak (1996) have even suggested that segments are not really psychological entities but merely inspired by the fact that orthography uses discrete symbols.

[2] Some linguists claim that schwa is a defective vowel. For instance, Van Oostendorp (1995) considers it a vowel with no feature specifications and Booij (1995) considers it an empty vowel. Zonneveld (1993) even claims that schwa is not present in the underlying representation but inserted later.

important for the account of Dutch spelling. Firstly, /aj/, /uj/, /oj/, /iw/ and /ew/, written as *aai*, *oei*, *ooi*, *ieuw* and *eeuw*, are sometimes considered diphthongs, see for instance Te Winkel (1860:7), Vieregge (1985:71), [*Woordenlijst 1995*], p. 14. However, they do not behave as diphthongs, which can be followed by any consonant, but as vowel-consonant combinations that have a restricted distribution. The sequence */ajk/, for instance, does not occur in Dutch, see Booij (1995:19). I will therefore consider them to be combinations of vowels and glides.

Secondly, [ŋ] is considered one sound despite the fact that it is often seen as being derived from a cluster /nɣ/ or /ng/ and changed into one sound by a rule, see for instance Kager (1989:208).[3] However, Trommelen (1983b:177) convincingly showed that the synchronic postulation of a cluster is unwarranted.

Finally, the phoneme inventory contains both voiced and voiceless fricatives even though the voiced-voiceless distinction, which has little functional load, is disappearing, see Gussenhoven (1983). However, phonologically the distinction still exists as illustrated by the minimal pairs in (2):

(2)   **Contrast between voiced and voiceless fricatives**

[**v**e]	*vee*	'cattle'	[**f**e]	*fee*	'fairy'
[ɛl**v**ə]	*elven*	'eleven (pl.)'	[ɛl**f**ə]	*elfen*	'elfs'
[**z**ɛnt]	*zend*	'to send'	[**s**ɛnt]	*cent*	id.
[kɪ**z**ə]	*kiezen*	'molars'	[ki**s**ə]	*kiese*	'delicate'
[**ɣ**lor]	*gloor*	'to dawn'	[**x**lor]	*chloor*	'chlorine'
[lo**ɣ**ə]	*logen*	'steep in lye'	[lo**x**ən]	*loochen*	'to denie'

Survey (2) contains pairs such as *kiese-kiezen* which show that the distinction between voiced and voiceless fricatives does not depend on the length of the preceding vowel as suggested in Wester (1987). The distinction of voiced and voiceless fricatives is thus not predictable, see also 3.4.1 and Appendix G. [4]

In addition to the thirty-five native phonemes, there are some phonemes that only occur in foreign words (/ɛː/ occurs in one native word, the onomatopoeia *blèren*-[blɛːrə] 'to bawl'):

---

[3] Alternations such as *koning-koninklijk* and the absence of /ŋ/ at the beginning of morphemes and after long vowels are often seen as evidence for this analysis. The spelling of this sound as two letters also seems to support the postulation of a cluster. However, these facts may also be accounted for by the fact that /ŋ/ is historically derived from a cluster, see for instance Te Winkel (1884:211), Van Bree (1987:163).

[4] The voicedness contrast is also supported by its role for the choice of verbal allomorphs, e.g. in *stoffen-stofte* versus *grieven-griefde*, see Kager (1989:211). With velar fricatives, the contrast is also relevant to the choice of diminutive allomorphs *lach-lach*je versus *vlag-vlag*g*etje*, see Wester (1987:65). Only [ɣ] occurs at the beginning of native morphemes, except for the word *chijl* which is of non-native origin, see Te Winkel (1860:20).

(3)   **Non-native phonemes**

iː	anal**y**se	yː	centrif**u**ge
uː	r**ou**ge	εː	s**e**rre
∅ː	fr**eu**le	ɔː	r**o**ze
ɑː	p**a**ss	ɛ̃	mannequ**in**
œ̃	parf**um**	ɔ̃	plaf**ond**
ɡ	**g**oal		

Following Booij (1995:7), the phonemes [ʒ] (*journaal*), [dʒ] (*jeep*) and /tʃ/ (*chip*) are analysed as derived from the underlying phoneme combinations /zj/, /dzj/ and /tsj/. For this reason these sounds are not listed under (3).

2.2.2 Dutch graphemes

Dutch spelling uses the 26 letters of the Latin alphabet supplemented with the ligature *ij*. Although originally the combination of *i* and *j*, *ij* now behaves as a single letter. Contrary to letter combinations it is capitalized as a whole: *IJssel/*Ijssel* versus *Aalst/*AAlst* see Te Winkel (1884:52–53).

(4)   **Dutch letters**
        a b c d e f g h i j k l m n o p q r s t u v w x y z ij

The letters *c* (except in *ch*), *q*, *x*, and *y* are only used in loan words and names, e.g. *Cremer*, *Quaadvlied*, *Bakx*, *Kuyper*. The letters in (4) are not sufficient to provide a unique representation for all 35 native phonemes in (1). The problem is most urgent in the case of vowels. Only six vowel letters (*a, e, i, o, u, ij*) are available to represent sixteen vowels. To overcome this lack of letters, Dutch uses the following fixed letter combinations:

(5)   **Fixed letter combinations**

geminates	aa, ee, oo, uu
vowel combinations	ie, oe, ei, ui, au, ou, eu
consonant combinations	ch, ng

Occasionally vowel letters are combined with *h* to denote a word-final short vowel: *bah, puh, goh, joh*. In a few cases, Dutch spelling uses diacritics. In *hé* and *oké*, the accent indicates a long vowel; in *hè* a short vowel ([ɛ]), and in *blèren* it indicates [ɛː]. Accents are also occasionally used to indicate word stress (often to prevent an incorrect pronunciation): *géren* ([ɣerə] 'to slant') versus *gerén* ([ɣərɛn], 'running'), *een* ([ən], 'a') versus *één* ([en], 'one'), ([vorꟷkomə], 'to prevent') versus *vóórkomen* ([vorꟷkomə] 'to exist'). Accents may also indicate

contrasts, e.g. *déze of díe man?* ('*thís* or *thát* one?'.

Since letter combinations are formed by letters which also occur independently some letter sequences are ambiguous: *zingen* ([zɪŋə] versus *ingang* [ɪŋɣɑŋ]). In case of ambiguous vowel sequences diacritics are sometimes used to prevent an incorrect segmentation, for instance, [ɣə-ɪnt] is written as *geïnd* to prevent an incorrect reading *[ɣɛint]. The rules that govern the distribution of these diacritics will be discussed in 5.5.1.

The properties of letter combinations cannot be predicted on the basis of the two letters, as observed for English by Venezky (1970). For instance, native words contain the combination <ch>, but <c> alone only occurs in loan words. Letter combinations behave like indivisible entities for instance with respect to hyphenation. The examples in (6) show that <ch> and <oe> may not be broken up by hyphens, unlike <o e> and <k s>:

(6)   [zuvə]   zoeven/*zo-even   ~   [zoevə]   zo-even
       [lɑxə]   la-chen/*lac-hen   ~   [ɔksəl]   ok-sel

The examples in (5) and (6) show that the orthographical counterpart of the phoneme may be either a letter or a fixed letter combination, see Te Winkel (1860:1–2). Such entities will be denoted as graphemes here. To be more precise, I will define graphemes as follows (for an alternative account by Wester (1985b), see 2.3.2):

(7)   **Definition of the grapheme**[5]
       a grapheme is the smallest orthographical unit, a letter or combination of
       letters, which behaves as an entity with respect to spelling rules

Since definition (7) does not define graphemes as letters or letter combinations that encode a phoneme, it excludes letter combinations such as <ng> and <bb> that do not form inseparable entities and may be hyphenated (*zin-gen*, *heb-ben*).[6] On the other hand, the fact that a given letter sequence is not hyphenated does not automatically imply that the letters form a grapheme, since there may be an alternative explanation. For instance, the fact that hyphenation *aa-i*, *oo-i* and *oe-i* is impossible can be accounted for by the fact that vowel letters which

---

[5] The term 'grapheme' is used in different ways in the literature. It is not used here to generalize over different realizations of the same letters (written, typed etc.), see for instance Venezky (1970:49–50), but as the orthographic counterpart of the phoneme, see for instance Haas (170:26), Carney (1994:xxvi).

[6] However, capitalization affects letters rather than graphemes: *Chris/*CHris, Aalst/*AAlst, Europa/*EUropa* (again, <ij> behaves as a single letter). Abbreviations sometimes use the initial letter of a word rather than the initial grapheme: *chronische aspecifieke respiratoire aandoeningen* is abbreviated as *cara*, not as **chara*).

correspond to consonants cannot be syllable-initial, see chapter 5, so the sequences *aai*, *ooi*, *oei*, *ieu*, *eeu* are considered combinations of graphemes. Consequently, Dutch has only one trigraph in native words, which occurs in names and the adjective *Bossche* from the name *Den Bosch* (<sch> used to be one of the spellings of /s/ before 1946).

Definition (7) thus yields the following set of native graphemes (in loan words there are other complex graphemes, e.g. *pathos*, *equivalent*, *team* and diacritics, e.g. *röntgen*, *crêpe*, *volière*, *logé*, *reçu*. These will not be discussed until chapter 4):

(8)   **Native graphemes**

         monographs:
         a e i o u ij; b d f g h j k l m n p r s t v w z
         digraphs:
         aa ee oo uu ie oe ei ui au ou eu; ch

According to Wester (1985b), the composition of digraphs is clearly systematic, since it is always the right-hand letter that modifies the left-hand one. In <ie> the <e> 'lengthens' the preceding <i>.[7] In <eu> the <u> rounds and lengthens the <e>, and in <oe> the <e> lengthens and raises the <o>. In <ch>, the only consonantal digraph, <h> makes a fricative of the stop <c>.

2.2.3 Sound-letter correspondences

In a strictly alphabetic system, every sound would correspond to exactly one letter and vice versa, see also Wester (1985a:205–206). However, this does not hold for Dutch. Even if sound-letter couplings in indigenous monomorphemic words only are considered, we still find sounds with multiple spellings or no corresponding letter at all ('0' in the following overview):

---

[7] In older stages of Dutch <e> was used to lengthen other vowel letters as well: <oe>, <ue>, <ae> represented /u/, /a/ and /y/. The same holds for <i> and <y>: <oi>, <ai>, <ui>, <uy> represented /o/, /a/, and /y/. Remnants of this spelling can be found in *heirbaan, oir, notoir*, and in names like *Bueren, Kraemer, Oirschot, Muyres* etc.

(9) Sound-letter couplings in native words

sounds	letters	examples
[p]	p, pp, b, 0	[trɑp]-*trap*, [kɑpər]-*kapper*, [wɛp]-*web*, [kɔmpt]-*komt*
[b]	b, bb, p	[bɑl]-*bal*, [ɣɑbər]-*gabber*, [ɔbduk]-*opdoek*
[t]	t, tt, d	[tɑk]-*tak*, [ɔtər]-*otter*, [hœyt]-*huid*
[d]	d, dd, t	[dun]-*doen*, [ɑdər]-*adder*, [œydbrɛŋ]-*uitbreng*
[k]	k, kk	[kɑl]-*kaal*, [ɑkər]-*akker*
[f]	f, ff	[fɛl]-*fel*, [ɔfər]-*offer*
[v]	v, f, w	[vel]-*veel*, [ɑvdun]-*afdoen*, [vret]-*wreed*
[s]	s, ss	[sɔk]-*sok*, [wɪsəl]-*wissel*
[z]	z, s	[zak]-*zaak*, [ɑzbɑk]-*asbak*
[x]	ch, g	[sxol]-*school*, [krɑx]-*kraag*
[ɣ]	g, gg, ch	[weɣə]-*wegen*, [wɛɣə]-*wegge*, [lɑɣbœy]-*lachbui*
[h]	h	[hɛlp]-*help*
[n]	n, nn	[nort]-*noord*, [lɪnə]-*linnen*
[m]	m, mm, n	[moj]-*mooi*, [ɛmər]-*emmer*, [ɪmpɛrk]-*inperk*
[ŋ]	ng, n	[ɑŋəl]-*angel*, [bɑŋk]-*bank*
[l]	l, ll	[lœyk]-*luik*, [ɑləs]-*alles*
[r]	r, rr	[ram]-*raam*, [bɔrəl]-*borrel*
[w]	w, uw, 0	[wer]-*weer*, [ew]-*eeuw*, [muwə]-*moeë*
[j]	j, i, 0	[jar]-*jaar*, [maj]-*maai*, [zejə]-*zeeën*
[ɪ]	i	[pɪt]-*pit*
[ɛ]	e	[pɛt]-*pet*
[ɔ]	o	[pɔl]-*pol*
[ʏ]	u	[pʏt]-*put*
[ɑ]	a	[jɑs]-*jas*
[i]	ie, i	[rit]-*riet*, [miɑu]-*miauw*
[y]	uu, u	[myr]-*muur*, [ny]-*nu*
[e]	ee, e	[ler]-*leer*, [etə]-*eten*
[ø]	eu	[løk]-*leuk*
[u]	oe	[rum]-*roem*
[o]	oo, o	[rot]-*rood*, [lopər]-*loper*
[a]	aa, a	[kas]-*kaas*, [adəm]-*adem*
[ɛi]	ij, ei	[wɛis]-*wijs*, [trɛin]-*trein*
[œy]	ui	[rœyk]-*ruik*
[ɑu]	ou(w), au(w)	[hɑut]-*hout*, [lɑu]-*louw*, [klɑutər]-*klauter*, [snɑu]-*snauw*
[ə]	e, i, ij, u, 0	[də]-*de*, [nodəx]-*nodig*, [olək]-*olijk*, [dɔkəm]-*Dokkum*, [ɑrəm]-*arm*

A striking divergence from the one-to-one correspondence in Dutch is the variation of the spelling of consonants (except for [w], [j], [h], [v] and [z]) and the vowels [a], [e], [o] and [y]. They can be written as geminates as well as single letters. Voiced sounds can also be represented by letters which correspond to unvoiced sounds and vice versa, [ŋ] and [m] are not only encoded by <ng> and <m> but also by <n>, and [v] is written as <w> in words such as *wreed*. The diphthongs [ɛi] and [ɑu] have more than one spelling as well. Glides can also be written in more than one way, cf. the spelling of [j] in *loei* and *joel* and of [w] in *eeuw* and *wee*. In *zeeën* ([zejə]) and *moeë* ([muwə]) the glides are omitted altogether.[8] Overview (9) does not express that some other sounds do not correspond to letters either, cf. for instance [lopə]-*lopen*, [sne]-*sneed*. Mismatches such as [ɣujə]-*goede/*goeie*, which are not illustrated in (9) either, are discussed in chapter 3.

However, overview (9) is needlessly complicated since it does not take into account two properties of Dutch orthography:

- Spelling abstracts from predictable aspects of the pronunciation as captured in Te Winkel's Principle of Received Pronunciation. For instance, <d> may encode [d] as well as [t], since spelling abstracts from the predictable devoicing at the end of syllables.
- It is often possible to predict the spelling of a given sound *in its context*. The choice between <i> or <j> as a code for [j], for instance, follows a pattern. Pairs such as *jaag-gaai* suggest that we write [j] as <i> at the end of a word and <j> at the beginning of a word. By referring to the context it is thus often possible to derive the spelling from the pronunciation and vice versa.

2.3 Literature on phoneme-grapheme correspondences

In this section, I will give an overview of the literature on the correspondence between phonemes and graphemes.

2.3.1 Prescriptive accounts

Traditional prescriptive accounts of the Dutch spelling system are Siegenbeek (1805a), Te Winkel (1863) and different editions of the *Woordenlijst van de Nederlandse taal* (literally: the word list of the Dutch language), henceforth

---

[8] The effect of Homorganic Glide Insertion is sometimes considered to be allophonic and thus irrelevant for orthography. Van Heuven & Hoos (1991) demonstrated that 'inserted glides' are systematically shorter and spectrally less extreme than lexical glides, which suggests that they do not have a phoneme status. However, inserted [w] and [j] can be written in certain contexts, cf. the <w> in *vrouwen*, and the <j> in 'progressively' spelled names such as *Tejo* (instead of *Theo*). Therefore, the fact that the effect of the rule is not represented in orthography is significant after all.

[*Woordenlijst 1866*], [*Woordenlijst 1954*], etc. The views of Siegenbeek and Te Winkel on the spelling system are essentially the same. Since Te Winkel is more explicit and elaborate and describes the spelling system which forms the basis for present day Dutch spelling, I will only summarize his views with respect to spelling variation for one sound (the differences between Siegenbeek's spelling and Te Winkel's spelling are described in Appendix H).[9]

Te Winkel (1863:8) captures the relation between spoken and written language in the Principle of Received Pronunciation:

> **Principle of Received Pronunciation** [Regel der Beschaafde Uitspraak]
> Represent by means of letters all the constituent parts that are heard in a word if it is pronounced correctly by civilized people. [… geef door letterteekens al de bestanddeelen op, die in een woord gehoord worden, wanneer het door beschaafde lieden zuiver uitgesproken wordt…]

This principle accounts for the fact that spelling abstracts from dialectical or idiolectical variation, such as differences in the quality of the final /r/.

Te Winkel acknowledges that this principle is too strict, since it does not allow for the predictable contextual variation of which most speakers are not aware, such as the quality of *aa* in *maal* and *maak*. He therefore relaxed it by allowing some deviation between spelling and pronunciation.[10] Spelling may deviate from the pronunciation as long as it will still enable us to derive the correct pronunciation, see Te Winkel (1884:27). This modification is known as the compatibility requirement ('verenigbaarheidseis'), see also Van Heuven (1978:45), but I will refer to this restriction with the term Readability Requirement (since compatibility may incorrectly suggest that the pronunciation

---

[9] Siegenbeek (1805a:17–38) uses a slightly different set of orthographic principles: "1. Write the way you speak [...] in writing, aim at the purest and most civilized pronunciation [Schrijf, zoo als gij spreekt [...] Rigt u in het schrijven naar de zuiverste en meest beschaafde uitspraak.] 2. "One should follow the closest and certain derivation" [Men volge in het schrijven de naaste en zekere afleiding.] 3. One should respect existing practice" [... men lette in de spelling op het algemeen erkend en aangenomen gebruik]. The first rule is the same as Te Winkel's Principle of Received Pronunciation; the second rule combines the Etymological and Morphological Principle. The third rule is the predecessor of Te Winkel's Principle of Common Practice, see Te Winkel (1865:25–26).

[10] See. Te Winkel (1863:9–10): "However, even in this way writing meets with almost insuperable problems with respect to the changing of most 'letter sounds', caused partly by the influence of neighbouring letters, partly by their position at the beginning, at the end or in the middle of words. However, if it is thus not possible to represent speech perfectly in writing, it is also unnecessary and would moreover be inefficient […] The aim of spelling is already attained if the reader can recognize the word." [Doch ook op dezen weg ontmoet het schrift bijna onoverkomelijke zwarigheden in de wijzigingen der meeste letterklanken, veroorzaakt deels door den invloed der naburige letters, deels door hunne plaats vooraan, achteraan of in het midden der woorden. Doch, is het niet mogelijk de spraak in het schrift volkomen juist weder te geven, het is ook onnodig en zou buitendien ondoelmatig zijn [...] Het doel van de spelling wordt reeds bereikt, wanneer de lezer het woord herkennen kan.]

of words must be compatible with their spelling). In combination with the Readability Requirement, the Principle of Received Pronunciation relates spelling to pronunciation and thus to phonology. Therefore this principle is called the 'Phonological Principle' by Cohen & Kraak (1972).

The Phonological Principle imposes a one-to-one mapping between phonemes and graphemes. However, the Phonological Principle is sometimes violated. For this reason Te Winkel formulated additional principles and rules.

The spelling of a word sometimes deviates from its pronunciation, as it does not represent the effect of a sound rule such as Final Devoicing. Although *strand* and *kant* are both pronounced with a voiceless final consonant in isolation, the underlying voicedness distinction, which surfaces in inflected forms, is still visible in orthography. Te Winkel (1863:12) accounted for these facts by supplementing the Phonological Principle with two more principles. The first one is the Principle of Uniformity:[11]

Principle of Uniformity [Regel der Gelijkvormigheid]
Give the same word and every constituent part the same shape, as far as the pronunciation allows this [Geef, zooveel de uitspraak toelaat, aan een zelfde woord en aan ieder deel, waaruit het bestaat, steeds denzelfde vorm…]

The Principle of Uniformity implies that spelling abstracts from the effect of sound rules such as Final Devoicing in order to provide a uniform representation for morphemes. If the effect of rules were represented, spelling would be less uniform: **strant-stranden*. Spelling thus represents the abstract sound representation that underlies both surface realizations. More examples of this phenomenon are given in chapter 3.

The Principle of Uniformity implies that orthography does not only abstract from the effect of allophonic rules, but also from the effect of sound rules that change one phoneme into another. Dutch spelling is thus not phonemic but morpho-phonemic. From this perspective, mismatches such as [wɛ**p**]-*web* are only apparent. If the spelling is compared with the abstract sound representation, i.e. the sound representation of the constituting morphemes of complex words, it becomes clear that there is a closer approximation to a one-to-one relation between phonemes and graphemes.

A second class of violations of the Phonological Principle consists of competing spellings such as *ij* or *ei* (*rijk ~ reik*) for /ɛi/. These spelling variants once corresponded to a sound contrast that has disappeared. To account for these facts, Te Winkel (1863:14) proposed the Principle of Etymology:

---

[11] Van Heuven (1980) calls this principle the 'Congruence Principle' and the Phonological Principle the 'Phonemic Principle'.

Principle of Etymology [Regel der Afleiding]
The choice between competing spellings for one sound is determined by the derivation or the older form that was used when pronunciations that are now identical could still be distinguished clearly. [Bij de keus der gelijkluidende letterteekens beslist de afleiding of de oudere vorm uit den tijd, toen de nu gelijk geworden klanken nog duidelijk onderscheiden konden worden.]

According to this principle, the choice between competing spelling variants is motivated by the history of the words. In some cases, however, similar sound contrasts have been neutralized in orthography. Therefore, this principle merely constitutes a justification in retrospect of existing practice, as was acknowledged by Te Winkel (1863:14): "We cannot question whether it would have been wiser to abolish one of the homophonous letter symbols. It suffices to know that this has not yet happened and we have to use the existing spelling. Abolishing it would be a radical change the effects of which are incalculable. It is wisest to respect existing practice."[12]

Other etymological spellings concern the spelling of schwa as <u>, <i> or <ij> in morphemes such as *Dokkum* (a minor generalization can be formulated for the choice between *-em* and *-um*, see Appendix F), *-ig* and *-lijk*, and the exceptional spelling of words such as in **th**uis, er**w**t, am**b**t in which the boldfaced letters are silent. I will exhaustively list such words in Appendix C. The 'older form' can also denote the original spelling of loan words in the language they were adopted from. The Principle of Etymology thus also accounts for the spelling of loan words, see chapter 4.

The last class of violations of the Phonological Principle that are not accounted for by the Principle of Uniformity, consists of facts such as *stations*straat instead of **stationstraat* and *hij word**t*** instead of *hij *word*. To account for such facts Te Winkel (1863:15) introduced the Principle of Analogy:

---

[12] "Het kan hier de vraag niet zijn, of wij niet verstandiger zouden gehandeld hebben met een der gelijkluidende letterteekens als overtollig weg te werpen; het is genoeg te weten, dat wij zulks niet gedaan hebben en nu met het bestaande moeten voortwerken. Het als ballast overboord te werpen, zou eene omwenteling zijn, wier nasleep niet te overzien is; het verstandigst is den bestaanden toestand te eerbiedigen."

Principle of Analogy [Regel der Analogie]
Words whose 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 that are apparently formed in the same way. […de woorden wier spelling noch door de uitspraak, noch door de gelijkvormigheid, noch door de afleiding wordt bepaald, worden op dezelfde wijze geschreven als andere, wier spelling met zekerheid bekend is en die oogenschijnlijk op overeenkomstige wijze gevormd zijn.]

This Principle is similar to the Principle of Uniformity, but it applies to affixes instead of free morphemes. It prescribes the spelling *stationsstraat* because of *stationsweg*, and *hij wordt* because of *hij loopt*, although two adjacent identical consonants are normally reduced to one by Degemination. Following Cohen and Kraak (1972), I will refer to the combination of the Principle of Uniformity and the Principle of Analogy as the Morphological Principle.

Te Winkel's principles thus suggest the following model of the relation between sounds and letters:

(10)   **Relation between sounds and spelling according to Phonological and Morphological Principles**
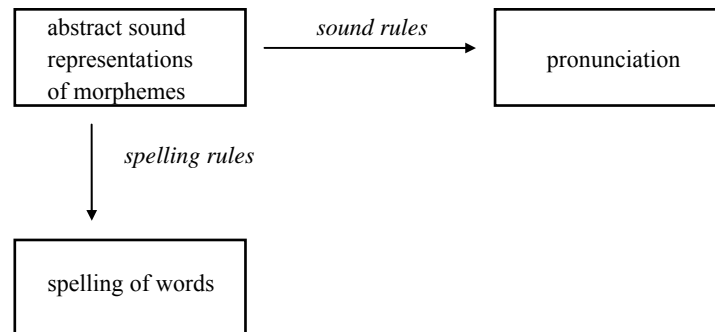
The four basic principles introduced by Te Winkel do not account for all deviations of the one-to-one correspondence illustrated in (10).[13] The remaining spelling variation can be predicted on the basis of context. Te Winkel discussed

---

[13] Te Winkel (1863:13) briefly considered introducing another principle that would increase the number of competing spelling variants, i.e. a requirement to orthographically distinguish all homophones (the Distinction Principle [Regel der Onderscheiding]), but eventually rejected it since it would give rise to arbitrary spelling distinctions. Te Winkel (1865:25–27) introduced two subsidiary spelling principles, the Principle of Melodiousness and the Principle of Common Practice. The first principle is an addition to the Principle of Received Pronunciation; the second one states that application of the spelling rules may not lead to major spelling changes.

two types of conditioned variation, namely the spelling of glides and the alternation of single letters and geminates.

First consider the spelling of glides. The phonemes /j/ and /w/ are not only written as <j> and <w> as expected, but also as <i> and <uw>: *loei* ~ *joel*; *eeuw* ~ *wee* (as mentioned above, the absence of <j> in *knieën* and of <w> in *moeë* follows from the Morphological Principle). Te Winkel motivated his decision to write /j/ as <i> rather than <j> in words such as *zaai* by the fact that it allows a uniform spelling of diphthongs (Te Winkel considered /aj/, /ɔj/ and /uj/ to be diphthongs). If all diphthongs were written with <j>, this would lead to problems in some cases: <ej> and <uj> would not correctly represent [ɛi] and [œy] before vowels, since *<ejeren> and *<ujer> suggest the pronunciation *[ejərə] and *[yjərə] rather than [ɛijərə] and *[œyjərə].[14] Therefore *zaaien* was preferred.

Te Winkel (1863:9) claims that consonantal variation such as <m> versus <mm> represents a sound contrast; geminate consonant letters correspond to geminate consonants. This implies that variation in the spelling of consonants is required by the Phonological Principle. However, he does not propose that variable spelling of vowels is motivated in a similar way and does not offer another explanation, see Te Winkel (1863:10). An exception is formed by the variation of <i> and <ie> for /i/ (*knie-knieën* ~ *neurie-neuriën*), which Te Winkel considers to be related to the stress pattern of the words, see chapter 4.

Te Winkel further discusses topics such as the choice between existing variants such as *schuifelen* or *schuiffelen*, *monniken* or *monnikken*, *wereld* or *waereld*, *zucht* or *zugt* etc. (in all these cases the first variant is chosen), the spelling of loan words (see chapter 4), and topics related to the spelling of compounds that are not discussed in this study. The distribution of <w> after <ou> and <au> (*koud-vrouw*) was not discussed either.

Te Winkel's view was elaborated by De Vries and Te Winkel in [*Woordenlijst 1866*], in which the rules were refined and supplemented by a word list. After the first edition, new versions of this spelling dictionary appeared in which the spelling of individual words and some spelling rules were altered because there were some minor reforms of the spelling of Dutch (see Appendix H). Despite these changes, Te Winkel's principles are still valid. The role of the Principle of Etymology became less prominent, but it was not given

---

[14] Te Winkel chose between the different variants that were used by writers in those days: *zaai-zaaijen, zaaij-zaaijen, zaay-zaayen* and *zaai-zaaien*. He rejected the Siegenbeek spelling *zaa-*zaaijen, since the hiatus filling letter is written in the inflected form (*zaaijen*) only, which violates the Morphological Principle. Furthermore, it would lead to *bij-bijjen*. The option *zaaij-zaaijen* was not chosen by Te Winkel since <j> was considered totally superfluous, even in inflected forms. Initially Te Winkel supposed that *zaaien* could be pronounced [zɑ-in] and would therefore require diaeresis (*zaaiën*). Therefore, Te Winkel (1863:26–28) prescribed *zaay-zaayen*. Te Winkel (1884:58–61) did no longer use <y> in native words or think that diaeresis was necessary, so the spelling *zaai-zaaien* was chosen. Te Winkel did not discuss the spelling of combinations with /w/, but the <w> was apparently not seen as totally superfluous.

up altogether. The variable spelling of /ɛi/ and /ɑu/, for instance, is still explained by the historical pronunciation in the most recent edition of this dictionary; see [*Woordenlijst 1995*], p. 21–22.

Discussion of Te Winkel's approach reveals that divergences of the one-to-one correspondence can be divided into three categories. The Morphological Principle causes spelling variants that are only apparent and disappear if we consider spelling to be a code for the pronunciation of morphemes. Real spelling variants can be subdivided into competing spelling variants described by the Principle of Etymology and conditioned, context-sensitive spelling variants for which no spelling principle was introduced. Examples of these three categories are listed in (11):

(11)   **Types of spelling variants**

Apparent variants

t, p, x	*t ~ d, p ~ b, ch ~ g*	*hond ~ bont, web ~ step, weg ~ toch*
s	*s ~ ss*	*stationsstraat ~ stationsweg*

Real variants
a   Conditioned variants

j	*j ~ i*	*joel ~ loei*
w	*w ~ uw*	*wee ~ eeuw*
ɑu	*ou ~ ouw*	*koud ~ vrouw*
	au ~ auw	*klauter ~ lauw*
a	*aa ~ a* etc.	*raam ~ ramen*[15]
i	*ie ~ i*	*neurie ~ neuriën*
r	*r ~ rr* etc.	*kar ~ karren*

b   Competing (etymological) variants

ɛi	*ij ~ ei*	*rijk ~ reik*
ɑu	*ou ~ au*	*louw ~ lauw*

It should be noted that Te Winkel's principles do not account for conditioned variants. Since the choice between the variants is predictable and regular, it seems likely that there is some other principle at work that Te Winkel either did not recognize or left implicit. In the next section, I will examine other descriptions of (part of the) Dutch spelling system to see whether they shed light on this principle.

---

[15] Some words seem to have a more constant spelling: *dak(en), bevel(en), god(en)* etc. In this case, however, the pronunciation is idiosyncratic: there is a vowel change in inflected forms. Other examples are the idiosyncratic plurals *stad-steden, schip-schepen* and *-heid-heden*. Sometimes vowels are also changed in the diminutive form where the vowel change causes spelling variation: *lot-lootje, blad-blaadje*.

2.3.2 Descriptive accounts

Non-prescriptive accounts of Dutch spelling fall into two categories: those that follow Te Winkel and claim that spelling is a code for the pronunciation as captured by Phonological Principle, and those that reject this assumption. I will now give a short overview of the general approach chosen by different authors and the implications for the way spelling variation is described by them.

Spelling as a code for the pronunciation
Examples of descriptions of the spelling system in which spelling is considered to be a code for the pronunciation are Cohen & Kraak (1972), Booij et al. (1979), Van Heuven (1980), Booij (1985, 1987, 1991, 1995), Wester (1985a-b, 1987, 1989). To illustrate this approach I will discuss the most elaborate proposals, namely those by Booij and Wester.

Booij
Booij has written about Dutch orthography in several publications: Booij et al. (1979), Booij (1985, 1987, 1991, 1995). To account for contrasts such as *kant-strand*, Booij claims that spelling aims at representing abstract representations of morphemes that underlie different realizations such as [strɑnt] and [strɑndə]. However, Booij points out that there are also words in which the effect of sound rules is visible e.g. in *lieve-lief/*liev* and *verbranden-verbrand/*verbrandd* (cf. *draaien-gedraaid*). Booij concludes that Dutch orthography is inconsistent in this respect.

With respect to conditioned spelling variants, Booij focuses on the alternations of single and geminate letters. Booij (1995) claims that the contrast between single and double consonant letters corresponds to a difference in prosodic structure. Geminate letters represent ambisyllabic consonants. To account for the absence of doubling after vowels representing the vowel schwa, in words such as *hannesen* ([hɑnəsə]) and *bezemen* ([bɛzəmə]), he postulates that schwa behaves as a long vowel in orthography. The variation of single and geminate vowel letters is described by context-sensitive spelling rules, which are sensitive to phonological syllables. Conditioned spelling variants are thus derived from sounds and prosodic structure.

However, Booij (1991) makes a different claim. He draws attention to the fact that there is no doubling in words such as *jongen* and *lachen* as expected on the basis of the phonological syllables ([jɔ-ŋə], [lɑ-xə]). He explains this by postulating that Consonant Doubling is not applied to phonological syllables, but to (autonomous) orthographical syllables. According to Booij, the orthographical syllable boundary is between the two letters: *lac.hen*, *jon.gen*, since *ng* and *ch* behave as two separate letters rather than as digraphs. This proposal leads to the correct hyphenation pattern in the case of *jongen*: *jon-gen*, but not in the case of

lachen: **lac-hen*, instead of *la-chen*. Booij considers the latter behaviour to be inconsistent.

Booij (1995:181) only mentions etymological variants such as *rijk ~ reik* in observations such as the following: "The lack of one-to-one correspondences is increased by the fact that the history of words plays a role. This does not only apply to loan-words that often keep their original spelling, but also to native words whose spelling may reflect an older stage of Dutch." Booij et al. (1979) argue that the Principle of Etymology may not always be motivated by the need to distinguish homophones such as *rijk* versus *reik*, and conclude that it merely protects existing practice.

Wester

Wester focuses on two topics: spelling variation caused by the Morphological Principle in Wester (1987) and conditioned spelling variants in Wester (1985a-b, 1989). Wester (1987) also claims that spelling represents abstract sound representations of morphemes, but where Booij eventually concludes that Dutch orthography does not do so consistently, Wester gives an alternative account for apparent inconsistencies such as *lief-lieve*. I will discuss Wester's account of the spelling of fricatives in 3.4.1 and Appendix G.

Wester (1985b) examines spelling phenomena such as Diaeresis Placement, Hyphenation and the alternation of single letters and geminates, and argues that these phenomena are best considered to be operations on the letter sequences that arise after sounds are converted to letters. According to Wester, more information is needed than is provided by mere letter sequences. For instance, to account for the fact that an intervocalic consonant is doubled in *menuetten* but not in *humeuren*, it is necessary to know that <eu> is a digraph, but <ue> a combination of two graphemes (Wester illustrates this phenomenon with loan words only, but the same observation can also be made on the basis of native pairs such as *geoffer-oefen*). For this purpose, Wester proposes that letters are associated to a CV-tier, a linear representation of abstract consonants (Cs) and vowels (Vs), by a set of extrinsically ordered rules. Cs and Vs are further organized into syllables (S). This results in the following structures for *menuetten* (/menyɛtə/) and *humeuren* (/hymØrə/):

```
(12)    S     S     S     S        S     S     S
        ∧     ∧     |    /|\       ∧    / \   /|\
        C V   CV    V    C V C      CV   C V   C V C
        | |   | |   |    | | |      | |  | ∧   | | |
        m e   n u   e    t e n      h u  m eu  r e n
```

According to Wester the syllables in (12) are phonological syllables (and Cs and Vs are phonemes, so letters form an entity (grapheme) when they represent one

sound). Spelling rules such as Consonant Doubling (which Wester calls Consonant Insertion) apply to structures such as (12). For instance, a consonant is doubled at the end of a syllable after a vowel letter that represents a short vowel, and vowel geminates are degeminated at the end of a syllable:

(13) **Wester's rules for the alternation of single letters and geminates**
   a   Consonant Doubling

$$0 \rightarrow C_i / \quad V \ \_\ ]_\sigma \ [_\sigma C_i$$

   b   Vowel Degemination

$$V_i \rightarrow 0 / V_i \ \_\ ]_\sigma$$

In Wester's view, spelling is an extra tier in the sound representation of words, which explains why spelling rules may refer to aspects of the pronunciation (e.g. syllable structure) as well as to spelling.[16] Wester's view may be represented as follows (to distinguish the two types of spelling rules used by Wester, I will refer to the rules that translate sounds into letters as phoneme-to-grapheme conversion rules):

(14) **Spelling model of Wester (1985b)**

```
┌─────────────────────────────────────┐
│        sound representations         │
└─────────────────────────────────────┘
                   │
                   │   phoneme-to-grapheme conversion rules
                   ▼
┌─────────────────────────────────────┐
│  combined sound-spelling representation  │
└─────────────────────────────────────┘
                   │
                   │   spelling rules
                   ▼
┌─────────────────────────────────────┐
│               spelling               │
└─────────────────────────────────────┘
```
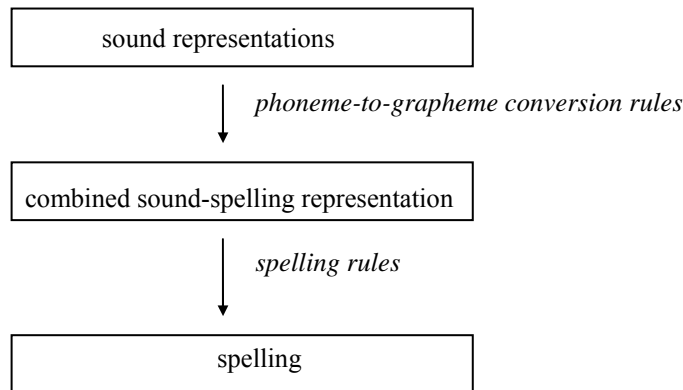
In contrast with Booij (1995), Wester considers spelling derivation to be a two-step process. First a combined sound-spelling representation is built, and subsequently letter sequences are modified. Both steps in the derivation are conditioned by the pronunciation, but the second rule type may refer to letter

---

[16] It is not very clear how abstract the sound representation in (12) is. However, representations such as (12) are not created for morphemes but for words, which implies that it is not a very abstract level. This is in contrast with Wester (1985b), which presents an account for the spelling of fricatives for which it is crucial that spelling encodes an abstract level of sound representation.

sequences as well.

Spelling as an autonomous rule system
There are two descriptions of the Dutch spelling system that are not based on the assumption that the spelling of a word is derived from its pronunciation, namely Zonneveld (1980) and Kerstens (1981). These authors assume that the traditional approach in which spelling is related to sounds is motivated by an implicit assumption that spelling cannot be described as an autonomous rule system. Their articles argue against this assumption and claim that spelling is not a code for phonemes but for morphemes. The spelling system consists of rules which convert abstract spelling representations to their surface realizations:

(15)   **Autonomous Spelling (Zonneveld, Kerstens)**

```
┌─────────────────────────────────────────────┐
│ abstract spelling of morphemes               │
│ (no relation with sound representations)      │
└─────────────────────────────────────────────┘

              │  grapheme-to-grapheme rules
              ▼

┌─────────────────────────────────────────────┐
│                                               │
│        spelling of words                      │
│                                               │
└─────────────────────────────────────────────┘
```
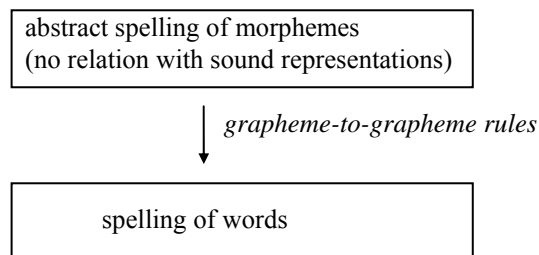
First consider Zonneveld's account.

Zonneveld
For an autonomous description, spelling variation is simply not relevant unless it concerns variation of the spelling of related words. Zonneveld therefore does not discuss contrasts such as *hond ~ bont* and *rijk ~ reik*. To account for alternations of written morphemes, he formulates a set of extrinsically ordered spelling rules (and postulates abstract spelling representations). For instance, the following two rules govern the distribution of single letters and geminates:

(16)   **Zonneveld's rules for the alternation of single letters and geminates**
        a   Vowel Doubling
            $0 \rightarrow V_i / C\ V_i\ \_\ C\ \#$
        b   Consonant Degemination
            $C_i / \rightarrow 0 / \_\ C_i\ \#$

The first rule says that a vowel letter following a consonant is doubled before a word final consonant letter. The second rule states that consonant geminates are degeminated word-finally. Applied to the underlying spelling representations

proposed by Zonneveld, *ram*, *ram+en*, *tamm* and *tamm+e*, these rules correctly derive the spelling *raam*, *ramen*, *tam* and *tamme*.[17] Note that Zonneveld's approach to the alternation of single and geminate letters is the reverse of that of Wester. For a comparison of both approaches see Nunn (1995), where it is argued that Wester's account is to be preferred, since it can easily account for the contrast between *hard* and *haard* (one or two underlying vowel letters) whereas Zonneveld's proposal implies that both words have one underlying vowel letter and are both realized as *haard*.

Zonneveld's analysis implies that certain sound alternations are duplicated in spelling, since they can no longer be explained by referring to the pronunciation. For instance, Zonneveld formulates orthographical counterparts to the alternation of <d> and <t> in past tense and past participle endings: *spieken-spiekte* versus *draaien-draaide*. To be able to describe such phenomena, Zonneveld postulates that spelling rules refer to phonological features such as [voice]. For instance, the contrast between *spieken-spiekte* versus *draaien-draaide* is accounted for by a rule that assimilates the voicedness of the initial consonant of the past tense or past participle ending to the final phoneme of the stem. Similarly, the contrast between *lief* and *lieve* is accounted for by 'spelling devoicing' that changes the features [+voice] into [−voice] for letters with the features [+cont, −back]. Since the corresponding pronunciation is not considered, the contrast between *lief-lieve* (alternation) and *strand-stranden* (constant spelling) does not form a problem as it does for Booij and Wester (but we will see in 3.4.3 that alternations such as *spiekte* versus *draaide* cannot be described adequately by autonomous spelling rules). Zonneveld acknowledges that the use of phonological features is an indication that there is some relationship between sounds and letters and postulates a diachronic relation rather than a synchronic relation; spoken and written language are two dialects of Dutch.[18]

Kerstens

Kerstens (1981) also proposes an autonomous spelling analysis. In contrast with Zonneveld, Kerstens does not discuss alternations such as *spieken-spiekte* and does not propose that spelling rules refer to phonological features. To account for the alternation of single letters and geminates, Kerstens proposes abstract

---

[17] Zonneveld (1980:527) assumes that there is a difference in the use of accents in geminates and other digraphs: *vóór* (one accent) versus *vóer* (two accents), and that this supports vowel doubling: the accent is copied with the letter. However this is not supported by the rules in [*Woordenlijst 1995*], p. 36.: "The stress marker is the symbol ` When a sound is encoded by more than one letter, the first two get an accent." [Het klemtoonteken is het teken ` Als een klank met meer dan één letter wordt weergegeven, krijgen de eerste twee letters een accentteken.].

[18] This is in line with the remarks of Van Haeringen (1962) on the differences between spoken and written language, see 3.4.2.

representations that do not only contain letters, but also '-', which represents orthographical syllable boundaries, and subsidiary letters. In these abstract representations geminates are represented by a letter and a subsidiary capital letter. Vowels representing schwa are denoted by subsidiary letters as well. For instance words such as *laat*, *later*, *watten*, *wie*, *de*, *kom* and *nu* are represented underlyingly as *laAt*, *laA-tEr wat-TEn*, *wiE*, *dE*, *komM* and *nuU*. Subsidiary vowel letters are spelled out as an identical letter within the same orthographical syllable (*laat*, *later*), as *e* after *i* or a consonant (*wie*, *de*) and deleted elsewhere (*nu*). The subsidiary consonant letter is spelled out as an identical letter after a heterosyllabic identical consonant (*watten*) and deleted elsewhere (*kom*). Together with rules that spell out the symbols, these abstract representations are sufficient to account for the alternation of single letters and geminates.

Summarizing, the literature offers different approaches to the spelling system of Dutch. The traditional view, first formulated by Te Winkel and adopted and elaborated by Booij and Wester, is that spelling encodes sounds. This view is worked out in two different ways. The account of Booij (1995) comes closest to that of Te Winkel, since he only uses one type of spelling rules, which convert sounds into letters. Wester (1985b) and Booij (1991), on the other hand, also use a second type of spelling rule that applies after phoneme-to-grapheme conversion and affects letter sequences rather than sounds. Consequently, they distinguish two levels of orthographic representation. The difference between these proposals is that Wester considers the second rule type to be influenced by the pronunciation, more specifically, phonological syllables, whereas Booij (1991) considers it to be autonomous. An alternative view, proposed by Zonneveld (1980) and Kerstens (1981), is that the spelling is not (synchronically) related to the pronunciation. In this view, the spelling of morphemes is listed in the lexicon. Spelling variation is only relevant if it concerns different occurrences of the same morpheme. Kerstens and Zonneveld also distinguish two orthographic levels.

2.4 A new model for the relation of sounds and spelling

On the basis of Te Winkel's spelling principles, the simple spelling model (10) was taken as a starting point. However, examination of the literature suggests that this model has to be extended.

Zonneveld (1980), Kerstens (1981) and Wester (1985b) argued that alternations of single and geminate letters can be better described as the result of operations on letter sequences instead of the result of phoneme-grapheme conversion rules. This approach has the advantage that it allows a more insightful account of spelling alternations. Vowels of which the spelling alternates ([a, e, ɔ, y]) and vowels of which the spelling is constant ([i, u, ∅, ɛi,

œy, ɑu]) form arbitrary subsets of sounds, but the corresponding graphemes fall into two well-defined sets: geminates are simplified, non-geminates are not (the alternation of <ie> and <i> is a separate phenomenon, which will be discussed in 5.6.1). Similarly, on the basis of the pronunciation it is not clear why /x/ and /ŋ/ are never written as geminates, but the fact that they are written as two graphemes (*ng*) or a digraph (*ch*) immediately explains this behaviour. I will not follow Zonneveld and Kerstens by claiming that spelling is autonomous, since there are too much aspects of spelling that are predictable on the basis of the pronunciation. I will therefore supplement model (10) with grapheme-to-grapheme conversion rules. This is also the approach chosen by Wester.

Wester considers adopting Zonneveld's autonomous spelling rules for this purpose, but eventually rejects this option. Zonneveld's rules refer to unstructured letters only and therefore are unnecessarily complex and sometimes incorrect. For instance, on the basis of pairs such as *schoenen-schoen*, Zonneveld would incorrectly predict *beogen-*beog* (instead of *beoog*). Wester convincingly shows that such incorrect predictions can be avoided by referring to phonological syllables and concludes that grapheme-to-grapheme conversion is not possible on the basis of letters only.

However, Wester's proposal runs into problems with the data discussed in Booij (1991): both *ng* and *ch* encode single phonemes, so we would expect *zi-ngen* as well as *la-chen*. For this reason Booij suggested that there are orthographical syllables next to phonological ones, and that spelling rules are conditioned by these orthographical syllables. Some other facts that point in this direction are the following:

(17)   **pronunciation**   **predicted spelling**   **actual spelling**
     a   ba-jərt   *   b**ai**erd   b**aa**ierd
     b   ba-jəs   b**a**jes   b**a**jes
     c   sno-dart   *   sn**o**daard   sn**oo**daard
     d   bo-ɣart   b**o**gaard   b**o**gaard
     e   lɑ-fart   *   la**ff**aard   la**f**aard
     f   drɔ-sart   dro**ss**aard   dro**ss**aard

The long vowels in (17a-d) all occur in open syllables and thus satisfy the conditions of Vowel Degemination (rule (13b) above), so we expect them to be written as single vowel letters. This is the case in (17ac) but not in (17bd). The contrast between (17e) and (17f) is not expected either, since both words satisfy the conditions set by Consonant Doubling (see rule (13a) above). The only difference between pairs such as *baaierd* and *bajes* is the spelling of the sound /j/. In the first word /j/ is written as <i>, while it is written as <j> in the second word. The choice between <i> and <j> correlates with the choice between <aa> and <a>, which suggests that Vowel Degemination is sensitive to orthographical

syllable structure. The contrast between *bogaard*, *drossaard* on the one hand and *snoodaard*, *lafaard* on the other also suggests that orthographical syllable structure of the latter deviates from the phonological syllable structure. We will see in chapter 5 that this is also suggested by the difference in hyphenation patterns: *bo-gaard*, *dros-saard* on the one hand and *snood-aard*, *laf-aard* on the other. Such a difference also exists for pairs such as *maai-en* versus *ba-jes*. The facts in (17) may not be adequately described by rules that only formulate conditions on the phonological context such as those of Wester.

I conclude that Wester is right in claiming that rules that modify letter sequences do not work well when applied to bare letter sequences, and that their performance can be improved by organizing letters into graphemes and graphemes into syllables. However, I will postulate that these syllables are orthographical syllables, which may deviate from their phonological counterparts, and that graphemes are defined on the basis of their spelling behaviour.

A hybrid model

The model proposed here consists of the model in (10) supplemented with a second set of rules comparable to the rules used in Autonomous Spelling, cf. model (15). To set them apart from phoneme-to-grapheme conversion rules, the second type of spelling rules will be denoted as autonomous spelling rules.[19] This term indicates that these rules do not directly refer to the pronunciation of words, in contrast to the second type of rules of Wester. Autonomous spelling rules only manipulate letter sequences.[20] The new model is given under (18):

---

[19] Different terms have been used for autonomous spelling rules. Zonneveld (1980) only has one rule type which is simply called spelling rule. Kerstens (1981) uses different terms: 'adaptation rule', 'formation rules', 'subcategorization rules', 'deletions', 'spelling out rules' and a diaeresis convention. In his description of English, Carney (1994:67) uses 'adaptation rules' (rules which adapt the spelling of a morpheme to the structure of complex words, e.g. (*eight+th → eighth/*eightth*) and 'graphotactic rules' or 'letter distribution rules' (rules which restrict possible letter sequences). The term 'autonomous spelling rule' as used here refers to all letter-based processes.

[20] 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 et al. (1986) also use these two rule types as well as letter-to-letter rules which insert morphological boundaries.

(18)   **Hybrid model**

```
 ┌────────────────────┐      sound rules       ┌──────────────────┐
 │  abstract sounds   │ ─────────────────────▶ │  pronunciation   │
 └────────────────────┘                        └──────────────────┘
         │
         │   phoneme-to-grapheme conversion rules
         ▼
 ┌────────────────────┐
 │  abstract spelling │
 └────────────────────┘
         │
         │   autonomous spelling rules
         ▼
 ┌────────────────────┐
 │      spelling      │
 └────────────────────┘
```
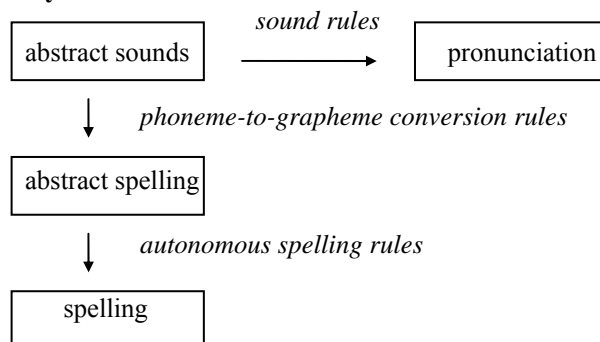
In chapter 5, I will argue that autonomous spelling rules apply to fulfil graphotactic requirements or prevent ambiguity and I will introduce a new spelling principle to account for these facts.

The assumption of the second rule type enables us to account for contrasts between words that are written differently even though they may be pronounced in a similar way. The absence of Consonant Doubling and Vowel Degemination in *lafaard* and *snoodaard* can be accounted for by the fact that *-aard* is treated as a compound member in spelling (cf. also the hyphenation pattern *laf-aard*, *snood-aard*), even though this is obviously not the case in the pronunciation. The contrasting words do not contain the suffix *-aard: drossaard*, *tabbaard* and *bogaard* are not derived from *dros*, *tab*, or *boog*, and are therefore treated as monomorphemic words.

Similarly, the different spelling of /a/ in *bajes* and *baaierd* is not arbitrary since the consonant immediately following the relevant vowel is written in different ways. The difference between <i> and <j> leads to a different orthographical syllable structure: <aa> is syllable final in *ba.jes* but not in *maai.en*, so only the first words satisfies the structural description of Vowel Degemination. I will discuss the computation of orthographic syllables in chapter 5. For the moment it suffices to know that the mismatches in (17) can be accounted for by Wester's rules (in a slightly modified version, see 5.2, Appendix E) if these rules are not governed by phonological syllables, but by their orthographical counterparts.

CV-structure
As observed by Wester (1985b), letter sequences alone do not provide enough information to account for all aspects of spelling behaviour. Consider for instance the following examples:

(19) **doubling**   **no doubling**
      a   bi**tt**er   [bɪtər]   mij**t**er   [mɛitər]
      b   be**ll**en   [bɛlən]   twijfe**l**en   [twɛifələ]

In these examples a single consonant letter follows a single vowel letter (*ij* is one letter, see 2.2.2), but in some cases the expected Consonant Doubling stays out. To be able to correctly apply Vowel Degemination and Consonant Doubling in such cases, I will supplement spelling representations with hierarchical structures.[21] The fact that <ij> patterns with digraphs such as <oe> in *foeter* and not with other single vowels such as <i> in *bitter* (**mijtter*) suggests that the factor that decides whether or not consonants are doubled is not the number of letters, but by the number of V-positions associated to a vowel letter. This suggests that short and long vowels are represented as follows in orthography:

(20) **Orthographic representation of short and long vowels**

short vowels	long vowels			
V	VV	VV	VV	CV-structure
|	V	||	||	
a	ij	oe	aa	spelling

Although the use of structure seems to eliminate the need for an underlying distinction between single letters and geminates such as *a* versus *aa*, I will not propose that they are both represented by a single letter for the following reasons. Firstly, a new deviation of the one-to-one correspondence between phonemes and graphemes is created if long and short vowels are both converted to single vowel letters. Secondly, this approach would involve reformulating Vowel Degemination as Vowel Doubling, which implies that we can no longer account for the fact that only geminates alternate with single letters.

    Now let us consider how we may account for the fact that a consonant is not doubled after a vowel letter that represents schwa as illustrated in (19b). For this purpose we need to be able to distinguish an <e> that encodes schwa from an <e> that encodes an [ɛ]. CV-structure does not immediately help. Schwa is phonetically a short vowel, so we expect it to be converted to a single vowel letter with one V-position, in which case we still expect doubling. I will therefore account for the contrast in (19b) by postulating that schwa has no V-position:[22]

---

[21] Another possibility would be to follow Kerstens (1981) and use subsidiary segments to encode additional information in spelling. However, in Kerstens it is crucial that orthographic syllable boundaries are given, whereas in the approach chosen here, they are computed on the basis of CV-structure. In this respect Kerstens's approach misses some generalizations.

[22] Another option would be to postulate an empty V-position, i.e. not to assign it a letter but this

(21)   **Orthographic representation of /e/, /ɛ/ and schwa**

```
VV      V                    structure
||      |
ee      e        e           spelling
```

Since Consonant Doubling applies after vowels with a single V-position, it is not applicable after vowels that represent schwa if this vowel has no V-position. I propose that <e> without a V-position be incorporated in syllable structure at the end of the derivation. Words such as *kennissen* and *salarissen* in which the relevant vowels are sometimes pronounced as schwa now seem problematic. These facts will be discussed in Appendix F, where it will be argued that <i> denotes full vowels, so the doubling is regular.

So far I have discussed what spelling representations should look like in order to account for spelling alternations. We also have to make sure that we can account for the fact that spelling alternations are often predictable on the basis of the pronunciation. Since the hyphenation of words such as *vin-ger* shows that the number of skeletal positions in spelling and pronunciation is not always isomorphic, I propose that CV-positions are assigned on the basis of the sounds that they represent, but that phonological and orthographical structure need not be identical. The relation between phonological and orthographical structure is analogous to the relation of morphological and prosodic structure, see Booij (1995:52). Orthographical structure is constructed on the basis of the pronunciation during phoneme-to-grapheme conversion. Short vowels are converted to letters with one V-position, long vowels to letters with two V-positions, and schwa is converted to a letter without a V-position:

(22)
```
            VV                  V              V
            ||                  |              |
      /a/ → aa          /ɑ/ → a         /p/ → p          /ə/ → e
```

The use of orthographic CV-structure based on the pronunciation accounts for the fact that spelling needs more phonological information than can be encoded by letters only, without stating that all phonological information has to be available. Thus, certain aspects of spelling behaviour can be predicted on the basis of the pronunciation: short vowels and consonants are assigned a single V-position or C-position, and long vowels two V-positions. However, since

---

approach seems inappropriate because of different spellings for schwa in words such as *Dokkum ~ stiekem*. Such spelling differences are generally accounted for by phoneme-to-grapheme conversion rules.

spelling and pronunciation are related by a mapping procedure, we can also account for mismatches, such as the fact that the letters that encode the sound /n/ do not form a grapheme.

Summarizing, for an insightful account of the alternation of single letters and geminates, the relevant rules should apply to letter sequences rather than to sounds. That way cases in which spelling and pronunciation are not isomorphic may be accounted for. The rules should not be governed by phonological structure as proposed by Wester, but by letters supplemented with extra information in the form of an orthographical CV-structure. Although this model implies that the computation of spelling proceeds in two steps, which seems more complicated than some previous models, this is compensated for by the fact that we have a more accurate and insightful description of spelling alternations.

2.5 Conditioned and competing spelling variants

In this section, I will discuss remaining conditioned spelling variants and competing spelling variants, but not the sound-letter mismatches caused by the fact that spelling represents an abstract level of phonological representation. Sound-letter mismatches such as *opdoek*-[ɔbduk] and *bank*-[bɑŋk] will be discussed in chapter 3.

Conditioned spelling variants
Apart from the alternation of single letters and geminates that were already discussed in the previous section, there are two more types of conditioned spelling variant for which rules have to be formulated: the distribution of <w> after <au> and <ou>, and the spelling of glides.

First consider the distribution of <w> after *ou* and *au*:

(23)   **The distribution of <w> after <au>**

/ɑu/	ou or ouw	koud versus vrouw
	au or auw	klauter versus lauw

The pattern in (23) is accounted for in [*Woordenlijst 1954*], p. XLI by the following statement: 'Finally and before vowels *w* is added to *au* and *ou*, except for final *ou* which is derived from *oude* [...] *kou, hou, wou, nou, jou.*'[23] This rule has to be made more specific: does 'finally' mean word-finally or morpheme finally? This can be decided on the basis of derived forms such as

---

[23] "Aan het eind en vóór klinkers wordt [aan <au> en <ou>] *w* toegevoegd [..] behalve, wat finale *ou* betreft, in gevallen waarin de tweeklank uit *oude* is ontstaan [...] *kou, hou, wou, nou, jou.*"

vertrouwd. This word shows that the context outside the morpheme is ignored (otherwise we would write **vertroud*). I conclude that 'finally' means morpheme-finally:

(24)   **Rules for the distribution of <w>**
    a   /ɑu/   → <auw> or <ouw> / _ {V, +}
    b   /ɑu/   → <au> or <ou> elsewhere

Comparison of forms without and with <w> suggests that <w> is the representation of the glide [w] that is inserted to solve hiatus:

(25)   **with w**   **without w**
    [kɑud]   koud   ou**w**el   [ɑuwəl]
    [sɑus]   saus   lau**w**er   [lɑuwər]

This assumption is supported by the fact that <w> is only written in words that are pronounced with a glide (e.g. *ouwel*), or have related forms that are pronounced with a glide (e.g. *vrouw-vrouwen*). However, the <w> is not only written in inflected forms where [w] is actually inserted, but in the corresponding stems as well. This suggests that <w> has become part of the spelling of morphemes such as *vrouw* although it does not correspond to a sound (even in the surface realization). Possibly <w> is added to the stem to satisfy the Morphological Principle: if <w> only appeared in inflected forms, the spelling of morphemes would not be constant. From this perspective, the absence of <w> in words such as *nou* that has no inflected forms, or words such as *hou* where /d/ or /də/ is deleted is as expected: insertion of <w> is not necessary to avoid differences between stem and derived forms. Note that the spelling <auw> or <ouw> is compatible with the pronunciation, since /w/ does not occur after sounds other than /i/, /e/ or /y/, see Booij (1995:44). Consequently, <w> can only be interpreted as a mute vowel here.

    Now consider the spelling of glides. The phonemes /j/ and /w/ may be written as <i> and <j>, and <uw> and <w>, respectively:

(26)   jaar   bajes   kanjer   aai
        wee   leewieken   sperwer   eeuw

Parallel to *zaai-zaaien* we would expect the spelling *eeu-eeuen* but the actual spelling is *eeuw-eeuwen*. The <w> in *eeuw* and *nieuw* may be accounted for the same way as the mute <w> of words like *vrouw* i.e. as a representation of the inserted glide in inflected forms.[24]

---

[24] Support for this assumption may be found in the spelling of South African, which is historically

The literature yielded two types of rule for this phenomenon. The first was proposed by Te Winkel and adopted in subsequent editions of the *Woordenlijst*, see e.g. [*Woordenlijst* 1995], p. 14. These rules prescribe the spelling <i> in semi-diphthongs and <j> elsewhere. They could be formalized as the following pair of rules that are ordered by the Elsewhere Principle:

(27)   **Rules for the spelling of glides according to [*Woordenlijst 1995*]**
        /aj, oj, uj/   → <aai, ooi, oei>   /ew, iw/   → <eeuw, ieuw>
        /j/   → <j>   /w/   → <w>

However, as argued above, combinations such as /aj/ are not considered diphthongs here. Another disadvantage of these rules is that they predict an incorrect spelling for words such as *bajes*: **baaies*. This implies that another generalization has to be formulated.

Booij (1991:33) formulates another rule (where the analogous case of /w/ was not mentioned):

(28)   **Rule for the spelling of /j/ from Booij (1991)**
        write /j/ as <j> before a tautosyllabic vowel, and as <i> after a
        tautosyllabic vowel

This rule also predicts the incorrect spelling for words such as *bajes*. I will therefore propose an alternative solution.

On the basis of the distribution in (26) the following generalization can be formulated:

(29)   **Rules for the spelling of glides**
        a   /j/ is written as <i> morpheme-finally, and as <j> elsewhere
        b   /w/ is written as <uw> morpheme-finally, except after <u> and as
            <w> elsewhere

The difference between words such as *baaierd* and *bajes*, *majem* and *jajem*, thus follows from the morphological structure: *bajes* etc. are monomorphemic. *Leewieken* is a pseudo-compound (cf. *leewater*) and *wiewauwen* a reduplicative compound in which the [w] is not morpheme-final, so these words are written differently than *eeuwen*.

There are a few apparent exceptions to rule (29) such as *poeier*, *ooievaar*,

---

related to Dutch. In this language the inserted glide is not written. The sound /ɑu/ is always written as <ou>; variants with <w> do not occur and no <w> is added to <u> after <ee> or <ie> either, see for instance the contrast between Dutch *kieuw-kieuwen*, *leeuw-leeuwen* and *leeuwerik* with South African *kieu-kieue*, *leeu-leeue* and *leeuerik* or *lewerik* (see [*Afrikaanse woordelys*]). These facts argue in favour of a similar treatment of <w> in *louw* and *eeuw*.

kooiker, *lichterlaaie*; *krieuwel*, *Leeuwarden* with <i> or <uw> in non-final position. However, some of them are not really monomorphemic, for instance, *ooievaar* is a pseudo-compound, *kooiker* is related to *kooi*, and *poeier*, a variant of *poeder*, encodes the effect of d-weakening that is applied only before a suffix, see 5.4.2. With rule (29), *bajes* and *baaierd* are both regular. I conclude that rule (29) best captures the spelling of glides.

Note that the rules in (24) and (29) apply to morphemes rather than words (the latter option would lead to spellings such as *trouw-*getroud* and *draai-*gedraajd*). This appears to be one of a set of more general properties of phoneme-to-grapheme conversion to be discussed in the remainder of this study. Grapheme-phoneme conversion rules also apply to morphemes, see Nunn & Van Heuven (1993:89). This observation is another reason to reject Wester's combined sound-spelling representations that were created for words rather than for morphemes, see also Nunn (1992). I will discuss the domains of spelling rules in more detail in chapter 5.

Competing spelling variants

Finally, let us consider competing spelling variants. Traditionally, this kind of spelling variation is accounted for by historical sound rules. From a synchronic viewpoint, etymological variation is arbitrary, as illustrated by homophones (*reik ~ rijk, louw ~ lauw*), contrasts between semantically related words (*weifelen* and *twijfelen*) and spelling variants such as *rouwdouwer/rauwdouwer*, *rijschaaf/reischaaf*, *stampei/stampij*. These variants were introduced by [*Woordenlijst 1995*]. More variation can be found in names, e.g. *Arnoud/Arnaud, Marjolein/Marjolijn*. These facts show that we cannot predict the spelling of /ɛi/ and /au/ in native words. For this reason I will assume that the distribution of such spelling variants is unpredictable and has to be listed in the lexicon.

An alternative account would be the following. In similar cases where sound contrasts have been neutralized by sound rules but remained visible in spelling, linguists have assumed that the contrast is still present in the lexical sound representations of native speakers and neutralized by productive sound rules. For instance, Chomsky (1970:4) claimed that where English spelling deviates from pronunciation, spelling represents an abstract phonological level. In Chomsky's view lexical representations remain stable, even when the surface forms change as drastically as in English.[25] If we could demonstrate that such an analysis were

---

[25] A related proposal would be to postulate that sounds and spelling are two instances of a third representation which contains all information of both levels. Phonology and orthography both neutralize some of this information. This view, known as 'glossematics', has been developed for English in the thirties and forties by Hjelmslev, see Anderson (1985:140–168). In this view, we need not postulate abstract phonological representations, but this analysis has the disadvantage that a new

plausible for the facts at hand, we could propose that <ei> and <ij> or <au> and <ou> have distinct abstract sound representations. However, such a proposal has some serious drawbacks.

Firstly, the analysis forces us to postulate that historical rules and representations are still available to speakers of Dutch on the basis of spelling only. It is unlikely that speakers are able to learn such abstract sound representations, especially as spelling is not available when children learn to speak. It is not even clear what abstract sound representations should look like. For instance, words written with <ij> originally had a /i/ that was affected by diphthongization (*pilum* became *pijl*), see for instance Van Bree (1987:121–123). However, if we represented *pijl* as /pil/ to distinguish it from /pɛil/ written as *peil*, we cannot predict which /i/ sounds are affected by diphthongization. On the basis of similar arguments (i.e. the fact that Chomky's analysis implies that historical sound changes such as the Great Vowel Shift must be duplicated in the phonological competence of speakers) most linguists have abandoned Chomsky's view on English orthography, see for instance Derwing (1992:194). Secondly, this analysis incorrectly predicts that etymological spellings do not pose a problem for someone who has mastered Dutch phonology.

I will therefore not propose lexical sound distinctions between the sounds encoded by *ij* or *ei* and *au* or *ou*. In Appendix G I will discuss an analysis by Wester (1987b) of some types of spelling variation in non-native words (e.g. the sound /s/ written as *s* or *c* and /z/ written as *s* or *z*) which also involves the postulation of abstract lexical distinctions, and argue that it must be rejected for the same reasons.

Instead of listing all words containing sounds of which the spelling is unpredictable we could also try to find out if one of the competing variants is less frequent (the frequency referred to is lexical frequency), and only list the words with this spelling. This way, the number of words that have to be listed as exceptions can be reduced. To find out the proportion of the two spelling variants, let us consider all word types marked as monomorphemic in the test lexicon (see Appendix A). Thirty-five per cent of native monomorphemic words with the sound /ɛi/ are written with <ei> and 65% with <ij>. Thus, the sound /ɛi/ is most often written as <ij>. Now consider the spelling of the sound /ɑu/. Thirty-eight per cent of the native words with this sound are written with <au> and 62% with <ou>. Thus, native words with /ɑu/ in the lexicon are most frequently written with <ou>. I will therefore only list native words with <au>

---

level of representation has to be introduced and the nature of this representation which underlies both spelling and pronunciation is not clear. It also has the disadvantage that it does not do justice to the nature of orthography as a code for the pronunciation. No one has proposed such an extra level for Dutch. However, when we give it a non-synchronic interpretation, it corresponds to the view of Zonneveld (1980) and Kerstens (1981) in which orthography and phonology are considered different dialects of Dutch.

and <ei> in the lexicon. An exhaustive list of native exceptions in the test lexicon is given in Appendix B.

Apart from these classes of etymological spellings, there are some individual exceptions such as *oir*, *misschien*, *thuis*, *erwt*, *ambt* etc., see Appendix B. Note that some etymological spellings prompt an incorrect pronunciation, e.g. *-lijk*, *oir*, *bijzonder* as recognized by Te Winkel, see also Van Haeringen (1962).[26] Apparently, the Readability Requirement only governs the relation between the Phonological Principle and the Morphological Principle, not between the Phonological Principle and the Principle of Etymology.[27]

To conclude the discussion of the spelling of native words, I will give an overview of phoneme-grapheme relations, where abstract sound representations are related to the abstract spelling representations (i.e. where Vowel Degemination, Consonant Doubling and other autonomous spelling rules have not yet been applied).

---

[26] Te Winkel (1884:27) concludes that there are three types of spelling with respect to the pronunciation: 1. a spelling which is totally in accordance with the pronunciation, e.g *al, bast* ([ɑl, bɑst]) 2. a spelling, not totally according the pronunciation, but compatible with it because the pronunciation follows automatically from the mutual influence of letters, e.g. *abt, geenszins* ([ɑpt], [ɣɛnsɪns]); 3. a spelling, partially or completely contrary to and incompatible with the present day pronunciation, but necessary as the result of common use, e.g. *menschen, tusschen, thans* ([mɛnsə], [tʏsə], [tɑns]), *vijftig, zestig* ([fɛiftəx], [sɛstəx]).

[27] There are indications that the most frequent spelling is also considered the default case by writers. In some words of German origin <ei> has changed into <ij>, for instance, German *eifer* and *eilgut* became *ijver* and *ijlgoed*, see Van der Sijs (1996). Similarly, loan words with <au> have sometimes changed to <ou>. For instance, German *lauter* and Malayan *karbau* became *louter* and *karbouw*, see Van der Sijs (1996). Other indications are the inflected form *Nassouwe* of the name *Nassau,* and the fact that South African, which is related to Dutch, only has one spelling for /ɑu/, namely <ou>.

(30)   Phoneme-grapheme relations in native words

phonemes	graphemes	examples
[p]	p	[trɑp]-*trap*
[b]	b	[bɑl]-*bal*
[t]	t	[tɑk]-*tak*
[d]	d	[dun]-*doen*
[k]	k	[kal]-*kaal*
[f]	f	[fɛl]-*fel*
[v]	v	[vel]-*veel*
[s]	s	[sɔk]-*sok*
[z]	z	[zak]-*zaak*
[x]	ch	[sxol]-*school*
[ɣ]	g	[weɣə]-*wegen*
[h]	h	[hɛlp]-*help*
[n]	n	[nort]-*noord*
[m]	m	[moj]-*mooi*
[ŋ]	ng, n	[ɑŋəl]-*angel*, [bɑŋk]-*bank*
[l]	l	[lœyk]-*luik*
[r]	r	[ram]-*raam*
[w]	w, uw	[wer]-*weer*, [ew]-*eeuw*
[j]	j, i	[jar]-*jaar*, [maj]-*maai*
[ʏ]	u	[pʏt]-*put*
[ɪ]	i	[pɪt]-*pit*
[ɛ]	e	[pɛt]-*pet*
[ɔ]	o	[pɔl]-*pol*
[ɑ]	a	[jɑs]-*jas*
[i]	ie	[rit]-*riet*
[y]	uu	[myr]-*muur*
[e]	ee	[ler]-*leer*
[ø]	eu	[lØk]-*leuk*
[u]	oe	[rum]-*roem*
[o]	oo	[rot]-*rood*
[a]	aa	[kas]-*kaas*
[ɛi]	ij, ei	[wɛis]-*wijs*, [trɛin]-*trein*
[œy]	ui	[rœyk]-*ruik*
[ɑu]	ou(w), au(w)	[hɑut]-*hout*, [lɑu]-*louw*, [klɑutər]-*klauter*, [snɑu]-*snauw*
[ə]	e, i, ij, u	[də]-*de*, [nodəx]-*nodig*, [olək]-*olijk*, [dɔkəm]-*Dokkum*

When we compare this overview with the overview in (9), we see that phoneme-grapheme relations are less complex than sound-letter relations. Most phonemes are related to one grapheme or one combination of graphemes. Exceptions are competing and conditioned spellings. Another exception is formed by the fact that there is no unique representation of schwa. However, this may be appropriate since this vowel is considered absent in underlying sound representations, see Zonneveld (1993) and footnote 6.

The simplification of (30) as compared with (9) is partly due to the fact that the effect of sound rules need not be taken into account. Another reason is that the effect of autonomous spelling rules is not visible in abstract spelling representations. Of course, (30) only presents part of the derivation of abstract sound representation to surface spellings, which should be kept in mind when comparing it with (9): the coupling of underlying phonemes with underlying graphemes is easier, but a second type of rule is necessary. In the remainder of this study I will refer to rules that relate abstract sounds to letters as phoneme-to-grapheme conversion rules. The phoneme-to-grapheme conversion rules that derive the spelling of monomorphemic native words are listed in Appendix B.

2.6 Conclusion

This chapter investigated how the spelling of native words can be derived from their pronunciation. I started out with a very simple view of the relation between sounds and letters:

(31) | (abstract) pronunciation |

                ↓   *phoneme-to-grapheme conversion rules*

      | (surface) spelling |
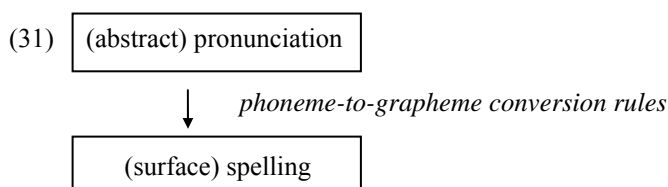
Discussion of the data showed that this model had to be modified. I argued that the alternation of single letters and geminates (*jaar-jaren*, *kar-karren*), which is accounted for in various ways in the literature, can best be described by means of spelling rules applied to letter sequences without reference to the corresponding pronunciation. However, unlike authors such as Zonneveld, I do not postulate autonomous spelling rules only. Rather, I propose that model (31) be extended with a second stage in which the result of phoneme-to-grapheme conversion rules is modified by autonomous spelling rules. Since phoneme-to-grapheme conversion rules neutralize some distinctions that are relevant to the proper application of rules such as Vowel Degemination, I propose that these distinctions be encoded in abstract spelling representations. In line with current

practice in phonology this was done by relating letters to an orthographical CV-tier.

If we apply these modifications to model (31) we get the following model of the relation between sounds and letters:

(32)   **Two-level spelling model**
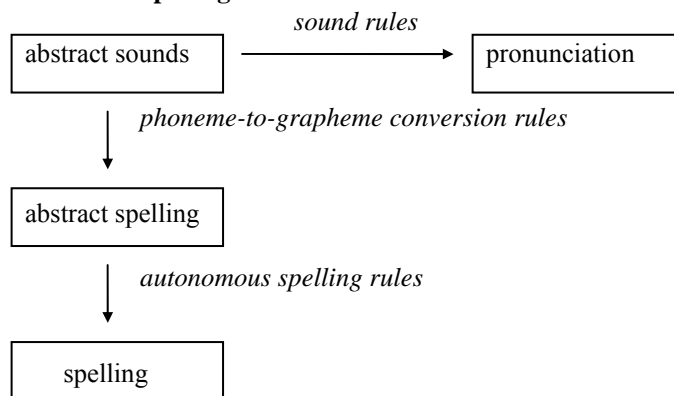
This extended model combines the advantages of spelling as a code for the pronunciation and Autonomous Spelling: it has rules that relate (abstract) sounds to spelling as well as rules that describe the relation between abstract and surface spelling. This way, it accounts for the fact that spelling can be predicted on the basis of the pronunciation (by phoneme-to-grapheme conversion rules) as well as for mismatches between pronunciation and spelling (by autonomous spelling rules). The result is a more accurate and restrictive description of the spelling of monomorphemic native words than would be possible with model (31) or Autonomous Spelling (15). Another advantage of the new model is that the introduction of the second type of rule enables us to formulate phoneme-to-grapheme rules that need not account for some types of spelling variation. This implies that the resulting spelling system comes closer to the ideal one-to-one relation between sounds and letters (where we relate abstract sound representations to abstract spelling representations). The role of context in phoneme-to-grapheme conversion rules is thus reduced to some historically motivated cases (*houd* ~ *vrouw*, *aai* ~ *ja*). What remains is only a small residue of words with an unpredictable (etymological) spelling (*wei*, *lauw*). These cannot be derived by rule.

Chapter 3

The uniform spelling of morphemes

3.1 Introduction

In the preceding chapter, we have seen that written morphemes are often less variable than their spoken counterparts. Firstly, the spelling of morphemes remains constant where the pronunciation alternates. Some examples are given under (1):

(1)   [ɪn]-[ɪmpɑkə]   in-**in**pakken (*impakken)
        [kni]-[knijə]   knie-knieën (*kniejen)
        [ɑls]-[zowɑls]   als-zoals (*zowals)
        [ɣan]-[ɔpɣan]   gaan-op**g**aan (*opchaan)
        [ɔp]-[ɔpdun]   op-o**p**doen (*obdoen)

The words in (1) illustrate that spelling abstracts from the effect of the sound processes Nasal Assimilation, see Zwaardemaker & Eijkman (1928:228–229), Homorganic Glide Insertion, see Zonneveld (1978:64–73) (but cf. footnote 8 in chapter 2) and Progressive and Regressive Voice Assimilation, see Zwaardemaker & Eijkman (1928:224–227). Secondly, the spelling of morphemes remains constant even where spelling is context-dependent and addition of an affix crucially changes the context. This is illustrated under (2):

(2)   trouw   ge*trouw*d (*getroud)   cf. koud
        aai   *aai*en   (*ajen)   cf. bajes

In chapter 2, the uniform spelling of words such as those under (1) was accounted for by the following provisional statement: spelling encodes the abstract sound representation of morphemes. This implies that context outside

the morpheme is irrelevant to phoneme-to-grapheme conversion rules, and that the effect of sound rules that operate across morpheme boundaries is not encoded in spelling. In this chapter, I will examine this claim in more detail.

There is reason to doubt that phoneme-to-grapheme conversion rules are restricted to the morpheme domain, since the data are not as straightforward as suggested by (1) and (2). Firstly, some spelling rules apply to domains larger than the morpheme, and thus result in variation of the spelling of the same morpheme:[1]

(3)   a   ra**men** (r**aa**m+en)   ra**mm**en (ra**m**+en)
       b   ge**ï**nd (ge#in+d)   pa**'**s (pa+s)
            neur**ië**n (neur**ie**+en)   koni**n**kje (koni**ng**+tje)

However, it was argued in chapter 2 that the alternation of single letters and geminates in (3a) can be described more adequately and insightfully by autonomous spelling rules which are conditioned by orthographical context rather than phonological context. In chapter 4, it will be argued that the same holds for the alternations in (3b). By postulating that only phoneme-to-grapheme conversion rules are restricted to the morpheme domain, it is possible to maintain the claim that spelling encodes the sound representation of morphemes and still to account for variable spelling of morphemes in (3).

There is a second, more serious challenge to the claim that spelling encodes the sound representation of morphemes. As pointed out by linguists such as Booij et al. (1979), Dutch orthography does not abstract from the effect of all sound processes which apply across morpheme boundaries. To be more precise, the effect of morpholexical rules, i.e. sound rules which only apply to specific (sets of) morphemes, is visible in writing. This phenomenon is illustrated under (4) and (5). Examples (4a-d) show the effect on spelling of Diminutive Allomorphy, see Trommelen (1983b), alternations in verbal endings -t(e)/-d(e), see Zonneveld (1978), Prevocalic Schwa Deletion, see Booij (1995:67), and alternation of -er and -der, see Booij (1995:73–75):

(4)   **Variable realizations of morphemes**
       a   [bɛit**j**ə]   ~   [bɔm**p**jə]   bij**tje** (bij)   ~   boom**pje** (boom)
       b   [kɑm**d**ə]   ~   [kɑp**t**ə]   kam**de** (kam) ~   kap**te** (kap)
       c   [lØk**ə**r]   ~   [rar**də**r]   leuk**er** (leuk) ~ raar**der** (raar)
       d   [sxadə]   ~   [bəsxadəx]   schad**e**   ~   beschadig

The examples in (5) illustrate the effect on spelling of the following sound

---

[1] Variation of the spelling of related non-native words such as *muziek-musicus*, *synoniem-synonymie*, *trochee-trocheïsch* etc. will be discussed in chapter 4.

changes: [ŋ]-[ŋk] Alternation, see Trommelen (1983b), Devoicing before *-elijk* and *-enis*, see Wester (1987), Vowel Lengthening, see Zonneveld (1978) and Booij (1981:69–72), Glide Insertion, see Gussenhoven (1980) and Verbal Ablaut, see Booij (1981:142). These rules do not apply to all the relevant words as illustrated by the examples in (5b):

(5) a [tuɣɑŋkələk]   toegan**k**elijk (toegang)
        [beltənɪs]   beel**t**enis (beeld)
        [lotjə]   l**oo**tje (lot)
        [kujə]   ko**ei**en (koe)
     b [bɑŋələk]   ban**g**elijk (bang)
        [bəlɛidənɪs]   belij**d**enis (belijden)
        [pɔtjə]   p**o**tje (pot)
        [muwə]   mo**e**ë (moe)

Note that the unpredictable glides in words such as *koeien* (*koe+en*) and *vlooien* (*vlo+en*), are visible in spelling, unlike the predictable inserted homorganic glides in words such as *knieën*.

   Van Heuven (1978) and Booij et al. (1979:82–87) also observed that the effect of some purely phonologically conditioned sound processes is reflected inconsistently. Survey (6) illustrates this for Final Devoicing, see Zonneveld (1983), Degemination, see Zwaardemaker & Eijkman (1928:231), D-deletion and D-weakening, see Zonneveld (1978:73–86), Booij (1995:91–93). The examples under (6a) show that Final Devoicing can be observed in the spelling of a word ending in /v/ or /z/, but not in a word ending in a plosive or /ɣ/. The words in (6b) illustrate that the effect of Degemination is represented word-finally, but not word-internally. Finally, the examples in (6c) show that D-deletion and D-weakening are visible in some cases only:

(6) **Effect of sound rule ignored**   **Effect of sound rule represented**
   a /strɑnd/   [strɑnt]   stran**d**   /lev/   [lef]   lee**f**
      /hɛb/   [hɛp]   he**b**   /vrez/   [vres]   vree**s**
      /draɣ/   [drax]   draa**g**
   b /hat+tə/   [hatə]   haa**tt**e   /ɣəhat+t/   [ɣəhat]   gehaa**t**
      /lɑnd+də/ [lɑndə]   lan**dd**e   /ɣəlɑnd+d/   [ɣəlɑnt]   gelan**d**
      /wɪl+loz/ [wɪlos]   wi**ll**oos   /bɪts+stə/   [bɪtstə]   bit**st**e
   c /ɣud+ə/   [ɣujə]   goe**d**e   /ɣud+ərd/   [ɣujərd]   goe**i**erd
      /ɑud+ə/   [ɑuwə]   ou**d**e   /ɑud+ələk/   [ɑuwələk]   ou**w**elijk

In addition to these sets of words with a variable spelling, there are also some individual examples in which the spelling of morphemes varies. For instance, spelling normally abstracts from Final Devoicing of plosives but not in *ab**t***

([ɑpt]), cf. *ab dij* ([ɑbdɛi].

The data in (1) and (2) can only be described adequately if spelling is derived from the sound representation of morphemes, but the facts in (4) and (5) suggest that spelling encodes the pronunciation of words. In the case of (6) both options lead to incorrect predictions. Stated otherwise, whether spelling is derived from abstract morphemes or from a less abstract sound representation of words, it will always lead to incorrect spelling of parts of the data.

In this chapter, we will investigate how to derive the correct spelling for all the examples given above. Section 3.2 gives a survey of the literature on the topic of the uniform spelling of morphemes. In 3.3 I will show that the different approaches to the uniform spelling can be captured in the form of spelling constraints. Since these constraints have disadvantages, a new constraint will be proposed. The facts in (4)-(6), which seem to contradict this new constraint, are accounted for in 3.4. Concluding remarks are given in 3.5.

3.2 Previous accounts for the uniform spelling of morphemes

Te Winkel (1863)
Te Winkel accounted for the fact that spelling abstracts from sound rules such as Final Devoicing by the postulation of the Principle of Uniformity and the Principle of Analogy. As mentioned in chapter 2, both principles will be referred to by the term Morphological Principle, which runs as follows:

> **Morphological Principle**
> Spelling reflects the sound representation of morphemes, as far as the pronunciation allows this.

The first clause of the Morphological Principle forbids all variation in the spelling of morphemes, but the second clause, the Readability Requirement (see 2.3.1), allows spelling to abstract from the effect of sound processes to avoid a spelling which is not compatible with the pronunciation. For instance, we write *boompje* and *vreselijk* since *boomtje or *vrezelijk would cause an incorrect pronunciation: *[bomtjə], *[vrɛzələk]. However, the Readability Requirement does not account for the fact that spelling inconsistently reflects the effect of Final Devoicing and Degemination. In the case of Final Devoicing, this is recognized by Te Winkel and accounted for by the assumption that plosives are not pronounced totally voicelessly, but fricatives are, see Te Winkel (1863:9, fn.), so the Phonological Principle accounts for the difference of *hebben-heb* and *leven-leef*.

Booij et al. (1979), Booij (1985, 1987, 1995)
Booij discusses the abstractness of the sound representation encoded by spelling in several publications. Booij et al. (1979:82) claim that spelling encodes the underlying representation of morphemes rather than the surface realization of words, but that the variable spelling of morphemes (4) and (5) is necessary since readability would be impaired by an abstract spelling. They note that some morpholexical processes are not totally predictable on the basis of the pronunciation (*bloempje/bloemetje*). In this case, the abstract spelling *bloemtje* would be ambiguous. This also suggests that the spelling variation is imposed by the Readability Requirement. However, this requirement does not account for those cases where sound rules are reflected inconsistently. Booij et al. (1979) point out that there are words whose spelling encodes the abstract representation, surface level or even an intermediate representation, where only part of the rules has been applied. For instance, in words such as *vreesde* the spelling abstracts from Voice Assimilation, but not from the earlier rule of Final Devoicing. Booij et al. (1979) therefore conclude that Dutch spelling aims at uniformity of morphemes, but that this principle is not applied consistently.

What is not noted by Booij et al. (1979) is that in their analysis no stage in the derivation of words such as *verhuisd* corresponds to the spelling, see Van Heuven (1978:57), Nunn (1992). The derivation of *vreesde* and *verhuisd* is given under (7):

(7)	/vrez + də/	/vər+hœyz+d/
Syllabification	vrez-də	vər-hœyzd
Final Devoicing	**vres-də**	vər-hœyzt
Assimilation	vrez-də	vər-hœyst
	[vrezdə] <vreesde>	[vər-hœyst] <verhuisd>

The pronunciation of the last two letters of *verhuisd* is [zd], [zt] and [st] at different points in the derivation, but never [sd] as it is in spelling. We will see below that Booij (1995) accounts for the alternation of [d] and [t] in another way (i.e. by Laryngeal Spreading), but in that analysis there is no intermediate stage with [sd] in the derivation either.

Booij (1985, 1987, 1995) emphasizes that the typology of sound rules plays a crucial role for the characterization of the degree of abstractness of Dutch spelling. Spelling abstracts from the style-dependent rules of connected speech, but encodes the effect of morpholexical rules.[2] In the case of the (obligatory)

---

[2] Some optional style-dependent processes whose effect is not represented in spelling are Vowel Shortening (/kanal/ → [kɑnal] *kanaal/*kannaal*), see Cohen et al. (1978:49), Kager & Zonneveld (1989), Vowel Reduction (/banan/ → [bənan] *banaan/*benaan*) see Kager (1989:297), Vowel Raising (/video/ → [vidijo] *video/*vidio*), see Kager & Zonneveld (1986), /wr/ → [vr], (/wred/ →

rules of word phonology, the situation is more complicated. This is illustrated by the following survey from Booij (1995:187) which summarizes the effect of strictly phonologically conditioned sound rules on orthography:[3]

(8) **Survey of orthographical representation of sound rule effects of Booij (1995:187)**

Rule	Effect represented	Effect not represented
Final Devoicing	/v, z/	/b, ɣ, d/
Laryngeal Spreading	always	
Nasal Assimilation	for [m]	for [ŋ]
Prevocalic Schwa Deletion	always	
Degemination	word-finally	word-internally
Homorganic Glide Insertion		always

Final Devoicing, Prevocalic Schwa Deletion, Degemination and Homorganic Glide Insertion were discussed above, cf. the examples in (1), (4c) and (6b). Laryngeal Spreading is the name of the rule introduced by Booij to account for the alternations in the verbal endings -t(ə)] and [d(ə)], which was illustrated under (4b). These alternations cannot be accounted for by regressive assimilation as proposed by Booij et al. (1979), since the /d/ normally gives its voicedness to the preceding obstruent (e.g. /knop+duk/ → [knobduk] *knoopdoek*), but *knoop + de* becomes *knoopte* (/knop+də/ → [knoptə]). Booij (1995) proposes that the relevant consonant be unspecified for voice, and that this feature is copied from the preceding consonant, see Booij (1995:62).[4] Nasal Assimilation was not discussed yet, since it never leads to variation in the spelling of related words (except in rare cases such as *hennep-hemp* which may not even be synchronically related).

On the basis of survey (8), Booij concludes that the effect of the rules of word phonology is represented inconsistently in spelling.

[vret] *wreed/*vreed*), see Zwaardemaker & Eijkman (1928:178), Nasal Deletion (/ɔnzekər/ → [ɔzekər] *onzeker/*ozeker*) and N-deletion (/opənən/ → [opənə] *openen/*opene*), see Zwaardemaker & Eijkman (1928:230), I-deletion (/spesial/ → [speʃal] *speciaal/*specaal*), see Booij (1995:138), T-deletion (/lʏxt/ → [lyx] *lucht/*luch*), see Zwaardemaker & Eijkman (1928:233), /sxr/ → [sr] (/sxrɛiv/ → [srɛif] *schrijf/*srijf*), see Zwaardemaker & Eijkman (1928:236), Schwa Deletion (/mɑkələk/ → [mɑklək] *makkelijk/*maklijk*) see Zwaardemaker & Eijkman (1928:236), intrusive stops (/kɔmt/ → [kɔmpt] *komt/*kompt*), see Zwaardemaker & Eijkman (1928:232), Schwa Epenthesis (/ɑrm/ → [ɑrəm] *arm/*arrem*), see Booij (1995:127).

[3] Booij also includes the palatalization of /n/ in *mandje*, which is omitted here since spelling abstracts from allophonic processes.

[4] Zonneveld (1983) claims that the first consonant of the suffix is a fricative, so that the alternation can be accounted for by Progressive Voice Assimilation, cf. *raadzaam* (/rad+zam/ → [ratsam]).

Wester (1987)

Wester (1987) criticizes Booij's conclusion that spelling aims at representing the underlying representation of morphemes and that the spelling of words such as *boompje*, *vrees* and *haatte* is exceptional. According to Wester, one could also draw the opposite conclusion on the basis of these facts, i.e., that words such as *boompje*, *gehaat* and *vrees* show that Dutch spelling aims at representing the surface form of words, and that the spelling of words such as *strand* and *haatte* is exceptional. Wester argues that seemingly exceptional facts should not immediately be seen as inconsistencies, but as a problem for the analysis. This approach is illustrated by an alternative analysis of one of the topics discussed by Booij, namely the representation of Final Devoicing. Like Booij, Wester claims that spelling is constant because it represents the underlying representation of morphemes and argues that the spelling difference between plosives and fricatives is caused by a difference in their underlying representation.

Wester points out some differences between the distributions of the fricatives /v/ and /z/ (but not /ɣ/, see Wester 1987:75) and other obstruents. Firstly, voiced and voiceless fricatives are in complementary distribution before vowels. A short vowel is typically followed by a voiceless fricative, and a long vowel (or a vowel followed by a sonorant) by a voiced fricative, see also Booij (1977):

(9) **The distribution of voiced and voiceless fricatives**

After long vowel or sonorant consonant		**After short vowels**	
voiced:	[boːzə], [bɔnzə] (boze, bonzen)	*[bɔzə]	(bozzen)
	[laːvə], [lɑrvə] (laven, larven)	*[lɑvə]	(lavve)
voiceless:	*[bosə] (bose, bonsen)	[bɔsə]	(bossen)
	*[lafə] (lafen, larfen)	[lɑfə]	(laffe)

Another difference between the two types of obstruents noticed by Wester is the fact that the fricatives /z/ and /v/ are always voiceless before the suffix *-elijk* whereas plosives are voiced:

(10) **The voicedness of obstruents before *-elijk***

[liːvə- liːfələk] lieve-liefelijk [hɛbə-*hɛpələk] hebben-hebbelijk
[vreːzə-vreːsələk] vrezen-vreselijk [doːdə-*doːtələk] doden-dodelijk
[draːɣə-draːxələk] dragen-dragelijk

Wester concludes that the sound pattern of fricatives is not derived in the same way as that of plosives, but that fricatives are underlyingly voiceless and become voiced between a long vowel or sonorant consonant and a vowel by Fricative

Voicing. For a related proposal, see Kooij (1983). Wester formulates Fricative Voicing as follows:

(11)   **Fricative Voicing**
       [−son, +cont] → [+voice] / [+voc] [+son] __ [+voc]

This rule is blocked before the suffix *-elijk*, so the underlying voiced sound surfaces in these contexts.

    This reanalysis of the phonological facts allows another account of the different spelling of plosives and fricatives in which the ordering of rules plays a crucial role. If fricative voicing precedes Final Devoicing, spelling corresponds to an intermediate stage in the derivation, between these two rules:

(12)   /strɑnd/   /lef/   /lefən/
         voicing   levən (*phoneme-to-grapheme conversion*)
         devoicing   strɑnt
         other rules   levə
                              [strɑnt]   [lef]   [levə]

This, however, incorrectly predicts that a word such as [drɑxələk] is written as **draachelijk*, but Wester explains the actual spelling *dragelijk* by claiming that <g> represents both voiced (*dragen*) and voiceless velar fricatives (*dragelijk*).

    In Wester's view, the difference between *strand* and *leef* thus corresponds to a difference in the underlying sound representation, and the effect of Final Devoicing is consistently represented in orthography.

3.3 Introducing a new account

Te Winkel's Morphological Principle suggests that Dutch orthography is subject to the following constraint:

> **Constraint A**
> The spelling of morphemes must be constant (as far as the pronunciation permits this).

This constraint posits a one-to-one relationship between visual representations and morphemes, which may facilitate perception, see also Van Heuven & Birkhäger (1983). This constraint is very strong; not only does it forbid that spelling represents the effect of sound processes, but it also rules out all spelling variation which is not forced by the Readability Requirement. Constraint A does not imply that spelling encodes abstract phonological representations, but in

most cases this happens to be the only way to ensure a uniform morpheme spelling. In principle it is also possible to encode one of the surface realizations as long as this does not violate the Readability Requirement. In fact we have already seen that pairs such as *vrouw-vrouwe*, the effect of glide insertion (which only occurs if a vowel initial suffix is added) is encoded instead of ignored in all instances of the word *vrouw*. Another example is that Nasal Assimilation is represented within morphemes ([rɑmp]-*ramp/*ranp*) but not across morpheme boundaries ([ɪmpɑkə]-*inpakken/*impakken*)

However, Constraint A is too strong, since it also prohibits the variation introduced by autonomous spelling rules, i.e. phenomena such as the alternation of single letters and geminates (in 5.2 we will see that such alternations cannot be explained by the need to provide a readable code for the pronunciation). Alternations such as *mogen-mocht* also violate Constraint C, since the uniform spelling *mogt* would not lead to reading difficulties. Te Winkel accounted for the inconsistent representation of Final Devoicing by the assumption that the *d* in words such as *haard* is not pronounced totally voicelessly, but the *s* in *huis* is. However, even if there was a phonetic difference (in an earlier stage of Dutch), this would not be relevant since spelling encodes phonemes, not allophones, see 2.2.1.

Booij and Wester propose a different approach to the uniform spelling of morphemes. They claim that the uniform spelling of morphemes in related words is a consequence of the fact that spelling encodes the sound representation of morphemes as listed in the lexicon, which is more stable than the surface realization. This could be formulated as follows:

Constraint B
Phoneme-to-grapheme   conversion   must   be   applied   to   underlying representations.

This constraint suggests the following relationship between phonology and orthography:[5]

---

[5] This type of account of the constant spelling of morphemes was first proposed for English by Chomsky. Chomsky (1970:4) made the following claim about English orthography: "Conventional English orthography in its essentials is to a large extent a direct point-by-point transcription of a system that the speaker of English has internalized and uses freely, a system that I will refer to as 'lexical representation'." According to Chomsky (1970:12), it is not surprising that orthography corresponds to the underlying representation. Firstly, the underlying representation is stable across time and across dialects, and is therefore more suitable to be encoded in writing than the variable surface realizations. Secondly, by representing underlying representations and eliminating all phonetic detail which is predictable by rules, orthography can be directly mapped onto the smallest semantically significant units, i.e., morphemes. In other words the abstract sound representation and spelling are isomorphic codes for morphemes.

(13)

| underlying representation | *rules A...Z* | orthography |

↓ *rules a...z*

| pronunciation |
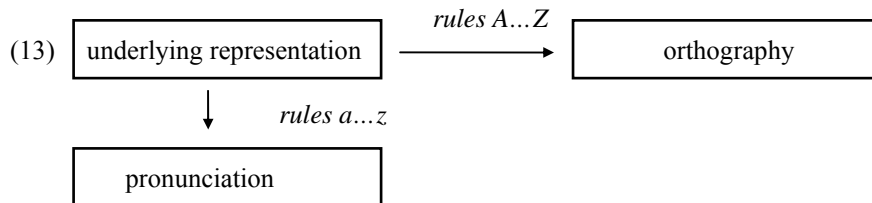
However, it should be noted that neither Booij nor Wester succeeds in demonstrating that spelling consistently represents underlying representations. At the most their work suggests that spelling encodes an intermediate level in the sound derivation. This suggests that Constraint B must be somewhat relaxed:

Constraint B'
Phoneme-to-grapheme conversion rules must be consistently applied to an intermediate level in the phonological derivation.

This proposal has three disadvantages. Firstly, as soon as one assumes that spelling does not represent the underlying representation, there are many possible intermediate levels. In this approach there is no restriction on the possible relations between different phonological levels and orthography, see Nunn (1992). Secondly, unlike Constraint A, Constraint B' does not account for the fact that the morpheme-internal application of the rule nasal assimilation is visible in orthography, but the effect across morpheme boundaries is not. Finally, Constraint B' does not account for the fact that phoneme-to-grapheme conversion rules apply to morphemes.

The first disadvantage can probably be resolved by an innovation of phonological theory, namely Lexical Phonology, see for instance Kiparsky (1982) and Mohanan (1985). In this theory a distinction is introduced between rules that apply in the lexicon (and have access to morphological information) and post-lexical rules. In the case of Dutch, Booij & Rubach (1987) and Booij (1995:53–57) distinguish three rule types: (i) cyclic rules that have access to morphological information and are therefore applied in the lexicon, (ii) word rules or post-cyclic rules (Final Devoicing, Voice Assimilation and Degemination) which apply after morphology but still within the lexicon, and which can refer to prosodic words and (iii) post-lexical processes which can only be applied after the creation of phrases in syntax.[6] If we ignore the

---

[6] Final Devoicing cannot be a cyclic rule, because in that case we would apply it to *strand* before the addition of *-en* and derive *[strɑntə]. It cannot be a postlexical rule either, since it is insensitive to the effect of postlexical resyllabification, so it is a word rule, see Booij (1995:55). Since Voice Assimilation and Degemination are fed by Final Devoicing (*handtas*: /hɑnd#tɑs/ → hɑnttɑs →

inconsistencies in (6) for a moment, we see that the distinction between cyclic rules and word rules corresponds almost perfectly to the division between rules whose effect is visible in spelling and rules of which orthography abstracts.

However, the other two disadvantages of Constraint B' cannot be solved by this approach. Firstly, we now expect that the effect of cyclic rules on derived words is visible in orthography as well, but this is not the case with Nasal Assimilation. Secondly, if spelling corresponds to the output of the cyclic level where rules may refer to derived words, there is no reason why phoneme-to-grapheme rules should be restricted to the morpheme domain.

Let us therefore consider an alternative account for the fact that spelling represents the effect of some sound rules only, which is closer to the constraint suggested by Te Winkel's Principle of Uniformity:

Constraint C
Spelling encodes the sound representation of morphemes.

Like Constraint B', this constraint accounts for the uniformity of written morphemes by the secondary nature of spelling. However, since it only requires that spelling encodes the sound representation of morphemes it does not forbid that autonomous spelling rules cause spelling alternations or that the morpheme-internal effect of rules is represented. I conclude that the uniform spelling of Dutch morphemes can be accounted for most adequately with Constraint C. However, Constraint C requires that spelling abstracts from the effect of all rules which operate across morpheme boundaries. The problem thus becomes how to reconcile C with the fact that spelling (sometimes inconsistently) encodes the effect of rules which involve domains larger than the morpheme. This will be the topic of the following sections.

3.4 Accounting for the variation in the spelling of morphemes

According to Constraint C, variation in the spelling of morphemes is not allowed if it represents the effect of sound rules, or if it is caused by the application of phoneme-to-grapheme conversion rules to domains larger than the morpheme. This implies that two types of variation are not in conflict with constraint C:
- spelling variation which is caused by autonomous spelling rules;
- spelling variation between pairs of words which although semantically related, are not derived from the same morpheme but form separate lexical entries.

In 3.4.1 and 3.4.2 I will examine which facts can be accounted for this way.

[hɑntɑs]; *hebzucht*: /hɛp#zɤxt/ → hɛpzɤxt → [hɛbsɤxt]) they must be word rules as well.

Section 3.4.3 discusses the remaining violations of Constraint C.

3.4.1 Variation as the result of autonomous spelling rules

Booij and Wester try to account for the alternations in (6) by a phonological analysis, but this proved to be difficult. Wester claims to have at least solved one of the problems, i.e. the deviating spelling of fricatives. However, on closer investigation Wester's analysis is not tenable as pointed out in Booij (1991) and Nunn (1992). The main argument against Wester's approach is formed by the realization of the past tense suffix and past participle suffix. These suffixes are voiced or voiceless depending on the final phoneme of the stem: [pɑkə-pɑktə] (*pakken-pakte*), [kɑmə-kɑmdə] (*kammen-kamde*). If Wester's analysis were correct, it should be *[leftə] (*leefte*) and *[vrestə] (*vreeste*) instead of the actual [levdə] (*leefde*) and [vrezdə] (*vreesde*). For this reason I will not adopt Wester's analysis.

The spelling model proposed in this study allows another option: alternations can also be the effect of autonomous spelling rules. This, in fact, has already been proposed in Zonneveld (1980) in whose autonomous model spelling alternations are accounted for by autonomous spelling rules only.

First consider alternations such as *huis-huizen*. Zonneveld accounts for these facts by means of a spelling 'devoicing' rule that only affects <v> and <z>, see Zonneveld (1980:532). This approach eliminates all inconsistencies: spelling does not inconsistently represent the effect of devoicing, but the spelling system has its own rule which accounts for *z/s*- and *v/f*-alternations. The data in (7) illustrated that there is no stage in the derivation of words such as *verhuisd* which corresponds to its spelling, which suggests that the alternation of *z* and *s* is not the result of the representation of a sound rule. A further indication that we are dealing with an orthographic rule is the following: the choice between <v> or <f> and between <z> or <s> is governed by syllable structure, and cases where phonological syllables and orthographical syllables differ (as shown by their hyphenation positions), show that the orthographical syllable determines the spelling choice:

(14)		**Pronunciation**	**spelling**		
	a	[r∅-**z**ɑxtəx]	*reu.**za**chtig	~	reu**s**.achtig
		[ɣrɛi-**z**art]	*grij.**za**ard	~	grij**s**.aard
	b	[vɛin-**z**art]	vein.**za**ard	~	*vein**s**.aard
		[ɣrɛin-**z**art]	grijn.**za**ard	~	*grijn**s**.aard

In phonology, all words in (14) are syllabified as monomorphemic words (this is the normal pattern for words with the suffix -*aard*, and *reusachtig* is a lexicalized compound), but in orthography -*aard* and -*achtig* are treated as separate syllabification domains, see [*Woordenlijst 1954*], p. LII. The words in (14b) have an exceptional orthographic syllable structure according to [*Woordenlijst 1954*], since they are morphologically irregular, i.e. derived from verbs instead of adjectives, see Te Winkel (1884:213). The sound /z/ is written as <z> in these words. The pronunciation of the fricatives in (14a) and (14b) is the same, but the spelling differs: the examples show that /z/ is written as <s> at the end of an orthographic syllable, and as <z> at the beginning, and that phonological syllabification is irrelevant. The description of the alternations in [*Woordenlijst 1954*], p. XLIII also suggests that we are dealing with an orthographical alternation: 'The symbols *v* and *z* are only written at the beginning of a syllable. Where the principle of congruence would require a final *v* or *z*, they are replaced by *f* and *s*'.[7] The alternations can be captured by the following rules:[8]

(15)   **Spelling devoicing**
         $z \rightarrow s \quad / \_\, cons_0]_s$ e.g. *huis, gevreesd*
         $v \rightarrow f \quad / \_\, cons_0]_s$ e.g. *leef, hoofd*[9]

Unlike the traditional sound-based analysis, (15) gives an accurate description of the actual patterns.[10] Spelling devoicing has some exceptions, but these are all

---

[7] "De tekens v en *z* worden alleen aan het begin van een lettergreep geschreven. Waar volgens het gelijkvormigheidsprincipe *v* of *z* op het einde zou moeten worden geschreven, worden deze door *f* en *s* vervangen [...]"

[8] Similar rules exist in English where they are clearly independent of the pronunciation. For example, syllable final <z> is allowed, but syllable final <v> is not. Therefore, only <z> can be doubled: *puzzle* versus *gravel*. It is clear that the alternation cannot be accounted for by the pronunciation here, since voiced fricatives are not devoiced in English (with only a few exceptions such as *calf-calves*), and freely occur after short vowels in words such as *quiver* and *busy*. Words that are pronounced with a final [v] are not written with <f> as in Dutch, but with the addition of an <e>, for instance in words such as *hive* and *give*, see Schane (1977).

[9] Zonneveld (1980:532) formulates the rule differently: [+cons, +cont, −back] → [−voice] / _ #. This rule, which presupposes that graphemes have distinctive features, is not simpler than the one proposed here, since <v> and <z> do not form a natural class anyway. I will therefore not adopt Zonneveld's rule.

[10] A possible motivation for spelling devoicing is the following. Remember that there is a gap in the distribution of /v/ and /z/: these sounds do not occur after short vowels. Since /v/ and /z/ do not occur after short vowels, <v> and <z> are never doubled. Finally, <v> and <z> do not occur after vowel geminates or sonorants (in contrast with letters that correspond to plosives). Next to words such as *ambt*, there are no words such as **beezd* or **hoovd* (cf. *Beesd* and *hoofd*). The following pattern now emerges: <v> or <z> only appear at the beginning of a syllable, <f> and <s> elsewhere. It is possible that writers assume that this distribution is the result of a requirement that <v> and <z> may only appear at the beginning of syllables in orthography and that word-final <v> and <z> are changed to

loan words, see chapter 4.

The analysis of words with *-elijk* presented here implies that words such as *dragelijk*, *walgelijk* have an idiosyncratic spelling: since the alternation of voiced and voiceless sounds before this suffix is unpredictable, we have to be postulate that the complex words are listed in the lexicon, and that the written form represents the 'devoiced' consonants in words such as [wɑlxələk], but we write *walgelijk* instead. The expected spelling **draachelijk*, **walchelijk* may have been rejected because <ch> very rarely occurs after vowel geminates or consonants (only before voiceless sounds, e.g. *burcht*).

Another type of spelling variation which can be accounted for by an autonomous spelling rule is the alternation of single and double consonants at morpheme boundaries. In cases where the sound rule Degemination is applied across word boundaries or fed by other sound processes such as Regressive Voice Assimilation and T-deletion, the corresponding written words have geminates:

(16)   /ɪn nem/   [ɪnem]   *ineem   inneem
         /ɪn mak/   [ɪmak]   *imaak   inmaak
         /vər brɑnd t/ [vərbrɑnt]   *verbrant   verbrandt
         /kʏnst stɔf/   [kʏnstɔf]   *kunstof   kunststof

The postulation of an orthographic Consonant Degemination rule is further supported by the fact that Degemination also occurs after silent letters in French loan words:

(17)   **singular**   **plural**
         [burzjwɑ]   bourgeois   [burzjwɑs]   bourgeois *bourgeoiss
    cf. [sɪrkəl]   cirkel   [sɪrkəls]   cirkel**s**

In the pronunciation of the word *bourgeois* (singular) there is no /s/, so the fact that we write the plural form with one <s> rather than two can not be accounted for by phonological Degemination. An orthographical Degemination rule would derive the correct result for words such as *bourgeois* (other examples are *chassis*, *compromis*, *parcours*, *pardessus*, *permis*, *relais*, *remous*, *rendez-vous*, *revers*, *sousbras* and *vis-à-vis*). This suggests that there is an orthographic Degemination rule next to the sound rule.[11]

---

satisfy this requirement. This analysis is supported by the contrast between *grijs-aard* and *grijn-zaard* and by the fact that, contrary to what is claimed by Wester, only <v>/<f> and <z>/<s> but not <g>/<ch>   are   in   complementary   distribution: **lozze/*haz.bel/*beezd/*huiz*; **grovve/*huv.der/*hoovd/*leev* but: *vlugge/vreug.de/voogd/maag*. This implies that Spelling Devoicing concerns exactly those letters which correspond to sounds with a restricted distribution.

[11] Similar rules exist in other languages: cf. for instance English *eighth/*eightth* (*eight+th*), see

With the postulation of an autonomous rule of spelling Degemination, contrasts such as *haatte-gehaat* cease to be inconsistencies. For instance, consider the facts in (6b), repeated under (18):

(18)   [hatə]   haa**t.t**e   [ɣəhat]   ge.haa**t**
         [lɑndə]   lan**d.d**e   [ɣəbrɑnt]   ge.bran**d**

On the basis of the spelling the difference between application and non-application of Degemination is obvious. The spelling rule is restricted to tautosyllabic consonants (remember that orthographic syllables sometimes deviate from their phonological counterparts):

(19)   **Consonant Degemination**
        $C_i \rightarrow \emptyset$ / [$_s$ …C_i_]$_s$ e.g. *verbran**d***

In this way we can account for the contrast between *verbrand-de* and *verbrand*. This approach is also suggested in [*Woordenlijst 1954*], p. XLIII.[12] The spelling rule thus differs from the corresponding sound rule with respect to its domain of application. Note that the rule seems to work across syllable boundaries in words such as *wijste* (*wijs+st+e*). This can be accounted for by postulating that Orthographic Syllabification and Consonant Degemination apply each time a suffix is added, see 5.3.

The final type of spelling variation that can be accounted for by an autonomous spelling rules concerns alternations such as *schade-schaden*. Instead of assuming that spelling represents the effect of Prevocalic Schwa Deletion, it could be assumed to be the result of an orthographic rule that deletes vowel letters in this context. This rule is needed anyway for pairs like the following: [duʃ-duʃə] *douche-douchen/*doucheen*; [rɛs-rɛsə] *race-racen/*raceen*. Since the final <e> of these examples is mute, its deletion cannot be the reflection of a sound process. Prevocalic Schwa Deletion applies to a vowel before a schwa and /rɛsən/ does not meet this structural description:

(20)   **Prevocalic E-deletion**
        e → 0 / _ + V
        CONDITIONS: e is not linked to a V-position, V is part of a suffix

---

Carney (1994:67), and, before the most recent spelling reform, German *Schiffahrt-*Schifffahrt* (*Schiff+Fahrt*), see [*Duden 1991*], p.:61.

[12] "Within a syllable, consonants are not doubled: hij praat; vermoord (verl. deelw.); *gedood; gehaat; gezet* etc. (despite *hij zingt; gehoord; getapt*); *een Fries, Parijs, Goes, Jutfaas, Waas meisje*; *dat is het wijst* enz." [In eenzelfde lettergreep wordt een medeklinker niet verdubbeld: hij praat; vermoord (verl. deelw.); gedood; gehaat; gezet enz. (ondanks: *hij zingt; gehoord; getapt*); *een Fries, Parijs, Goes, Jutfaas, Waas meisje; dat is het wijst* enz.].

The conditions prevent the rule from applying to the second <e> of *weeën* or to *beantwoord*.

3.4.2 Variation because of variation in lexical representations

Another type of spelling variation that does not violate constraint C involves pairs of words which although semantically related, are not derived from the same morpheme but form separate lexical entries.

First consider the alternations denoted as D-deletion and D-weakening by Booij (1995:91–93). The effect of these rules is sometimes visible in spelling. For instance, *glij* represents [ɣlɛi], derived from /ɣlɛid/. This would be problematic in the framework presented here if we were dealing with rules. However, as noted by Booij, this is not the case. The alternations are lexically governed and in some cases more than one realization is possible:

(21) **Optional application of D-weakening and D-deletion**

	pronunciation	spelling
a	[ɣlet-ɣle]	gleed
	[kɑudə-kɑuwə]	koude
	[ɣudə-ɣujə]	goede
	[rodə-rojə]	rode
	[bredə-brejə]	brede
b	[ɣlɛit-ɣlɛi]	glijd/glij
	[hɑut-hɑu]	houd/hou
	[kwadə-kwajə]	kwade/kwaaie
	[pudər-pujər]	poeder/poeier
	[ɑudə-ɑuwə]	oude/ouwe

In some cases, written forms abstract from the effect of D-deletion or D-weakening, see the examples in (21a), but in other cases two spellings are allowed, as illustrated in (21b). If we assume underlying morphemes as input for allomorphy rules that derive the surface realizations, we are forced to conclude that the spelling in (21a) is inconsistent with the spelling in (21b). The analysis with competing allomorphs allows another account. In written speech only the more formal allomorph is allowed in some cases. In this view we are not dealing with the inconsistent representation of the effect of rules, but with lexical differences between spoken and written language, see also Booij (1991:34).

Apparently, allomorphs such as *glee* and *kouwe* are judged too informal to be used in written language. This is purely a writing convention: when we write *gleden* and *koude*, we sometime read these words as [ɣlejə] and [kɑuwə]. On the basis of such mismatches Van Haeringen (1962) concluded that pronunciation

and writing are intrinsically different; although spelling is derived from the
spoken forms of words, it is not meant to be an accurate representation thereof,
since written language is more formal, see Van Haeringen (1962:43). It is not
clear why *glij* and *glijd* are both possible while only *gleed* is allowed (cf. also *rij*
~ **ree*, *snij* ~ **snee*). *Ouwe*, *rooie*, *poeier* and *kwaaie* also are informal but in
these cases it could be argued that they have different meanings than *oude*, *rode*,
poeder and *kwade*, which accounts for the fact that only the form affected by the
rule is used in fixed expressions such as *'kouwe ~ *koude kak'* ('posh'), *'de
kwaaie ~ *kwade Pier'* ('scapegoat'), *'geen rooie ~ *rode cent'* ('not a sou to his
name') and *'iemand een poeier ~ *poeder geven'* ('give someone a whack').

The competing allomorph analysis predicts that where D-deletion or D-
weakening are obligatory (not only in compounds as noted by Booij (1995:90),
but also in some derivations) spelling always reflects the effect of these rules:

(22)   **Obligatory application of D-weakening and D-deletion**

[ɣlɛiban]	glijbaan	[snɛiblum]	snijbloem
[hɑuvɑst]	houvast	[lɛidrat]	leidraad
[ɑuwəher]	ouweheer	[ɑuwələk]	ouwelijk
[kɑuwələk]	kouwelijk	[kwajəxhɛit]	kwaaiigheid
[ɣujərd]	goeierd	[ɣujəx]	goeiig[13]

In a derivational analysis this would just be a coincidence, since spelling
abstracts from obligatory as well as optional sound processes.

Since the sound alternations in (21) and (22) are not the effect of sound rules,
pairs such as *goede-goeie* and *koude-kouwelijk* do not form counterexamples to
the claim that spelling encodes underlying sound representations.

We can account for the facts in (5), repeated below, in the same way:

(23)   a   [tuɣɑŋkələk]   toegan**k**elijk (toegang)
            [beltənɪs]   beel**t**enis (beeld)
            [lotjə]   loo**t**je (lot)
            [kujə]   koe**i**en (koe)
        b   [bɑŋələk]   bang**e**lijk (bang)
            [bələɛidənɪs]   belij**d**enis (belijden)
            [pɔtjə]   po**t**je (pot)
            [muwə]   moe**ë** (moe)

---

[13] *Goedig* is not listed in [*Woordenlijst 1995*], but it does occur in [*Woordenlijst 1954*]. Neither
dictionary contains *goeiig*.

These alternations are not predictable, and must be described by listing different allomorphs in the lexicon, see chapter 4 of Booij (1995). If such alternations are lexicalized and available as the input of phoneme-to-grapheme conversion, spelling variation such as *toegang-toegankelijk* is unproblematic (the deletion of *g* in *toegankelijk* will be discussed in 5.6.4).

A final example of sound alternations that have to be analysed as the selection of competing allomorphs is formed by the spelling of clitics. For instance, the pronoun [hɛi] has a weak form [i] that is phonologically dependent on the preceding prosodic word: [hɛi kɔmt] (*hij komt*) versus [kɔmti] (*komt-ie*), see chapter 8 of Booij (1995). The weak form or clitic has a different distribution, determined by syntactic factors and is therefore listed as a separate entry in the lexicon. We always write the strong and more formal variant irrespective whether the strong or weak form was used in the pronunciation (except where a transcription of spoken language is intended).

The lexicon also plays a crucial role in the account of individual words of which the spelling seems to represent the effect of the sound rules Final Devoicing, Nasal Assimilation and Voice Assimilation:

(24)   a   [samə ~ təzamə]   **s**amen ~ te**z**amen

       b   [mɛt ~ medə]   me**t** ~ me**d**e

               [brabɑnt ~ brabɑndər]   Braban**t ~** Braban**d**er

               [moɣə ~ mɔxt]   mo**g**en ~ mo**ch**t

               [an ~ ambɛlt]   aa**n** ~ aa**m**beeld (also aa**n**beeld)

       c   [hɑndə ~ bɛidəhɑntə]   han**d**en ~ bijdehan**t**e

               [rɪt(ə) ~ rɛidə]   ri**t**(ten) ~ rij**d**en

               [zɑt(ə) ~ vərzadəɣd(ə)]   za**t**(te) ~ verzadig**d**(e)

               [vart(ə) ~ kopfardɛi]   vaar**t**(en) ~ koopvaar**d**ij

               [ɑf(ə) ~ avərɛxt]   a**f**(fe) ~ a**v**erecht

               [bəhɛpt(ə) ~ hɛbə]   behe**p**t(e) ~ he**bb**en

               [ɑpt(ə) ~ ɑbdɛi]   ab**t**(en) ~ a**b**dij

Instead of assuming that these words have a regular phonological derivation (/mɛd/ → [mɛt]) while their spelling corresponds to the surface realization, we could also consider the effect of the sound rules to be lexicalized in which case spelling is regular (/mɛt/ → <met>). For the word *samen* this is the only possible account. This word is derived from *tzamen* (cf. *gezamenlijk*) and the [s] is the result of Voice Assimilation, but since the conditioning /t/ has disappeared we must assume that the abstract sound representation has changed. In the case of (24b), we cannot be sure that the words are lexicalized. However, the inflected forms of the words in (24c) show that reanalysis sometimes occurs, even if there is a related word in which the underlying sound surfaces. Other examples are the variants *grint-grinten*, *bout-boute* next to *grind-grinden*, *boud-boude*. These

examples show that historically related words need not be derived from a common sound representation. For this reason, the spelling is not uniform either. The spelling alternations in (24b) could also be attributed to reanalysis of the sound representation of words which are no longer considered to be related to the forms where that underlying sound surfaces. Consequently, spelling does not reflect the effect of a productive sound rule and constraint C is not violated.

The words in (25) are similar to those in (24). Again, these words are no longer part of a paradigmatic set with words with voiced consonants, so there is no synchronic evidence for a underlying voiced consonant. The inflected forms in (25) show that the consonant in question has been lexicalized. The examples in (25c) are comparable to *samen* where the initial consonant has been devoiced under the influence of a voiceless consonant which is now absent. However, in this case the spelling seems to encode the underlying sound representation:

(25)　a　[ɑxt]　　　　　　　　　　aa**g**t (cf. also geneu**g**t(e), oo**g**st, no**g**)

　　　　　[fɔnts]　　　　　　　　fon**d**s (cf. also gi**d**s, loo**d**s, va**d**sig, -er**d**)

　　　　　[ɑmpt]　　　　　　　　am**b**t

　　　b　[hɛlmɔnt-hɛlmɔntər]　　Helmon**d**-Helmon**t**er

　　　　　[lelistɑt-lelistɑtər]　　Lelysta**d**-Lelysta**tt**er

　　　　　[sɪmpəlvɛlt-sɪmpəlvɛltər]　Simpelvel**d**-Simpelvel**t**er

　　　c　[fertəx-vir]　　　　　　**v**eertig-**v**ier

　　　　　[fɛiftəx-vɛif]　　　　　**v**ijftig-**v**ijf

　　　　　[sɛstəx-zɛs]　　　　　**z**estig-**z**es

　　　　　[sevəntəx-zevə]　　　　**z**eventig-**z**even

Apart from the spelling there is no reason to postulate a phonological difference between *aagt* and *mocht*; *zestig* and *samen* or *Helmond* and *met*, and a phonological analysis should not be motivated by spelling only. I conclude that the spelling of these forms does not correspond to an abstract sound representation, but is etymological: *aagt, abt, ambt, -erd*, and *oogst* are related to *aagjesappel, abdij, ambacht, -aard*, and *augustus*, respectively. In the case of the examples in (25c), the spelling often leads to a spelling pronunciation.

I conclude that spelling abstracts from the effect of Final Devoicing of plosives, Nasal Assimilation and Voice Assimilation, unless the effect of these rules has been interpreted as part of the underlying representation of words. The spelling of the words in (25) is idiosyncratic.[14]

---

[14]. The same holds for loan words such as *plebs, wodka, labskaus, drugstore, podsol, ombudsman.* In the following words the voicedness of obstruents or the presence of geminates is only motivated by the morphological structure: *absent, adstructie, obsceen, subsidie etc.; excellent, ex(s)tirpatie, transsubstantiëren, interruptie.* If non-native prefixes are no longer recognized as morphemes, these words have an etymological spelling as well. I will return to this subject in Appendix F

3.4.3 Remaining alternations

In the preceding section, it was argued that spelling consistently abstracts from the effect of Final Devoicing, Degemination and Voice Assimilation. This implies that the only sound rules whose effect on domains larger than the single morpheme is visible in spelling are the following:

(26)   Alternation of [ər] and [dər]
            Diminutive allomorphy
            Alternation of [d(ə)] and [t(ə)]

These alternations cannot be accounted for by autonomous spelling rules as illustrated below for the choice of past tense suffixes:

(27)   a   [vlɛi]-[vlɛidə]   vlei-vleide
           [sle]-[sledə]   slee-sleede
       b   [res]-[restə]   race-racete/*racede
           [slɑjs]-[slɑjstə]   slice-slicete/*slicede

Orthographically, the examples in (27a) and (27b) are comparable: both end in a vowel, so we would expect <d> to be chosen in both cases. However, this is not the case. If we only look at spelling, we cannot account for this contrast, but if we look at the pronunciation of the words we predict this difference: the <e> of *race* is mute.[15] The examples in (27) show that the sounds rather than the letters predict the choice between the different realizations of the suffix. The wrong spelling would also be predicted for diminutive allomorphs with a spelling-based analysis: on the basis of pairs such as *boom-boompje*, *dame-dametje*, *koord-koordje* and *doos-doosje*, we could formulate the following (simplified) generalization: write <pje> after <m>, <je> after <s> or <d> and <tje> elsewhere. However, this rule would predict an incorrect spelling for the examples in (28):

(28)   *crèmetje   crèmepje   [krɛːmpjə]
       *racetje   raceje   [resjə]
       *crapautje   crapaudtje   [krɑpotjə]
       *bourgeoisje   bourgeoistje [burzjwɑtjə]

---

[15] A well-known aid to memory to write the suffix correct in this case is the following: write <t> if the verb stem before the suffix -en ends in one of the sounds in *'t kofschip*, and write <d> elsewhere. This rule will only work when applied to sounds rather than letters: <x> is not in *'t kofschip*, so when the choice between <d> and <t> is based on the spelling, the wrong spelling is predicted, for instance **mixde* instead of *mixte*.

Again, the correct spelling can only be predicted on the basis of the pronunciation. The diminutive form *bourgeoistje* illustrates that Diminutive Alternation is of a different nature than Consonant Degemination. Spelling alternations such as Consonant Degemination are sensitive to the silent letter (*bourgeois/*bourgeoiss*) but morpholexical alternations are not.

The fact that spelling does not abstract from the sound rules in (26) can of course be accounted for by invoking the Readability Requirement. Readability sometimes requires that the effect of rules which involve domains larger than the morpheme is taken into account. Spellings such as **boomtje*, **pakde* and **rarer* would prompt an incorrect pronunciation.

A second option is the following. We have seen in 3.4.2 that some types of idiosyncratic morpholexical variation can only be described as the lexically governed choice between competing allomorphs. The fact that this variation is represented in orthography is therefore compatible with Constraint C. In the case of the more regular alternations under (26) we are not forced to posit a competing allomorph analysis, but it is at least possible. In fact, this has already been proposed by Van Zonneveld (1983). Van Zonneveld assumes that we write *raampje* because /-pjə/ is listed in the lexicon. However, this approach is often rejected with the argument that it would suggest that morpholexical variation is arbitrary from a phonological point of view, see for instance Booij (1995:62).

Booij & Van Santen (1995:132) draw another conclusion after discussing the distribution of *-er/-aar:* "[…] there may be a phonologically conditioned complementary distribution of affixes, without implying that these affixes should be viewed as each other's allomorphs, that can be deduced to one underlying form." This suggests that even if the occurrence of different forms of the same morpheme is conditioned by phonological factors, we need not explain this by postulating one underlying form. After all, the selection of plural morphemes *-s* or *-en*, where a common underlying form is not likely, are also conditioned by phonological factors (select *-s* after stem ending in a unstressed syllable (e.g. *vaders*, *opties*) and *-en* elsewhere (*ballonnen*, *fantasieën*), see Haeseyn et al. (1997:772, 180–181). An analysis that reduces allomorphs of one underlying form and an analysis that uses competing allomorphs are equivalent. An abstract underlying form has the advantage that we need only list one form in the lexicon, whereas the competing allomorphs variant has the advantage that we do not need abstract forms. In the case of Diminutive Allomorphy the postulation of different allomorphs is supported by the presence of variants such as the following, see Booij (1995:7172):

(29)   [blumpjə-blumətjə]   bloempje-bloemetje
       [vlɑɣjə-vlɑɣətjə]   vlagje-vlaggetje
       [mɑsjinə-mɑsjin(ə)tjə]   machine-machientje/machinetje

Thus, there are some indications that even regular morpholexical alternations can be analysed as a choice between competing allomorphs. Since the postulation of competing allomorphs does not prevent us from making generalizations about the phonological context in which they appear, I conclude that this analysis is equivalent to the derivational approach as far as phonology is concerned. From a spelling perspective this analysis is superior. It allows a more restrictive spelling account, since it allows us to describe variations such as *zeetje-boompje* and still assume that spelling encodes morphemes.

To be able to account for the alteration of [d(ə)] and [t(ə)] and of [ər] and [dər]as a choice between competing allomorphs it is crucial that we do not follow Booij who claims that these rules (as well as Prevocalic Schwa Deletion) are purely phonologically conditioned rules.[16] As argued above, an analysis with competing morphemes can express distributional generalizations as well: in case of the alternation of [d(ə)] and [t(ə)] this has the advantage that we need not stipulate underspecified segments in a specific morphological context as proposed by Booij. The reformulation of [ər]-[dər] alternation as a morpholexical rule is supported by the presence of variants: the verb *scoren* has an irregular variant *scorer* as well as regular *scoorder*, and the names *Durer* and *Prinsterer* do not undergo the rule, which argues against Booij's claim that d-insertion in the context /rər/ is purely phonologically conditioned.

I conclude that it is possible to account for the cases where spelling does not abstract from the effect of the rules that apply across morpheme boundaries by postulating that we are dealing with competing allomorphs. If further research reveals facts that prove that some of these alternations are the result of purely phonologically conditioned rules, it is still possible to account for the fact that their effect is visible in writing by invoking the Readability Requirement. It is not necessary to conclude this beforehand. I will therefore consider alteration of [d(ə)] and [t(ə)] and of [ər] and [dər] and Diminutive Allomorphy to be the result of selection of allomorphs rather than of modification of abstract morphemes.

3.5 Conclusion

In this chapter old and new insights were combined to account for the fact that the spelling of written morphemes is often constant across related words.

In the literature, we found two accounts for the uniform spelling of

---

[16] Section 3.4.1 revealed that alternations such as *kade-kaden* can be accounted for by a spelling rule that is needed anyway for the alternation of *douche*. Another option would be to reanalyse Prevocalic Schwa Deletion as a morpholexical rule. Support for this analysis can be found in the fact it does not define possible prosodic words, that is a property which distinguishes sound rules from morpholexical rules, see Booij (1995:57–58). Prevocalic schwa is not always deleted but occurs in some prosodic words (where a glottal stop is inserted): *geantwoord/*gantwoord*, *beoog/*boog*.

morphemes. Te Winkel (1863) assumes that spelling aims at a uniform spelling for morphemes (Constraint A). This constraint is contradicted by spelling alternations such as *raam-ramen*. Booij (1985) and Wester (1987) claim that spelling aims at consistently encoding one (possibly intermediate) level in the phonological derivation (Constraint B). This constraint is not violated by pairs such as *raam-ramen*, but it does not account for the fact that the effect of Nasal Assimilation within morphemes is visible, whereas Nasal Assimilation across morpheme boundaries is not, as illustrated by pairs such as *ramp-inpakken* (*in+pakken*). Neither does Constraint B account for the fact that phoneme-to-grapheme conversion rules are restricted to the morpheme domain (cf. *aaien* (*aai+en*) versus *bajes*). Thus, neither account is compatible with all types of spelling variation that occur in Dutch spelling.

For this reason, I proposed an alternative constraint (Constraint C), which requires that spelling encodes the sound representation of morphemes. Like Constraint B, this constraint accounts for the uniformity of written morphemes by the secondary nature of spelling, but it does not forbid the representation of morpheme-internal effects of sound rules, just like Constraint A. It is thus compatible with variation introduced by autonomous spelling rules, as well as with contrasts such as *ramp-inpakken*, while it still accounts for the uniform spelling of *aai-aaien* and *hond-honden*.

Some spelling alternations that seem to contradict Constraint C have been given an alternative analysis. Spelling variation such as *lieve-lief* or *haatte-haat* was accounted for by the autonomous spelling rules Spelling Devoicing and Consonant Degemination. In the case of alternations such as *hebben-behept* we assumed that the lexical representation contains voiceless sounds. Alternations such as *goed-goeie* do not reflect the effect of sound rules either. Rather we are dealing with competing allomorphs of the word *goed*. The remaining spelling alternations, i.e. alternation of *-er* and *-der*, Diminutive Allomorphy and alternation of *-d(e)* and *-t(e)*, can be accounted for by postulating that we are dealing with competing allomorphs rather than with different realizations of one abstract morpheme, or by invoking the Readability Requirement.

It should be noted that for the account presented here, the combination of autonomous spelling rules and phoneme-to-grapheme conversion rules is crucial. Without autonomous spelling rules, we could not account for contrasts such as *haatte-gehaat*, but without phoneme-to-grapheme conversion rules incorrect predictions would be made with respect to morpholexical alternation, e.g. **racede* instead of *racete* and **crèmetje* instead of *crèmepje*.

Chapter 4

The spelling of loan words

4.1 Introduction

In this chapter loan words will be examined in order to find out whether we can make generalizations about their spelling or whether their spelling is unpredictable.

Like all languages, Dutch occasionally adopts loan words. Since these words are often written as they are in the donor language, their spelling deviates from the indigenous pattern. For instance, loan words use letters and letter combinations that do not occur in indigenous Dutch words (*centraal* and *quintet*), or use the same letters to encode different sounds, e.g. native *zoom*-[**zo**m] versus non-native *zoom*-[**zu**m]. The spelling rules for Dutch words often make incorrect predictions for these words, for instance **lieter, *sent, *teze* instead of *liter, cent, these*. By using the spelling rules for native words, the correct spelling is predicted for 25% of the loan words only. In order to capture the spelling of loan words, Te Winkel formulated the *Principle of Etymology* that states that the spelling of loan words is given and does not follow from spelling rules. A strict interpretation of the Principle of Etymology implies that the spelling of loan words cannot be derived from their pronunciation by (Dutch) spelling rules. Te Winkel's view can also be illustrated by the following quotation:

> The spelling of foreign words and partly adapted words need not be determined by new and special rules. It is given: as far as the stem is concerned by the foreign spelling, as far as the suffixes are concerned by the normal Dutch spelling rules. [1]   Te Winkel (1863:183)

---

[1] "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."

There are facts that seem to support the assumption that the spelling of loan words cannot be predicted from their pronunciation, e.g., words where the same sound in a similar context is written differently, e.g. *vizier ~ visie, tractie ~ traktaat*, **k**olom ~ **c**olonne, **k**ompas ~ **c**ompassie, **c**arbon ~ **k**arbonade, **k**astanje ~ **c**astagnet and **c**apituleer ~ **k**apitaal, minimal pairs that are distinguished by their spelling only such as *se**c**reet* ('secret') ~ *se**k**reet* ('toilet'), *apsis* (id.) ~ *abscis* ('abscissa') and *colli* ('packages') ~ *collie* (id.), and spelling variants such as **choqu**eren or **shock**eren, **poe**let or **pou**let, **sch**ibbolet or **sj**ibbolet, and *talm**oed*** or *talm**ud***.

However, I will show that in many cases the spelling of loan words can be predicted by spelling rules, although not by the same rules that apply to indigenous words. Stated differently, non-native words sometimes have a regularity of their own. An example of this phenomenon is given in (1a), where the rules formulated for indigenous words incorrectly predict that long vowels are written as vowel geminates instead of single letters:

(1)   **Predicted spelling of long vowels**
         a   *   *cap**uu**chon*   *   *Ra**a**chel*
             *   *G**uu**staaf*   *   *po**o**ster*
         b   *go**o**chel*
                 *ro**o**ster*

(1b) illustrates that it is not the sequence of letters that is ungrammatical in (1a). The same sequence occurs in indigenous words. The spelling rules produce the incorrect result here since they were meant for indigenous words. It is possible to formulate a rule that yields the correct spelling for the loan words in (1a): write long vowels as geminates in final syllables, and as single letters elsewhere. In the literature such a regularity of loan words is only recognized for the spelling of the sound /i/. In this chapter, I will show that more aspects of the spelling of non-native words are predictable by context-sensitive spelling rules. Although the spelling system is somewhat complicated by the postulation of distinct rule sets, the result is a better description of the spelling of loan words with fewer exceptions. We will also see that the description cannot be further improved by formulating spelling rules for subsets of non-native words such as hybrid (partially changed) words and foreign words, or words of different origins, e.g. Greek words, French words, etc.

If there were no explicit criteria to distinguish indigenous words from loan words, or if these criteria referred to the spelling itself, distinct rule sets would not improve the description of the spelling system. However, Te Winkel already recognized that it is possible to classify words by referring to certain phonological and morphological properties. Te Winkel's observation was repeated in recent studies on the distinction between native and non-native

words in Dutch phonology and morphology. After a discussion of these studies, I will conclude that it is possible to recognize two distinct groups, i.e. native and non-native words.

This chapter is organized as follows. Section 4.2 gives an overview of the literature on the spelling of loan words. Section 4.3 investigates how we can recognize loan words. In 4.4 it will be examined whether phoneme-to-grapheme conversion rules and autonomous spelling rules distinguish between native words and loan words. In 4.5 the question will be addressed how the phoneme-to-grapheme conversion rules from 4.4 apply to complex words. The conclusions of this chapter are formulated in 4.6.

4.2 Literature on the spelling of loan words

This section gives an overview of the treatment of loan words in the literature. In each case, it will be examined how loan words are defined, whether loan words are further divided into smaller groups and whether special spelling rules have been formulated for (subsets of) loan words.

4.2.1 Prescriptive accounts

Te Winkel (1863)
Siegenbeek (1805a) and Te Winkel (1863) discussed loan words in order to establish whether their spelling should be adapted to the Dutch system. I will only summarize Te Winkel's account, since it is more explicit and elaborate.

Te Winkel compared loan words of which the spelling had already been adapted and unadapted words, and concluded that the regularity of the spelling is predictable from the extent to which words are adapted to the Dutch linguistic system, irrespective of their origin. Loan words such as *beschuit* (from French *biscuit*) and *kelder* (from Latin *cellarius*) have been totally adapted with respect to pronunciation and morphological behaviour. Such words can no longer be distinguished from Dutch words such as *bedrog* and *melder*, and the spelling of these words has been adjusted as well, see Te Winkel (1863:60–63). Words such as *misère* and *museum*, on the other hand, which betray their foreign origin by the fact that they contain a foreign phoneme ([mizɛːrə]) or have a foreign inflected form (*musea*), have also kept their foreign spelling and have not been changed into *miezère or *muzeum.

These examples suggest that loan words have either been totally adapted or have not been adapted at all. However, Te Winkel also observes that some loan words have undergone sound changes that were accompanied by spelling changes to prevent an incorrect pronunciation. For this reason, the spelling of some loan words has been partly adapted. An example is formed by the word

artikel. The spelling **articel* would be closer to the original *articulus*, but since this spelling would cause an incorrect pronunciation ([ɑrtisɛl]), the <c> has been changed to <k>.

On the basis of such observations, Te Winkel divided the lexicon into subgroups on the basis of the extent to which their pronunciation and morphological behaviour was comparable to that of Dutch words. Totally adapted words such as *kelder* were classed as native words. Te Winkel considers a word native if it contains native phonemes only, if it has maximally one full vowel in the stem, if it has a native stress pattern and if it is exclusively combined with native affixes. Partially adapted words such as *artikel* were denoted as hybrid words ('bastaardwoorden' literally means words of mixed origin), and the remaining, unaltered loan words such as *misère* as foreign words, see Te Winkel (1863:60–62). The distinction between hybrid and foreign words is thus not very clear.

A second factor that influences spelling adjustment according to Te Winkel (1863:74) is the familiarity of words: commonly used loan words that are not (completely) adapted to Dutch in sound and inflection still have an adapted spelling, e.g. *kazerne* (*caserne*) and *praktijk* (*practicum*). On the other hand specialized words kept their original spelling, for instance in case of words like *scrupel* and *pose*. Thus, in Te Winkel's view, the following types of spelling adjustment had taken place: total adjustment of native words; partial adjustment of hybrid words where necessary to prevent an incorrect pronunciation, and total adjustment of familiar words.

Te Winkel did not investigate whether partial spelling adaptation of hybrid words had resulted in a new spelling pattern and only recognized a native and a non-native spelling ('in- en uitheemsche spelling'). He claimed that only the spelling of native words can be derived by means of rules and that the spelling of hybrid and foreign words is idiosyncratic. Consequently, he does not recognize that it is possible to make generalizations about the spelling of loan words, not even for the spelling of /i/.[2] Te Winkel claimed that it is stress that determines the choice between <i> and <ie> in native as well as loan words. The phoneme /i/ is written as <ie> in stressed syllables (*kn<u>ie</u>*, *fa<u>brie</u>k*, *jezuï<u>et</u>*), but as <i> in unstressed syllables (*kr<u>i</u>oel*, *fabr<u>i</u>kant*, *jezuï<u>ti</u>sme*, *pro<u>za</u>ïsch*).[3]

Te Winkel eventually decided not to further adapt the spelling of loan words,

---

[2] However, in Te Winkel (1860:28–29), an overview of Siegenbeek's spelling rules, Te Winkel mentions the different spelling of /i/ in native words and loan words, and refers to stress only when discussing the spelling of endings like *-iek* for pairs like *fabriek-fabrikant*.

[3] It is hard to establish whether Te Winkel's predecessor Siegenbeek also made this distinction: Siegenbeek (1805b) gives no summary of spelling rules and it does not contain verbs like *ruziën* or plural forms of words like *bezie*. In Siegenbeek (1805a) and Te Winkel (1860), an overview of Siegenbeek's spelling rules, this alternation is not mentioned either. It is likely that Siegenbeek did not prescribe alternations of <i> and <ie> and that they are introduced by Te Winkel.

so the spelling of loan words that still betray their non-native origin by their pronunciation or inflection, remained subject to the Principle of Etymology.

[Woordenlijst 1954]

In [*Woordenlijst 1954*] the distinction between native words, denoted as Dutch words, hybrid words and foreign words still plays a role. After all, this dictionary was published because of a spelling reform of (among other things) hybrid words that therefore need to be distinguished from native and foreign words. Hybrid words are defined as loan words that have been partially adapted, foreign words as "words that are entirely foreign such as *alibi, meeting, thriller*, which have often become international in their original form […]".[4] Apart from the hybrid-foreign distinction, certain etymologically defined sets of words such as French words and Latin words were also referred to in the discussion of the spelling of loan words.

[*Woordenlijst 1954*] does not give special spelling rules for hybrid words except for the rule for the spelling of /i/, and a rule that prescribes the spelling of words like *stationeren*, see Appendix H. As already observed by Kollewijn (1916:141–142), it is not stress that determines whether we write <ie> or <i> as proposed by Te Winkel, see also Neijt & Zuidema (1994b:262). The vowel written as <ie> is not stressed in words such as *sub<u>sie</u>die* and *po<u>li</u>tie*, whereas <i> denotes a stressed vowel. If stress shift accounts for the spelling variation between *mark<u>ies</u>* and *mark<u>izaat</u>* we would also predict **mark<u>izin</u>* instead of *mark<u>iezin</u>*, and alternations such as *emp<u>irie</u>-*emp<u>ierisch</u>*. Kollewijn argues that the generalization is that /i/ is written as <i> in non-native words, except morpheme-finally. The generalization was adopted in [*Woordenlijst 1954*].

[*Woordenlijst 1954*] does not give rules for the spelling of other sounds in loan words, but only outlines the effect of the spelling reform, cf. for instance the following remark about the spelling of /e/. "*Ae* is mostly replaced by *e* (*equator, prestatie* etc.). Sometimes *ae* remains an allowed variant (*prefix, prehistorie* etc. also *praefix, praehistorie* etc.).", see [*Woordenlijst 1954*], p. XLVII.[5]

[Woordenlijst 1995]

In [*Woordenlijst 1995*] loan words are denoted as non-native words ('uitheemse woorden' or 'woorden van vreemde herkomst'), which are subdivided into etymologically defined subtypes. The distinction between hybrid and foreign words is no longer mentioned. The treatment of loan words has also been

---

[4] "Als bastaardwoorden beschouwen we niet de geheel vreemde woorden zoals *alibi, meeting, thriller*, die in hun oorspronkelijke vorm vaak internationaal zijn geworden.' [*Woordenlijst 1954*], p. XLVI).
[5] "*Ae* is veelal vervangen door *e* (*equator, prestatie* enz.). Soms is *ae* als bijvorm gehandhaafd (*prefix, prehistorie* enz. ook *praefix, praehistorie* enz.)." [*Woordenlijst 1954*], p. XLVII.

changed. Firstly, the only rule for non-native words that regulated the spelling of the sound /i/, was reformulated. Unfortunately, the rules for /i/ again refer to stress, with the result that words such as *liter*, *divan*, *politie*, *empirisch*, etc. are incorrectly considered to be exceptions, see [*Woordenlijst 1995*], p. 24. Secondly, the overview of the spelling reforms of 1954 has been replaced by an overview of spelling variants in loan words without generalizations about their distribution. No new rules were formulated for the spelling of loan words.

4.2.2 Advisory reports on spelling reform

We have seen that Te Winkel subdivided the lexicon in order to establish which loan words are candidates for reform. The same has been done by several committees that proposed spelling reforms of loan words in 1966–1969 and 1990–1994 (neither reform proposal was actually carried out, see Appendix H).

Pée et al. (1969)
In Pée et al., loan words were divided into native and hybrid words by formal linguistic criteria similar to those of Te Winkel: adaptation to the phoneme system, creation of derived forms with native affixes and development of meaning independent of the donor language.

   The spelling reform of Pée et al. can be characterized as subjecting hybrid words to the spelling rules for native words (except in the case of /i/; no change was proposed for words like *rivaal*). Some words were excluded from reform on the basis of formal properties, such as having a foreign phoneme or having more than one pronunciation. The spelling was also used to define words which should not undergo spelling reforms, for instance the spelling of words with /ɛ/ written as *è*, *ê*, *ei* or *ai* was not affected. Internationally used words, names etc., were also excluded from the reform.

Neijt & Zuidema (1994b)
In Neijt & Zuidema (1994b), loan words were divided into native, hybrid and foreign words.[6] The distinction between native words on the one hand and hybrid and foreign words on the other was made by criteria that refer to native phonemes, stress pattern and syllable structure and inflected forms. They also proposed restrictions on possible inflected forms and on the sequence of syllables in native words (the stem may contain one full vowel only). The distinction between hybrid and foreign words was made by a second set of less restrictive formal criteria. Hybrid words must contain native phonemes only, but need not have one full vowel, and have more possibilities with respect to syllable structure. Words that violate the second set of criteria are classified as foreign.

---

[6] In Geerts et al. (1988), a preliminary investigation of the spelling issue, no spelling rules were

Although the proposed new spelling rules hold for native as well as hybrid words (but crucially not for foreign words), some rules are formulated in such a way that they only affect hybrid words. An example is formed by the spelling of /ks/. Neijt & Zuidema convert these sounds to <x> in words with two full vowels and as <ks> elsewhere to avoid writing a native word such as *heks* as **hex*. This approach also implies that for all sounds except /i/, the default spelling is the same as in native words. In some cases, this implies formulating intricate rules. For instance, in order to ensure that <k> is the default spelling for the sound [k], Neijt & Zuidema (1994b) formulate the rule as follows: 'write [k] as <c> between a full vowel or [s] and [a], [o], [y] (*vacature*, *esculaap*), between a short vowel and [t] (*pact*) and in the prefixes *co-*, *con-*, *com-*, *col-*, *cor-* (*commissie*)'. However, although the 'native spelling' is always default, this spelling reform does not simply adapt hybrid words to native words, but improves the consistency of the spelling of hybrid words.

4.2.3 Descriptive accounts

Most descriptive accounts of Dutch spelling, e.g. Booij et al. (1979), Zonneveld (1980), Kerstens (1981), Wester (1985) and Booij (1985, 1995), do not pay much attention to the distinction between indigenous and loan words. They do not develop new criteria to distinguish native from non-native words (Booij et al. (1979) merely adopt criteria from Pée et al. 1969). Instead of spelling rules for non-native words, we only find observations such as the following: "one writes *s* except in those originally foreign words that are written with <c> in the original language.", see Booij et al. (1979:50). No rules are given to predict where we write <s> and <c>. Wester ignores the origin of a word when discussing the spelling of the sound [z], which is written differently in indigenous words and loan words. However, Wester only discusses stem final [z] in native words, e.g. *huizen* and internal [z] in loan words such as *mensa* and concludes that the different spelling of [z] is caused by the position in the morpheme. It is implicitly assumed that spelling alternations such as Doubling and Degemination are the same in indigenous words and loan words. For instance, Wester (1985) illustrates Vowel Degemination with native words like *apen* and loan words like *lianen*.

However, there is one study that focuses on the relevance of the origin of words for the formulation of spelling rules, namely Neijt (1994). This study tests the hypothesis that certain spelling variants are special for words of a given origin. If this hypothesis were correct, formulating rules for etymologically defined subgroups should diminish the possible spellings for a given phoneme. However, Neijt concludes that in the case of Latinate and Greek words this

---

given; only an inventory of spelling variants based on Horck et al. (1986) and Kempff et al. (1988).

approach hardly offers a more accurate description of the spelling patterns in use at the moment. If there are different spellings for a given phoneme, the spelling variations are not restricted to etymologically defined subgroups. For instance in words of Greek origin we find /t/ written as <t> and <th>, /k/ as <c> and <k>, /ɪ/ as <i> and <y>, etc. Neijt (1994) also observes that words are sometimes handed down to Dutch through some other language. Neijt (1994) therefore concludes that the etymological origin of words is a poor predictor for their spelling.[7]

4.3 Distinguishing indigenous words from loan words

The overview of the literature revealed that most accounts of Dutch spelling distinguish between indigenous words and loan words, and some even further divide the latter group into subgroups. These classifications are made by referring to etymological origin, to the extent to which words are adapted to the Dutch linguistic system, or to sound-spelling relationships.

In this study, the lexicon will be divided into subgroups in order to improve the description of Dutch spelling. Consequently, the first requirement for a useful division of the lexicon is that subgroups must only be distinguished if they comprise words that exhibit different spelling behaviour. However, if the division cannot be made on the basis of clear criteria, the difficulty is merely shifted from spelling computation to the classification of indigenous and loan words. Therefore, the second requirement is that the classification of words must follow from clear and applicable criteria. A third requirement, which follows from the aim to model the spelling rules used by speakers of Dutch, is that the classification can be made on the basis of information (consciously or unconsciously) available to native speakers.

I will now examine whether these requirements are satisfied by the classification of loan words by their origin, by the distinction of native and non-native words or by the classification of non-native words as hybrid or foreign. Classifying words by referring to sound-spelling relationships, as done in a few cases by Pée et al., would lead to circularity, since the classification is used to improve spelling rules. Consequently, this option will not be discussed here. Below, the terms native and non-native will be reserved for the distinction made by formal criteria, while the terms indigenous words and loan words will be used

---

[7] Te Winkel (1863:74–75) was aware of this phenomenon. He argues that it is not possible to write Greek words consistently according to the original Greek spelling (after transliteration), since many Greek words were adopted through Latin: we do not write *theokratie, *fysika, *filosofie, but *theocratie, physica, philosophie*. Te Winkel therefore decided to write words according to the spelling of the language they were immediately adopted from. This is the only way to account for those cases in which both the original and the changed form are used in Dutch, for instance both the Latin words *procurator* and *subject* and their French derivations *procureur* and *sujet*.

for the distinction on the basis of the origin of words.

Distinguishing loan words by their origin
The three requirements of a classification of loan words mentioned in above are not fulfilled by a classification on the basis of the etymological origin of words. Firstly, Te Winkel showed that absolute etymological origin is not a good marker of non-native spelling behaviour, since some words derived from foreign languages (e.g. *kelder*) have been totally adapted to the Dutch linguistic system. Secondly, Neijt (1994) showed that the spelling rules for (unadapted) loan words cannot be simplified by formulating distinct rules for words with different origins. Thirdly, the etymological origin of words is not normally known to native speakers. I will therefore not define loan words on the basis of their etymological origin.

Distinguishing loan words on the basis of formal criteria
Dividing the lexicon into native and non-native words on the basis of the extent to which they have been adapted to the Dutch linguistic system (given in full detail in Appendix B) fulfils the requirements that were formulated in the introduction.

Firstly, the formal classification criteria proposed by Te Winkel are clear and applicable. Van Heuven et al. (1994) formalized similar criteria and showed that with these criteria it is possible to automatically classify words as native or non-native (i.e. of Germanic or other origin according to an etymological dictionary) in 90% of the cases. There are two small sets of mismatches: totally adapted loan words such as *kelder* that are classified as native, and native words that are classified as non-native since they are frozen compounds or phrases (e.g. *deemoed*, *meneer*) and therefore violate the restrictions imposed on native morphemes.[8]

Secondly, the classification is made on the basis of data available to speakers of Dutch, which is supported by the fact that native and non-native words are treated differently in respects other than spelling: different phonological and morphological rules apply to both word types. This holds for word formation rules, see for instance Booij (1977:131–139). Some phonological rules also apply to non-native words only, for instance vowel shortening *ananas*-[ɑnanɑs], but *adelaar*-[adəlar-*ɑdəlar], see Booij (1995:36). The native/non-native distinction also plays a role in another way. Words composed with non-native affixes behave as if they were monomorphemic with respect to stress placement,

---

[8] The only criterion suggested by Te Winkel which is sometimes hard to apply, is that a native morpheme should not be combined with non-native affixes. This criterion is not used by Neijt & Zuidema (1994) or Heuven et al. (1994), and it will not be used here either, see Appendix B.

see complex *nor*<u>*meer*</u> (*norm+eer*) and monomorphemic *Mil<u>aan</u>*. This implies
that these affixes influence stress placement, while native affixes are stress-
neutral: <u>*hu*</u>*mor-*<u>*hu*</u>*morloos*, <u>*her*</u>*tog-*<u>*her*</u>*togdom*, see for instance Trommelen &
Zonneveld (1990). It appears that phonological and morphological rules also
distinguish between native and non-native words. This implies that the
distinction must be made in phonology and morphology anyway, and it can also
be used in the spelling system.

Te Winkel observed that the native/non-native distinction does not only
manifest itself by the application of different sets of phonological and
morphological rules, but also by the extent to which the underlying
representations satisfy structural conditions imposed on Dutch words. This
implies that native and non-native words need not be merely marked by diacritic
features, as was proposed by e.g. Chomsky (1968), Saciuk (1970) and Booij
(1977). We have seen above that Te Winkel claims that native words obey a set
of strict conditions as to their lexical representations that non-native words need
not satisfy, for instance native words may have one full vowel only. This
observation was also made in Zonneveld (1993).

The assumption that lexical representations determine whether a word is
native, is supported by the fact that abbreviations of native words undergo non-
native rules when they look like loan words: *buma* and *havo* are abbreviations of
***bu**rger**ma**n* and ***h**oger **a**lgemeen **v**ormend **o**nderwijs*, i.e. composed of native
words. Yet they are combined with non-native affixes: *bumatisch*, *havist* because
they have more than one full vowel just like loan words such as *struma* and
Bavo. Another indication is the fact that the suffix *-isch*, though of Germanic
origin, is mostly combined with non-native morphemes, probably because of its
non-native appearance. Like non-native affixes such as *-atie* it begins with a full
vowel, see Heynderickx & Van Marle (1994).

The last requirement imposed on a useful division of the lexicon, namely that
the classification can be used to improve the accuracy of spelling rules, is also
satisfied by the distinction between native and non-native words. Almost all
words whose spelling cannot be predicted by the spelling rules for native words
are non-native; there is only a small list of exceptions such as ***c**irkel*, see
Appendix B. These words are considered exceptional with respect to the
classification criteria only, but they are otherwise treated as other non-native
words. We will see in 4.4 that we can improve the accuracy of the description of
the Dutch spelling system with a separate set of phoneme-to-grapheme
conversion rules for non-native words.

I conclude that the native/non-native distinction meets the requirements
formulated above. A full account of the restrictions on native phonemes,
syllables, morphemes, stress patterns and inflection is given in Appendix B.

Hybrid versus foreign words

The criteria formulated above are not met by the further classification of non-native words into hybrid and foreign words. Firstly, it is not possible to formulate spelling rules for foreign words. This would be the case if we found that foreign words are subject to different spelling rules than hybrid words. An imaginary example would be: write /u/ as <oe> in native words, as <ou> in hybrid words and as <oo> in foreign words. However, I have found no indications that there are generalizations of this type. The distinction between hybrid and foreign words could be useful if the spelling of hybrid words is sometimes adapted, while that of foreign words is not (as proposed by Te Winkel). In 4.4 we will see that even words that contain foreign phonemes (the most obvious and undisputed indication of foreign words) are sometimes adapted. Words with foreign phonemes often keep their etymological spelling but this is a tendency, not a rule. Consider for instance the contrast between [kɔngɑ] *conga/*konga* that has a foreign phoneme and in which <c> is not changed to <k>, and [rɔːzə] *roze/*rose* that also has a foreign phoneme but where <s> was changed to <z> (I will return to this topic in 4.4). As observed by Te Winkel, the adaptation of the spelling of loan words is also determined by other factors than the extent of adaptation to the Dutch linguistic system only. It appears that the distinction of foreign and hybrid words cannot be exploited to improve spelling rules.

Secondly, although the criteria of Neijt & Zuidema (1994b) are clear and applicable, they draw the line between hybrid and foreign words for a specific purpose (i.e. in order to exclude foreign words from regularization), without independently motivating their division. Other divisions seem also possible.[9] In the case of the native/non-native distinction it was possible to isolate indigenous words on the basis of an etymological dictionary, make an inventory of their properties and then examine which loan words share these indigenous properties. In the case of the hybrid-foreign distinction there is no such way to define

---

[9] Booij (1983) and Neijt & Zuidema (1994), draw the line between hybrid and foreign words differently. Both Booij and Neijt & Zuidema require that hybrid words contain native phonemes only, but Neijt & Zuidema also consider words foreign when they show variation in the pronunciation, e.g. in *auto* where <au> is pronounced as [ɑu] or [o]. Booij (1983) requires that hybrid words obey all conditions on syllable structure of native words. Neijt & Zuidema assign hybrid words some extra possibilities. For instance, they allow hybrid words to contain certain clusters that violate sonority, e.g. initial [ft] as in *ftisis*, but exclude other clusters that comply with sonority, e.g. intervocalic [j] in words like *vanille*. Generally, they allow clusters that typically appear in words of Greek and Latin origin, but exclude clusters that appear in words of English or French origin. The authors also differ in what they consider hybrid and foreign morphological behaviour. Consider for instance plural inflection. According to Neijt & Zuidema, the plural affix *-s* in words which does not end in a schwa, /e/ or /i/ shows that a word is foreign, (again this mainly affects words of English or French origin such as *tip* and *cadeau*). Booij only considers plural endings other than -s or -*en* as foreign. Finally, there is no agreement on foreign stress behaviour. Neijt & Zuidema do not formulate conditions, whereas Booij assumes that final stress on syllables which are not superheavy marks words as foreign.

prototypical hybrid or foreign words.

Finally, there is no evidence that native speakers treat hybrid and foreign words differently with respect to phonological or morphological rules (if this were the case, we could exploit this difference to define hybrid and foreign words). Booij (1977, 1981) claimed that there is such evidence. Foreign words (or 'real loan words' as Booij calls them) undergo irregular inflection rules (e.g. *lemma-lemmata*) and irregular stress rules that assign final stress irrespective of syllable weight (*fon<u>due</u>*, *eufo<u>rie</u>* and *siga<u>ret</u>*). However, it is not clear whether we are dealing with rules here, since we cannot predict which words have an irregular stress pattern or inflected form. We might as well consider *lemmata* and *fondue* exceptions to (non-native) inflection and stress placement rules. Unlike what we have seen with non-native words, foreign lexical properties do not predict the application of foreign rules. For instance, the presence of a foreign phoneme does not single out words with irregular stress patterns, as illustrated by the contrast between [ɑrgo]-*argot* (foreign phoneme, irregular stress pattern) but [tɑngo]-*tango* (foreign phoneme, regular stress pattern). This suggests that we are not dealing with hybrid and foreign words that undergo different sets of phonological and morphological rules, but with non-native words only, some of which have idiosyncratic properties. This is to be expected; if we define native and hybrid words as words that are completely or partially adapted to the native system, it follows that foreign words are inherently outside the reach of Dutch linguistic rules.

Consequently, in this study the lexicon will only be divided into native and non-native words.

4.4 Phoneme-to-grapheme conversion rules for non-native words

Now that it has been established how the lexicon can be divided into native and non-native words, only non-native words will be discussed in this chapter.

If we survey the literature, we can conclude that there are two contrasting approaches to the spelling of non-native words. Te Winkel claims that these words are subject to the Principle of Etymology. A strict interpretation of this principle implies that the spelling of monomorphemic non-native words cannot be derived from their pronunciation by Dutch spelling rules. The spelling of non-native words is possibly derived by applying foreign spelling rules, and where this is not possible, spelling is simply arbitrary and must be looked up. Since the study by Neijt (1994) showed that Dutch spelling cannot be accurately predicted by foreign spelling rules, Te Winkel's approach implies that the spelling of non-native words cannot be computed but must be looked up; let us call this option

Option A: The spelling of non-native words is inherently irregular.

Linguists such as Booij (1985) and Wester (1987) on the other hand, who treat non-native words like native words, apparently assign more weight to the Phonological Principle, which states that the pronunciation determines the spelling, irrespective of the origin of the words in question. In their view we can derive the spelling of all non-native words by phoneme-to-grapheme conversion rules; we will call this option

> **Option B:**   The spelling of non-native words can be derived by
>   spelling rules.

The question is which approach leads to the most accurate description of the spelling of non-native words. In this section, I will argue that option B is to be preferred for three reasons.

An argument against option A is that non-native words that happen to be written in the same way as native Dutch words, for instance *pasta*, are still treated as exceptions, just like words with an idiosyncratic spelling such as **cynisch**. In option B, words such as *pasta* are automatically accounted for, and we need only store words such as *cynisch* as exceptions.

A second argument against option A is that if all non-native words are considered to be inherently outside the Dutch spelling system, one would expect that the spelling of non-native words is never (partially) adapted to the native pattern; regularization of these words can never take place. However, such spelling adaptations have occurred both before and after the formulation of official spelling regulations. Example (2a) lists some words cited by Te Winkel as examples of words that had already been adapted before he formulated his rules (Te Winkel 1865:178), and (2b) gives examples of words of which the spelling was adapted in [*Woordenlijst 1954*]. All examples are of French origin, see Van der Sijs (1996):

(2)   **Spelling changes**
      a   'old'
          mel**on**   → mel**oen**
          cigare   → **s**igaar
          républ**ique**   → republ**iek**
      b   'new'
          l**oupe**   → l**oep**
          entrée   → entr**ee**
          mi**lli**ard   → mi**lj**ard

Such spelling adaptations argue in favour of option B.

The most important reason to reject option A in favour of option B is the

existence of regularity within the spelling of non-native words. The existence of such patterns in Dutch non-native words will be shown in 4.4.1 and 4.4.2. There are different types of generalizations about the spelling of non-native words, all of which can be illustrated by the rules for the sound /k/ proposed in this study:

(3)   **Phoneme-to-grapheme conversion rules for /k/**

a   /ks/ → <ct> / _ in certain /i/-initial suffixes   (*functioneer*)

b   /ks/ → <cc> / + ɑ _ [−back]   (*accent*)[10]

c   /ks/ → <x> elsewhere   (*max*)

d   /kw/ → <kw> / _ …{-un, -ɛin, -ir, oj, aj}   (*kwartier*)

e   /kw/ → <qu> elsewhere   (*quasi*)

f   /k/ → <k> / _ (…k) morpheme-finally   (*kaduuk*)

g   /k/ → <k> / _ (…k) [−back]   (*spektakel*)

h   /k/ → <k> / _ …{-un,-ɛin, -ir, oj, aj}   (*kalkoen*)
     CONDITION: not in [-ylir] or [-yrir]   (*particulier*)

i   /k/ → <c> in words with foreign phonemes   (*courgette*)

j   /k/ → <k> / _…œy   (*kazuifel*)

k   /k/ → <c> elsewhere   (*cadeau*)

The most important type of generalization is that etymological spellings are adapted to the native pattern word-finally, see rule (3f). A second type of generalizations captures predominant etymological patterns. Rules (3a, b, c, e, g) are of this type. Finally, there are many exceptions to predominant etymological patterns. In some cases we can predict such exceptions on the basis of the following three tendencies that can be observed in the spelling of non-native words: spelling tends to be adapted to the native pattern in certain contexts, as captured in rules (3d, f, h, j), spelling tends to remain unchanged in words with foreign phonemes, see rule (3i), and the spelling of a sound tends to be consistent within morphemes. The latter tendency is captured by '…' in (3d, f, h). For instance, rule (3g) states that /k/ is written as <k> before a front vowel (*spektakel*), or in a word with a front vowel (*spektakel*). Similar rule sets can be formulated for other sounds as well.

    The rules that capture hybrid patterns and the other generalizations will be discussed in 4.4.1 and 4.4.2, respectively.

4.4.1 Spelling rules that capture non-native patterns

In this section, I will argue that in some cases a combination of etymological

---

[10] Rule (3b) only accounts for words beginning with *acc-*).Examples which are not captured by this rule, *succes, vaccin* and *occident*, are listed as exceptions since they are rare and because there are also words which do not follow this pattern such as *Luxemburg, taxi* and *oxide.*

spelling patterns and spelling adjustment led to a new hybrid pattern (a 'bastaardspelling'), which can be described by special spelling rules for non-native words.

The literature does not yield such generalizations about the spelling of non-native words; strictly speaking, Neijt & Zuidema (1994b) propose one set of spelling rules for hybrid and native words, although these rules are formulated in such a way that they have different effects on native and hybrid words. I will argue that there are similar generalizations for /z/, /k/ and long vowels.

The spelling of /i/
[*Woordenlijst 1954*] contains one spelling rule for non-native words, namely the rule for the spelling of /i/ (I will give the older rules here since the reformulation in [*Woordenlijst 1995*] is incorrect, see 4.2):

(4)   **The spelling of /i/ according to [*Woordenlijst 1954*]**
    i   *ie* is written in word final syllables, except in Latinate month names, *-isch* and foreign words.
    ii A syllable with the sound /i/ that precedes a native derivational or inflectional suffix, is treated like a final syllable, unless it ends in an unstressed *ie* followed by a schwa, written as *e*.
    iii in all other cases *i* is written.[11]

Subrule ii of (4) is in conflict with the framework developed in chapter 2, in which all other phoneme-to-grapheme conversion rules are restricted to the morpheme domain. However, it is possible to reformulate (4) so that it fits the framework. In this version alternations that refer to context outside the morpheme are accounted for by autonomous spelling rules:

(5)   **Phoneme-to-grapheme conversion rules for /i/**
    a   write /i/ as \ie\ in the last syllable of non-native morphemes (*balie*, *actief*, *artiest*) except in Latinate month names, prefixes, *-isch* and *taxi*, *alibi*, *fis*, etc.
    b   write /i/ as \i\ elsewhere (*limonade*, *vitaal*)

    Autonomous spelling rules
    c   change \ie\ into <i> in an unstressed syllable before a vowel (*oliën*, *neuriën*)

---

[11] "Afgezien van de gevallen waarin *y* wordt geschreven gelden voor de spelling van de klank *ie* de volgende regels: 1. In eindlettergrepen van woorden [...] wordt *ie* geschreven, behalve in de Latijnse maandnamen, in het achtervoegsel *-isch* en in woorden die als vreemd worden beschouwd. 2. Vóór Nederlandse afleidingsachtervoegsels en buigingsuitgangen wordt een lettergreep met de klank *ie* behandeld als eindlettergreep volgens regel 1, behalve wanneer ze eindigt op een niet-beklemtoonde *ie* die door een als *e* geschreven sjwa wordt gevolgd. 3. In alle andere gevallen wordt i geschreven."

    d   change morpheme final \i\ into <ie> before a consonant-initial native
        suffix (*taxiede*) except before the plural or genitive suffix -*s* where an
        apostrophe is added: *ski's*

I will return to (5c) and (5d) in chapter 5.

    Note that a generalization is missed in (4), since rules (4i) and (4ii) both refer
to the final syllable. The alternative formulation in (5) avoids the duplication of
the old rule. The reformulated rules also make more accurate predictions about
the type of exceptions that can be expected. With the original rules the difference
between idiosyncratic *taxi* and regular *taxiede* is just a coincidence; the spelling
of the word in isolation and the spelling of the inflected form are the result of
separate rules. It just happens that *taxi* is exceptional. The reverse is just as
likely; a regular word in isolation and an idiosyncratic inflected form, e.g.
*taxie-*taxide. The new version, on the other hand, predicts that the reverse is
marked.

    Rule (5) does not simply derive the original etymological pattern. Instead a
new hybrid pattern has evolved. The 'native' spelling <ie> sometimes already
occurred word-finally in the original spelling (*officier*, *genie*, *categorie*, cf.
French *officier*, *génie*, *catégorie*) and <i> was almost exclusively adjusted to
<ie> word-finally (*publique* → *publiek*, *plaisir* → *plezier*). Thus an asymmetry
arose: <ie> in the word final syllable, <i> elsewhere. The same pattern, adapted
spelling in the last syllable, etymological spelling elsewhere, holds for pairs such
as *synoniem ~ synonymie*, *stereotiep ~ stereotypie*.

    Apparent exceptions to (5b) are *johannieter*, *karmelieter*, *dansmarieke*,
petieter and *sodemieter* in which a native suffix -*er* or -*ke* can be found, and
dixieland, *iezegrim* and *faliekant* that are lexicalized compounds. The same
could hold for *fieselemie* that is not listed in [*Woordenlijst 1995*]. In other
exceptions <ie> is part of the etymological spelling: *sowieso*, *spielerei*,
spielmacher and *retrieval* that are not listed in [*Woordenlijst 1995*]. Exceptions
that can not be explained this way are *flottielje*, *mediene*, *trielje* from *flotille*,
medina and *trille* (and *gienje* (*guinje*) and *vlieseline* that are not listed in
[*Woordenlijst 1995*]). These facts suggest that -*e* and -*je* are interpreted as a
native suffix.

    If we ignored the generalization captured in (5b) and extended the rule for
native words to non-native words, we would not only have to consider a
relatively small group of words such as *taxi* as exceptional, but also the large
class of words such as *liter* and *motivator*. Of the non-native words with /i/ of
the CELEX-database that are labelled monomorphemic, 58% are written with
<i>. In 86% of the cases, this <i> occurs in a non-final syllable, and in 14% in a
final syllable. Rule (5b) thus accounts for the larger group. Although the
generalization is not exceptionless, it is strong enough to influence the new
spelling of words that were adapted in 1954:

(6) **1863**   **1945**   **1863**   **1945**
    anon**y**miteit   anonimiteit/*anoniemieteit   **cy**linder   cilinder/ *cielinder
    t**y**ran   tiran/*tieran   as**y**l   asiel/*asil

The spelling of /z/

A second example of a sound that is written differently in native and non-native words is /z/. This sound is always written as <z> in native words, but in non-native words we sometimes write <s>. Again, the distribution of the alternatives is partly predictable: we write <z> morpheme-finally, e.g. in *precieze* (*preciez+e*) and retain the etymological spelling <s> elsewhere, for instance in *precisie* or *isolatie* (given that the latter words are monomorphemic, see 4.5). The rules for the spelling of /z/ in native and non-native words are given under (7):

(7)   **The spelling of /z/ after a long vowel or sonorant consonant**
          write /z/   i   as <z> morpheme-finally
                         ii   as <s> elsewhere

The contrast between <s> and <z> is neutralized word-finally by Spelling Devoicing (see 3.4.1), so a word such as *advies* is not a counterexample to (7). The different underlying spelling only manifests itself in inflected forms: *adviezen* instead of **adviesen*.

If we tried to describe the lexicon as a whole by the rule 'write /z/ as <z>', we would ignore the fact that the occurrence of <s> is largely predictable. The number of exceptions would also increase. The non-native rule has some exceptions such as *ruzie* and *plezier*, so if we extended the native rule to non-native words, these words would become regular. However, in that case the much larger class of words such as *isolatie*, *presenteren* and *polarisatie* would become exceptional.

The rules proposed in Neijt & Zuidema (1994b:45–47) prescribe that /z/ (between a long vowel and a vowel) is written as <s> after *(p)re-* and before non-native suffixes such as *-eer*, *-isch*, *-iseer*, etc., and as <z> elsewhere. This rule was meant to reform the spelling of words such as *pauseer*, so it does not describe the current spelling of these words. On the one hand there are examples such as *pauzeer*, *vizioen*, *plezant*, *gazeuse*, *muzisch* and *jaloezie* where <z> occurs before non-native suffixes and on the other hand there are words such as *ba**s**alt*, *i**s**olatie*, *qua**s**i* and *the**s**aurus* where /z/ is written as <s> although it does not precede a suffix. For this reason, the generalization about the distribution of <s> and <z> is not adopted here.

The spelling of /k/

The spelling of the sound /k/ is notoriously unpredictable, see for instance Geerts et al. (1988), De Vriendt (1994b). However, even in this case there is one position where the spelling has been almost always adapted, i.e. morpheme-finally (*publiek*, *bruusk* etc.). In all other cases, the spelling is etymological (this is represented as <c> only):

(8)   **Some generalizations about the spelling of /k/**
        write /k/   i   as <k> morpheme-finally
                       ii   as <c> elsewhere

This rule again reduces the number of exceptions; words such as *publiek*, *augurk*, *obelisk*, *barak* etc., which constitute the majority of words with final /k/ are now regular. The only exception to (8i) is formed by a small group of words such as *chic*, *plaque* and *truck* that would also be exceptional if we used the native rules.

The spelling of long vowels
So far, I have shown that some spelling differences between native and non-native words are predictable by separate, non-native rules. However, even if the spelling of native and non-native words is almost the same, closer investigation of so-called exceptions could reveal two different spelling systems in native and non-native words. This is the case with the spelling of long vowels.

    The spelling of long vowels in non-native words is not always similar to that in native words, which is illustrated in (9):

(9) **The spelling of long vowels in native and non-native words**

		closed syllable	**open syllable**
native words		*raam*, *vaandel*	*kamer*, *ja*, *zee*
non-native words	a	*kanaal*, *bruuskeer*	*boleet*, *hallo*, *trochee*
	b	*tenor*, *luxe*	*cheeta*, *shampoo*

The non-native words in (9a) behave like indigenous words. Long vowels are written as geminate letters in closed syllables, and as single letters in open syllables, but there also are irregular non-native words, see (9b). In *tenor* and *luxe* a long vowel is written as a single letter although it occurs in a closed syllable, while some geminates occur in open syllables, e.g. in *shampoo* and *bazooka*. Until now, the rules for the alternations of double and single vowel letters (see the references in 4.2.3) were meant for non-native words as well as native words. This implies that all examples of (9b) are simply considered to be exceptions. If we excluded these exceptional words, the spelling of vowel letters seems to be the same as in native words:

(10)   geminate: *kanaal*, *boleet*, *suppoost*, *dressuur*
            single letter: *ahorn*, *thema*, *boleet*, *puree*

This suggests that the spelling of long vowels is derived in the same way as in native words and that single letters are the result of Degemination. However, in (10) we also observe a spelling difference between final and non-final syllables, which is reminiscent of the spelling of /i/. Just as there is a contrast between final <ie> and non-final <i>, there also is a contrast between final vowel geminates and non-final single vowel letters. This means that the spelling of long vowels in non-native words can also be accounted for by rules similar to the ones in (11): write geminates in the last syllable and single vowels elsewhere. The possible derivations differ but the result is the same:

(11)   **Two alternative accounts for the spelling of long vowel**
           Option 1:
           /a/ → <aa>
           /kanal/→ \kaanaal\ → <kanaal>
       cf. /hamər/   → \haamer\ → <hamer>

           Option 2:
           /a/ → <aa> / _ in the last syllable
           /a/ → <a> / _ elsewhere
           /kanal/→ kanaal
       cf. /sipir/   → cipier

Support for underlying single letters representing long vowels is found in the fact that the vowel letter <y> is never doubled when it denotes a long vowel: *python*, *asyl*, but this is not conclusive. In order to be able to choose between the two options, we must find other, less idiosyncratic words where the two analyses make different predictions. The spelling in contexts in which Vowel Degemination is not applicable is decisive in this respect. In native words Vowel Degemination does not apply in closed syllables (*koord*) or before the digraph *ch* (*loochen*), see 2.4. If non-native words must be analysed the same way, we would expect vowel geminates in the same contexts cf. the examples in (12a), but instead of geminates we find single vowel letters here, as illustrated in (12b):

(12)   **closed syllables**   **digraphs**
    a   *na**aa**ste/*n**a**ste*   *g**oo**chelen/*g**o**chelen*
         *me**ee**ster/*m**e**ster*
         *r**oo**ster/*r**o**ster*
    b   **mas**oo**chist/ mas**o**chist*   **m**ee**chanisch /m**e**chanisch*
         Guu**stave/G**u**stave*
         luu**xe /l**u**xe*   **st**oo**chast /st**o**chastisch*
         concuu**rrent /conc**u**rrent*   **cap**uu**chon /cap**u**chon*

The single letters in (12b) cannot be the result of the rule Vowel Degemination that holds for native words, which suggests that the spelling of long vowels is different in final and non-final syllables, just like the spelling of /i/. Thus, in the few cases where the two analyses make different predictions, the facts are in accordance with the second analysis. I therefore propose to account for the spelling of long vowels by means of the following rules:

(13)   **The spelling of long vowels in non-native words**
    a   write /a/, /e/, /o/ and /y/ as <aa>, <ee>, <oo> and <uu> in the last
         syllable of a morpheme
    b   write /a/, /e/, /o/ and /y/ as <a>, <e>, <o> and <u> elsewhere

Of course, we still need Vowel Degemination for inflected words such as *kan**a**len*. The separate non-native generalization also holds for the following words that are no longer formally recognizable as non-native words: *Rachel*, *p**o**ster* and *b**u**ste*. There are some exceptions to (13): the derivations *bastaarderen*, *bruuskeren*, *waarderen*, *standaardiseren*, *neerlandistiek* (these will be further discussed in 4.5), *vaandrig*, *maarschalk* that seem complex and *staatsie* and *tave(e)rne*.

The surprising conclusion is that long vowels are subject to different phoneme-to-grapheme conversion rules in native and non-native words. Long

vowels are always converted to geminates in native words, but to single letters in non-native words, except word-finally. The new analysis of vowel alternations in non-native words confirms the need for underlying vowel geminates in indigenous words. We need vowel geminates to distinguish regular native words from regular non-native words: *rooster-poster*, *loochen-masochist*, etc. Note that the representation of such contrasts would be problematic in a system such as that of Zonneveld (1980) in which all long vowels are represented by underlying single vowel letters. In that case non-native words must somehow be prevented from undergoing vowel doubling.

Summarizing, we found hybrid spelling patterns for /i/, /z/, /k/ and long vowels in non-native words. In most cases these patterns co-occur, i.e., whenever one of these generalizations holds, all other generalizations hold as well. This is illustrated in (14):

(14)   cad**uc** → kad**uuk**   *kad**uk**/*cad**uuc**
        exqu**is**e   *exqu**iz**e/*exqu**ies**e
        ch**ic**   *ch**iec**/*ch**ik**

Exceptions are words ending in *-uus* (*abstruus*, *confuus*, *diffuus*, *excuus*), in which <u> was changed to <uu> while /z/ is still written as <s> in inflected forms: *abstruse*, *confuse*, *diffuse*, *excuses*, not *confuze* etc. All other words are regular in this respect.

The spelling of non-native words as described above is neither completely adapted to the Dutch pattern (<ie>, vowel geminates, <k>, or <z> in all positions), nor completely etymological. Instead a new hybrid pattern has evolved. The generalization is that the spelling is adapted to the native system in final syllables and has remained unchanged elsewhere. Note that this implies that the default rule does not generate the native spelling, but the predominant etymological spelling. In this respect the rules for /k/ and /z/ differ from the rules in Neijt & Zuidema (1994b:41, 45).

The difference between final and non-final syllables is understandable from a diachronic point of view. As noted by Te Winkel, spelling changes are induced by partial sound changes that often take place word-finally. Therefore, spelling changes are more likely to take place word-finally as well. Another factor that may have influenced the asymmetry is the fact that a native spelling is often more important word-finally, to prevent inflected forms and derivations from becoming ambiguous. The original spelling sometimes fails in this respect (note that the actual spelling violates the Morphological Principle):

(15)   truc-tru**c**endoos   (<c> suggests [s])   → tru**k**endoos
        chic-chi**c**e   (<c> suggests [s])   → chi**qu**e

This could explain why word final /k/ is mostly written as <k> whereas /ks/ is often still written as <x> (*affix*, *lux*, etc.). <x> is not problematic in inflected forms. Of course we have seen that within words, there are some ambiguous sequences as well, sometimes even unresolved, e.g. in *pistool*, *milliade*. In the case of monomorphemic words we might learn the irregular letter-sound relationship, but this is not possible for all derived forms. Changing spelling in the last syllable can be seen as a compromise between the Phonological Principle and the Principle of Etymology, since it restricts the possible adaptations both in number and in position, while maximally ensuring a readable result by first applying adaptations in those positions where endings might be added.

4.4.2 Other generalizations

Until now, I have only discussed phoneme-to-grapheme conversion rules that are partly the result of spelling adjustments. However, there are also many cases where there has been no systematic spelling adjustment. We can predict the spelling with even less accuracy in these cases, since the spelling may still be the same as that in the donor language, or it may have been partially or entirely adapted to the Dutch spelling rules. For instance, no exceptionless rules can be given for the choice between spelling variants such as *th* or *t*, *ch* or *sj*, *ou* or *oe*, etc. Still, it is often possible to formulate rules that capture the predominant etymological spelling (which sometimes depends on the context).

Predominant etymological spelling patterns
Generalizations about predominant etymological spelling patterns formed the basis of the reform proposal in Neijt & Zuidema (1994b). However, the rules formulated in the present study are different in two respects: they are meant to describe the current spelling rather than prescribe a spelling reform, and they apply to non-native (i.e., hybrid and foreign) words, instead of to native and hybrid words. This implies that the rules formulated here are more elaborate, since they are also meant to capture minor generalizations (such as the fact that we write <cc> in words such as *accent*, see rule (3b) above). Another implication is that the default spelling is often not the same as in the native lexicon (for instance, the default spelling of /ks/, /kw/ and /k/ is <x>, <qu> and <c>, see the rules (3c), (3e) and (3k) above). This approach also implies that rules for foreign phonemes have to be formulated if possible. An example is formed by the following generalization about the spelling of /g/. This sound is written as <gu> before front vowels (e.g. *san**gu**ine*) and as <g> elsewhere (e.g. ***g**oal*).[12]

---

[12] However, no rules were formulated for some foreign phonemes because they do not occur in the transcription provided by CELEX, or only sporadically, or because their spelling is too idiosyncratic.

I will not discuss all the rules that capture predominant etymological patterns here. They are given in Appendix D, and some aspects of the spelling of /(t)s/, /z/ and /k/ are discussed in Appendix G. However, I will discuss one case where the spelling of the same sound, the diphthong /ɑu/, is different in native and non-native words. In native words /ɑu/ has the default spelling <ou>, and <au> is only written in a number of etymologically motivated exceptions. However, most non-native words are written with <au>, and part of the exceptions are predictable. Before the dental clusters /nt, nd, nz, ns, tsj/ we write <ou>: *compound*, *ounce* etc. A few exceptions remain, e.g. *fout*, *kabouter*.[13] Other exceptions are *down*, *cacao* that are also exceptional if <ou> is considered the default spelling. All other words are written with <au>: *causaal*, *rabauw*, *aula* etc.

(16)   a   /ɑu/ → <auw> / _ morpheme-finally   e.g. *kabeljauw*
       b   /ɑu/ → <ou> / _ n [+cor, -son]   e.g. *sound*
       c   /ɑu/ → <au> / _ C elsewhere   e.g. *laurier*

The different default spellings for the same sound, <ou> in native words, <au> in non-native ones, argues in favour of separate spelling rules for non-native words.

Although I have stressed that we need to formulate distinct sets of spelling rules for native and non-native words, this does not imply that there is no overlap. In some cases, spelling rules are the same as in the native lexicon, the only difference being that non-native words are more often exceptional. This holds for the sounds /p, b, d, f, v, x, ɣ, m, n, l, r, h, ɪ, ɔ, œ, ∅, ɛi, ə/. In other cases, the generalizations made on the basis of native words are still valid, but supplemented by new rules, for instance the rule 'write /j/ after a vowel as <i> morpheme-finally and as <j> elsewhere' for native words is supplemented with the rule 'write /j/ as <i> before non-native affixes', cf. for instance *koeioneer* versus *bajonet*.

Tendencies
Some of the generalizations made so far have many exceptions. For instance, of the non-native monomorphemic words with the sound /u/, 44% are written as <oe>, 37% as <ou> and 19% have yet another spelling. Even if we regard <oe>

For instance, the transcriptions provided by CELEX never contain the phonemes /uː/ or /ɑː/ (*rouge* and *basket* are transcribed as /ruʒə/ and /bɑskət/). Other phonemes occur sporadicly. /yː/ and /iː/, for instance only occur in *centrifuge* and words with *-lyse*, such as *analyse*. Even if a sound occurs more frequently, there often is not much regularity to be found. For instance, there is not much regularity in the spelling of the sound /ɔː/ in *roze*, *crawl*, *board*, etc.

[13] Words such as *kabouter* seem to suggest that we write <ou> before single dental consonants as well, but cf. *baud*, *fauna*, *pauzeer*, *auto* etc.

as the default spelling, the rule derives incorrect spellings in many cases. Therefore it is useful to be able to predict where the spelling will deviate from the predominant pattern. There are three observations that we can exploit for this purpose: firstly, spelling tends to be adapted in words with certain suffixes. Secondly, spelling tends not to be adapted in words with foreign phonemes and finally, all occurrences of a phoneme in a word tend to be either adapted or unaffected.

Firstly, the spelling of certain words with certain semi-suffixes or sound sequences is often adapted. The relevant endings are /ir/, /ɛin/, /un/ and the sequences /oj/, /uj/, /aj/. This can be explained by the fact that these endings are adapted versions of the French affixes *-ier* (/je/), *-on* and *-in(e* or *-aine*, and sequences such as <oi> (/wɑ/) and <ai> (/ɛ/). Consider for instance pairs such as the following:

(17)   **Pronunciation adapted**   **Pronunciation not adapted**
       *k*ohier   [ko**hir**]   *c*ahier   [kɑ**je**]
       *pavil*j*oen*   [pavəl**jun**]   *carri**ll**on* [kɑrɪl**jɔn**]
       rozijn   [roz**ɛin**]   *limousine*   [limu**zinə**]
       *k*onvooi   [kɔn**voj**]   *couloir*   [kul**war**]

Some adapted examples are given under (18):

(18)   **k**oerier, **k**ohier, b**oek**anier, ple**z**ier, tre**z**orier; ro**z**ijn, bi**lj**oen, bla**z**oen;
       kalk**oen, **k**ampioen, pavil**j**oen, ja**k**obijn, **k**onijn, tamb**oe**rijn; **k**abaai,
        konvooi, t**oe**rnooi

For sounds whose predominant spelling is an etymological spelling, namely /k/ and /z/ we can exploit this observation in rules such as 'write /k/ as <k> in words with certain semi-suffixes', cf. rule (3d) and (3h) above.

The second tendency is that (as has been mentioned in 4.4.2) words with a foreign sound (combination) often keep their etymological spelling. This does not only hold for the foreign sound (which is not surprising, since there are no spelling conventions for foreign phonemes) but also in the rest of the word:

(19)   /g/   congé, acajou, courgette, rouge
       /zj/   pinguïn, argot, gourmet, conga
       /ɛː, ɔː, ɪː/   crème, colbert, analyse, controle, enquête
       nasal vowels   raison, sousmain, mannequin, mocassin

Rule (3i) is based on this tendency. However, there are also a few exceptions (even if we ignore non-preferred variants from [*Woordenlijst 1954*] such as *k*olbert [kɔlbɛːr], *k*ontrole [kɔntroːlə], *k*omfort [kɔmfɔːr], *pingoeïn* [pɪŋgwɪn]).

The words in (20a) all have foreign sounds but the spelling is partially adapted:

(20)a	ele**k**tricien	[elɛktrisjɛ̃]
	katalyse	[kɑtalɪːzə]
	mayo**n**aise (not: *mayo**nn**aise)	[mɑjonɛːzə]
	f**oe**rage	[furazjə]
b	ro**z**e	[rɔːzə]
	hang**a(a)**r	[hɑngɑr]
	plafon(**d**)	[plɑfɔ̂]
	f**oe**rgon	[furgɔn]
	degout	[degu]
	genie	[zjəni]
	log**ee**	[lozje]
	negligé	[neglizje]

It could be argued that the words in (20a) are no real counterexamples, because the foreign phoneme is part of another morpheme than the sound of which the spelling is adapted (*kata+lyse*), but the words in (20b) cannot be explained this way.

The final tendency is the consistent spelling of the same sound. In 4.4 it was already observed that if a phoneme is adapted, other occurrences of the same phoneme in the same morpheme are adapted as well (of course, this does not hold in the case of context-sensitive phoneme-to-grapheme conversion rules, cf. *limiet, synoniem*; *desastreuze*).[14] More examples are given under (21), where the spelling variants which were allowed until 1995 are given in parentheses:

(21)   **Consistent spelling of the same sound**
   a   **k**linie**k**, **k**anunni**k**, **k**oza**k**, **k**aduu**k** (**c**adu**c**) etc.
   b   **c**a**c**ao, **c**as**c**ara, **c**loa**c**a, **c**on**c**aaf, ro**c**o**c**o, etc.
   c   **k**a**k**adoris, **k**al**k**oen, **k**ami**k**aze, **k**ara**k**ter, **k**o**k**arde, **k**o**k**os, **k**om**k**ommer, **k**ro**k**odil, **k**ur**k**uma, etc.
   d   **c**a**c**tus, (**k**a**k**tus), **c**ara**c**ole (**k**ara**k**ol), **c**aval**c**ade (**k**aval**k**ade), **k**leri**k**aal (**c**leri**c**aal), **c**on**c**urrent (**k**on**k**urrent), re**c**tifi**c**atie (re**k**tifi**k**atie) etc.

The examples in (21) suggest that the spelling of the same sound is consistent. There are also exceptions, some of which are listed in (22):

---

[14] The adaptation of the spelling of one sound in a word does not imply the adjustment of the spelling of other sounds. This can be illustrated by the fact that in 1954 the words *catheder, chrysanth* and *cylinder* were changed into *katheder, chrysant* and *cilinder*, and not into **kateder, *chrisant* and **silinder*.

(22)   **Variable spelling of the same sound**
       pina**c**othee**k**, o**z**oni**s**eren, **c**lownes**k**, a**c**robatie**k**, ga**z**eu**s**e, ele**k**trifi**c**atie,
       kineti**c**a, **k**walifi**c**atie, e**k**wino**c**tiaal, ma**c**ro**k**osmos, mi**c**ro**k**osmos,
       s**k**epti**c**us, ele**k**troni**c**a, **k**osmeti**c**a

The spelling of the words in (22) suggests that the consistent spelling of a
phoneme holds for morphemes rather than words (although in some cases
allowed variants contradict this). However, not all morphemes are treated the
same way. *Con-* does not form a separate domain (the only exception is
conventikel where the <k> is needed to prevent an incorrect pronunciation), but
micro- and *macro-* and affixes such as *-icus*, *-ica*, *-icatie*, *-iseren* and *-esk* do.
Sequences such as *elektr-* and *-catie* are always written the same way. I conclude
that *micro-*, *macro-*, *elektr-*, *-icus*, *-ica*, *-esk*, *-iseren* and *-icatie* form separate
domains with respect to this generalization, so only *gazeuse* is a real exception.
[15]

    The rules discussed here are of a different nature than those discussed in the
preceding section. The rules in 4.4.1 form a new regularity that is taken into
account in further spelling adjustments, but in the case of the rules of this section
this is less clear. *Cylinder* became *cilinder* according to the new hybrid pattern,
not **cielinder* as predicted by native rules, but *accent* was changed in the
direction of native words (in the allowed spelling variant): *aksent* not **axent*.
Non-native words such as *karbau* and the inflected form of *Nassau* became
karbouw and *Nassouwe*, (see 2.4). These facts suggest that the rules in 4.4.2,
contrary to those in 4.4.1, cannot be used to predict the spelling of new words.

4.5 The spelling of complex non-native words

In this section, the question will be addressed how the phoneme-to-grapheme
conversion rules proposed in 4.4 apply to complex words. In Chapter 3 we have
seen that native phoneme-to-grapheme conversion rules apply to morphemes
rather than words. Context-sensitive phoneme-to-grapheme conversion rules
ignore the context outside the morpheme, cf. for instance *koud ~ trouw* but
getrouwd/*getroud. On the basis of these facts we expect that non-native
phoneme-to-grapheme conversion rules apply to morphemes as well. In order to
find out if this is indeed the case, we must look at context-sensitive rules,
especially rules that are sensitive to the distance from the word edge, and
examine whether they take the context outside the morpheme into account. In

---

[15] The contrast between *con-* and *micro-* suggests that the productivity of the affixes determines
whether it forms a separate domain. We do not expect alternations such as *kassa-microcassa* (the
alternation *kosmos-microcosmos* was abolished in 1995). However, suffixes such as *-ica* are not
productive but still have a consistent spelling.

4.4.1, we have seen that the spelling of /j/, long vowels, /z/ and /k/ is described by phoneme-to-grapheme conversion rules that are sensitive to the distance from the morpheme boundary. The behaviour of non-native derivations with respect to these rules is thus decisive. [16]

First let us look at combinations of non-native stems with native suffixes. The examples in (23) show that the context outside the morpheme is ignored:

(23)		**phoneme**	**stem**	**derivation**
		/i/	lim**ie**t	lim**ie**ten
		/e/	troch**ee**	troch**ee**ën
		/z/	precie**s**[17]	precie**ze**
		/k/	klinie**k**	klinie**ke**n
		/j/	konvoo**i**	konvoo**i**en

The uniform spelling of the morphemes in (23) shows that words composed of non-native stems and native affixes behave the same way as native derivations.

However, words composed of a non-native stem and a non-native suffix present a different picture:

(24)		**phoneme**	**stem**	**derivation**
	change:	/i/	lim**ie**t	lim**i**teer
		/e/	troch**ee**	troch**eï**sch
		/z/	precie**s**	preci**s**ie
		/k/	klinie**k**	klini**c**us
	no change:	/j/	konvoo**i**	konvoo**i**eren[18]

The spelling of non-native words with non-native suffixes in (24) is less straightforward than that of non-native words with native affixes. The first four examples in (24) suggest that non-native morphemes do not form domains for phoneme-to-grapheme conversion, since the context outside the morpheme is not

---

[16] In the case of /j/ and long vowels, the rules make predictions about final and non-final positions (as observed for /e/ and /i/ in [*Woordenlijst 1954*], p. XLVIII-XLIX). In the case of the sounds /k/ and /z/, there is only a strong generalization about the spelling in morpheme-final position; in all other positions the spelling is less predictable. This implies that there is no rule which requires replacement of <k> by <c> or <z> by <s> when an affix is added to a stem which ends in /k/ or /z/. This accounts for cases such as *muziek ~ muzikaal, fabriek ~ fabrikant, markies ~ markizaat, studentikoos ~ studentikoziteit, leprozen ~ leprozerie*.

[17] The underlying spelling of *precies* is \preciez\; the <s> is the result of Spelling Devoicing, see 3.4.

[18] A similar relation could hold for *akkoord-accorderen*, if we interpret that fact that the derived form is pronounced with a short vowel as a spelling pronunciation.

ignored, but pairs such as *konvooi-konvooieren* point in the opposite direction. However, there are only three words of the latter type; the other examples are *renvooi-renvooieren* and *octrooi-octrooieren* (and until 1995 *essaai-essaaieren*). To account for the facts in (24), I will therefore claim that combinations of non-native stems and non-native suffixes are treated as single morphemes by phoneme-to-grapheme conversion rules.[19] Words of the type *konvooieer* can be seen exceptions, or we could consider the uniform spelling of these morphemes to be accidental since /j/ is generally converted to <i> before the suffix *-eren*, cf. *associëren*, *discussiëren*, *differentiëren*.

Now consider the following pairs that are composed of native stems (*fee* and *waarde* have only one full vowel, the others have more than one but *neerland* is a compound and *bastaard* and *standaard* contain the suffix *-aard*) and non-native suffixes:

(25) **Stem**   **Derivation**
     bast**aa**rd   bast**aa**rderen
     f**ee**   f**ee**ërie(k)
     n**ee**rland   n**ee**rlandistiek, n**ee**rlandicus
     stand**aa**rd   stand**aa**rdiseren
     w**aa**rde   w**aa**rderen

These pairs show that native morphemes form spelling domains whether combined with native or non-native affixes. This leads to the following overview of uniform or variable spelling in complex words:

(26)		**native suffix**	**non-native suffix**
	native stem	d**ie**r-d**ie**ren	stand**aa**rd-stand**aa**rdiseren
	non-native stem	lim**ie**t-lim**ie**ten	lim**ie**t-lim**i**teren

Survey (26) shows that only combinations of non-native stems with non-native suffixes are treated as a single domain for spelling rules. There is only one exception to this generalization in the test lexicon: the word *bruuskeren* that was introduced in 1954 when *brusk* was changed to *bruusk*. I will consider this spelling change to be a mistake.[20] Note that the spelling of non-native derivations from stems with an idiosyncratic spelling such as *mythisch*, *indexeer*,

---

[19] In this respect spelling rules behave like stress rules which also ignore the morphological structure of non-native derivations. Derivations such as *kristalliseer* are stressed the same way as underived words such as *individu,* see for instance Trommelen & Zonneveld (1989:183).

[20] Other exceptions are formed by the spelling of *Semitisch* and *Israëlitisch* in [*Woordenlijst 1914*]: *Semietisch*, *Israëlietisch*, the name *Markiezaat*, and new formations such as the trade name *dieetella/*dietella*.

therapeutisch can be predicted from the constituting parts (*mythe* and *-isch*, *index* and *-eer* and *therapeut* and *-isch*, respectively).[21] This means that exceptional properties of morphemes carry over to complex words.

4.6 Conclusion

This chapter focused on the question to what extent the spelling of loan words in Dutch is predictable by rules.

Before this question could be answered, it was necessary to define which words are loan words. Following Te Winkel's suggestion, I did not define words as loan words on the basis of their etymological origin, but on the extent of adaptation to the Dutch linguistic system. This way, we need not consider words such as *kelder* that can no longer be distinguished from indigenous words to be loan words because they are adopted from another language.

A set of phoneme-to-grapheme conversion rules for non-native words was proposed. The rules for /i/ are a reformulated version of the rules in Neijt & Zuidema (1994a). Other rules are modified versions of the rules in Neijt & Zuidema (1994b). New rules were proposed for the spelling of /z/, /ɑu/, long vowels and some foreign phonemes. These rules capture hybrid patterns that are the result of spelling adaptation in specific environments, predominant etymological patterns, generalizations about the context where spelling is adapted (in words with certain affixes) or not adapted (words with foreign phonemes), and generalizations about the consistent spelling of the same sound. With these rules, words such as *courgette*, *subsidie*, *mechaniek*, *speciaal*, *laurier* and *luxe* are regular, although their spelling deviates from that of native words. If we determine the spelling of non-native words by means of the spelling rules for native words, we can predict the correct spelling for only 25% of the words (for 30% of the words if we take the rule for the spelling of /i/ into account). All other non-native words are exceptional. However, if we use the spelling rules for non-native words proposed in this chapter it becomes possible to correctly predict the spelling of 73% of the words.

The fact that it is possible to formulate generalizations about the spelling of non-native words shows that a spelling that is in accordance with a strict interpretation of the Principle of Etymology is undesirable since it fails to

---

[21] Observe that the uniform spelling in pairs which are not related by productive rules such as *larynx-laryngaal* is only diachronically motivated, and that spelling contrasts can arise as soon as such pairs are no longer felt to be related. Some examples are **c**abine ~ **k**abinet, **c**liënt ~ **c**lientèle, **k**ader ~ en**c**adreer, **k**assa ~ in**c**asseren, **k**atabool ~ **c**atastrofe, **k**limaat ~ a**c**climatiseer, **k**oloriet ~ **c**olorist, **k**ompres ~ **c**ompressie, **k**orporaal ~ **c**orporale, **k**osmos ~ **c**osmetica, **k**ostuum ~ **c**ostumier, **kw**adraat ~ **qu**adrageen, **kw**estie ~ **qu**aestor, **kw**intet ~ **qu**into, mu**z**iek ~ mu**s**iceer, o**ct**et ~ o**kt**ober and pra**ct**icum ~ pra**kt**isch. Te Winkel (1863:76) already drew attention to such 'inconsistencies'.

account for the generalizations about the spelling of non-native words, or for the observation that the spelling of exceptional non-native words is often adapted. A set of phoneme-to-grapheme rules with special rules for non-native words is therefore to be preferred. I conclude that the Principle of Etymology is not an accurate statement about the spelling of loan words.

Finally, 4.5 revealed that combinations of non-native stems and non-native suffixes are treated as single morphemes by phoneme-to-grapheme conversion rules.

Chapter 5

Autonomous spelling rules

5.1 Introduction

The preceding chapters have shown that some spelling alternations can be described more adequately on the basis of letter sequences than on the basis of the pronunciation. An example is formed by the alternation between single letters and geminates, as illustrated in (1):

(1)   r**aa**m   r**a**men   *r**aa**men
       pa**k**   pa**kk**en   *pa**k**en

The letter sequences <aa> and <kk> encode /a/ and /k/ in some cases, e.g. in *raam* and *pakken*, but in other *orthographical* contexts <aa> and <kk> are ungrammatical. The alternations are systematic and predictable, but do not correspond to alternations in the pronunciation. The alternations cannot be covered by rules that refer solely to the pronunciation.

For this reason, I proposed to supplement the traditional phoneme-to-grapheme conversion rules with a second type of spelling rules, namely autonomous spelling rules. In chapter 3 we have seen that the introduction of such rules allows us to maintain the claim that phoneme-to-grapheme conversion rules are restricted to the morpheme domain. Only autonomous spelling rules operate in domains larger than the morpheme. An overview of the autonomous spelling rules discussed so far is given under (2):

(2)   **Autonomous spelling rules from chapters 2 and 3**
       a   Vowel Degemination   *raam-ramen*
           Consonant Doubling   *ram-rammen*
       b   Consonant Degemination   *brand-gebrand* (*ge+brand+d*)
           Spelling Devoicing   *huis-huizen, raaf-raven*

which are conditioned by orthographical context. Note, however, that they operate in domains larger than the morpheme. Conversely, the autonomous
spelling rules of (2b) were proposed to account for alternations that involve domains larger than the morpheme. However, these rules are also conditioned by orthographical context. In the case of Spelling Devoicing this can be illustrated by its sensitivity to orthographical syllables: *grijs-aard* versus *grijn-zaard*; in the case of Consonant Degemination by the fact that the inflected form of *bourgeois* with a mute <s> is written as *bourgeois*, not as **bourgeoiss*. In other words, exactly those rules that apply across morpheme boundaries are sensitive to orthographical context, whereas rules that apply to morphemes are conditioned by phonological context.

In this chapter, I will argue that the following spelling phenomena can also be described most adequately by means of autonomous spelling rules (the same holds for rules such as Hyphen Insertion and Capilization, see [*Woordenlijst 1995*], p. 29–33, which will not be discussed here since they apply to domains larger than the (prosodic) word):

(3)   **Additional autonomous spelling rules**
         Hyphenation   Diaeresis Placement
         Apostrophe Placement   Alternation of *i* and *ie*
         Vowel Doubling   Orthographic Diminutive Allomorphy
         Alternation of *ng* and *n*

The alternation of <i> and <ie> was already mentioned in chapter 4.[1]

Hyphenation, Diaeresis Placement and Alternation of <i> and <ie> have earlier been claimed to be phoneme-to-grapheme conversion rules, but it will be argued here that they can be reformulated as autonomous spelling rules. We will see that in some cases this leads to a more adequate account of the facts in question. It will be argued that the rules under (3) also refer to orthographical context and domains larger than a single morpheme.

This chapter is organized as follows: first, I will address some remaining questions related to the alternations already encountered in chapters 2 and 3. Section 5.2 addresses the question why Vowel Degemination and Consonant Doubling are used in Dutch spelling. Section 5.3 deals with complications that arise from Consonant Degemination. In 5.4–5.6 I will move on to phenomena that have not yet been examined. For these phenomena, I will summarize the accounts in the literature, formalize the rules and propose modifications where necessary. Section 5.7 discusses the computation of orthographical syllables. In 5.8 I will argue that autonomous spelling rules treat native and non-native words

---

[1] Zonneveld (1980) discusses none of these rules, while Kerstens (1981) only discusses some aspects of Diaeresis Placement, and insertion of apostrophe in cases such as *borst'len.*

the same way. In 5.9 the properties of autonomous spelling rules and phoneme-to-grapheme conversion rules are compared. The chapter ends with some concluding remarks on the nature of autonomous spelling rules in 5.10. This chapter deals with non-native as well as native words in anticipation of the conclusion in 5.8 that autonomous spelling rules treat all words alike.

5.2 Why Vowel Degemination and Consonant Doubling

Vowel Degemination and Consonant Doubling have already been discussed in chapter 2. The rules are repeated below (Vowel Degemination is a slightly adapted version of the rule proposed by Wester, in order to prevent *lochen*):

(4) **Vowel Degemination**

$$V_i \rightarrow 0 \, / \, V_i \_ \,]_s \quad [_s C$$

(5) **Consonant Doubling**

$$0 \rightarrow C_i \, / \, C_0 \, V \_ \,]_s \, [_s C_i$$

I will now address the following question: why does the spelling system use Consonant Doubling and Vowel Degemination instead of leaving the result of phoneme-to-grapheme conversion unchanged? Would this not result in a more straightforward correspondence between spelling and sounds?

These questions were raised by Wester (1985b) also addressed these questions in relation to the spelling system.[2] Wester claims that vowels are degeminated in order to use as few letters as possible, and that consonants are doubled to provide an unambiguous spelling. In Wester's view, Consonant Doubling and Vowel Degemination together generally ensure an unambiguous code. For this reason, Wester proposes that spelling is subject to two constraints: Economy (eliminate spelling-characters that are redundant in the face of the spelling system) and a principle which resembles the Readability Requirement (spelling must be formally predictable in its lexical letter-to-sound relationships), see Wester (1985b:209).

In accordance with Wester's generalization, we find that Vowel Degemination is blocked in those cases where a single vowel could be incorrectly interpreted as encoding a short vowel or a schwa in *the same spelling context*. Because of *ze*, *ram*, *mest*, *pochen* and *hoi*, pronounced as [zə], [rɑm],

---

[2] These questions were also raised by Van Heuven (1979) and answered from the perspective of the reader, whereas Wester (1985b) discusses them in relation to the spelling system.

[mɛst], [pɔxə] and [hɔj], there is no Degemination in *zee*, *raam*, *meest*, *goochem* and *hooi* ([ze], [ram], [mest], [ɣoxəm], [hoj]).[3]

However, there are some exceptions to Wester's generalization. The first is the redundant spelling of words such as *eeuw* ([ew]). In the absence of words such as [dyw] or [ɛw], a single letter would suffice to represent a long vowel, so we would expect **euw* and *duw*. However, only the geminate in words such as *duw* is degeminated. This could be explained by the fact that <eu> can be incorrectly read as [∅] (e.g. in *Euwe*), while *du-wen* cannot lead to an incorrect reading.

The second exception is that we do not write [ɣerə] and [ɣərɛn] as **geeren* and *geren*. Both words have the same spelling (before 1954, this only held for part of the words with <ee> in open syllables because of the etymological difference between pairs such as *steen-stenen* 'to sigh' versus *steen-steenen* 'stones'). The restricted distribution of schwa may have played a role here. In native words this sound only occurs in certain (semi-) affixes (*ver-*, *-el*, etc.); so given the morphological analysis of a word, <e> is not ambiguous. Apart from these cases, Degemination is indeed blocked where it would lead to an ambiguous spelling.

Although disambiguation can thus form the answer to part of the question addressed above, it does not explain why Consonant Doubling and Vowel Degemination apply when the underlying spelling already provides an unambiguous non-redundant code for the pronunciation. That this is the case is illustrated in (6) (this underlying representation of short and long vowels was motivated in 2.4):

(6)   a   **Output of spelling system without (de)gemination**
            raam, kaamer
            kam, jamer
       b   **Output of spelling system with (de)gemination**
            raam, kamer
            kam, jammer

Both the actual spelling in (6b) and the underlying spelling in (6a) use geminate letters in two cases and single letters in two other cases. It appears that a non-redundant unambiguous code is already available without the application of

---

[3] Dibbets (1983:67) notes that the current spelling of vowels before <ch> is an anomaly: the length contrast is expressed by single vowel letters versus geminates here instead of by single consonants versus geminates. This has not always been the case: [x] after short vowels has also been written as *(h)h*, *cch*, *chch*, *gg*, *chg* and *gch*. Siegenbeek (1805a) wrote *la**gch**en*. The present system was introduced by Te Winkel. He considered *lachchen* preferable over irregular *lagchen* and *lachen*, but since this spelling was not commonly used he has chosen *lachen* instead. This choice makes it necessary to refrain from degemination in open syllables in words like *goochel* in order to prevent a word such as *pochen* from becoming ambiguous.

rules, so there seems to be no reason to apply them.[4] Obviously, once there is Vowel Degemination, Consonant Doubling is needed to retain readability, but if double vowels were not degeminated, Consonant Doubling would be superfluous.[5]

If spelling tends to use as few letters as possible, we would expect that all letters that are not needed to prevent an incorrect pronunciation would be removed. However, this is not the case. For instance, the sequences <dd> and <ie> are not degeminated to <d> and <i> in *verbrand-den* and *neurie* (see 5.4 and 5.6.1). This cannot be explained by the need to avoid an ambiguous spelling since the alternative spelling **verbran-den* and **neuri* would not cause reading difficulties. In the case of *verbrandden*, the sequence <dd> provides crucial information for the reader (it distinguishes it from the present tense of the verb that is written as *verbranden*). The <ie> in *neurie*, however, cannot be motivated in such a way. Moreover, in other cases geminates that might provide information that is useful for the reader are simplified, cf. *verste* (*vers+ste*, 'freshest') versus *verste* (*ver+ste*, 'furthest') or *maatje* (*ma+tje* 'friend') versus *maatje* (*maat+je* 'measure'). This implies that the factor that determines the presence or absence of geminates is not economy or information value.

I will therefore propose a slightly different analysis in which Vowel Degemination is not motivated by the need to reduce the number of letters but by another requirement that is not violated by doubling of consonants. Consider the following overview of single letters and geminates:

(7)   **Distribution of single letters and geminates**
        Single letters   **Geminates**
        r**a**-men/*raamen   raam/*ram
        ha**m**/*hamm   ha**m-m**en
        gebran**d**/*gebrandd   bran**d-d**e

A generalization that holds for all facts in (7) is the following: sequences of identical letters are avoided *within one syllable*, unless forced by the Readability Requirement as in *raam*.[6] Stated otherwise, tautosyllabic geminates are forbidden in Dutch, unless required by the Readability Requirement. Avoidance

---

[4] However, according to Van Heuven (1980:63), Vowel Degemination occurs about twice as often as Consonant Doubling in average Dutch texts, so the effect is that average word length is slightly decreased in comparison with the underlying representation.

[5] It is unlikely that Consonant Doubling is caused by an orthographic counterpart to the phonological requirement that a rhyme must consist of at least two X-positions (VV or VC), since there are syllables which consist of one position only: *hè* ([hɛ]) and words in which a short vowel is followed by a digraph *pochen*.

[6] A similar constraint was proposed in Kerstens (1981:34). "Within a syllable only sequences of dissimilar letters are allowed." [In één lettergreep zijn alleen lettercombinaties toegestaan van ongelijksoortige letters.]

of geminates does not impose deletion of <e> after an <i> or of one of a sequence of heterosyllabic identical letters.

Summarizing, Vowel Degemination removes tautosyllabic geminates (in those contexts where there are no single vowel letters linked to one V-position) and Consonant Doubling ensures that the resulting spelling is not ambiguous.

5.3 Degemination of heterosyllabic consonants and <s>

In 3.4.1, I argued that spelling alternations that seem to represent the effect of the sound rules Degemination, Final Devoicing and E-deletion, can be described more adequately as alternations that concern letter sequences only. For this reason I proposed the autonomous spelling rules Consonant Degemination and Spelling Devoicing. The former rules were sufficiently discussed in 3.4, but there are two aspects of Consonant Degemination, repeated under (8), which remain to be discussed:

(8) **Consonant Degemination**

$C_i \rightarrow \emptyset$ / [$_s$ …C_i_]$_s$ e.g. *verbrand*

The first issue is why some heterosyllabic geminates are affected while others remain intact, and the second how we should account for facts such as *logischte* (*logisch+ste*) that are reminiscent of Degemination.

First consider the Degemination of heterosyllabic consonants. In 3.4.1, I concluded that the spelling rule Consonant Degemination only affects tautosyllabic consonants. However, the rule seems to work across syllable boundaries in some words but not in others:[7]

(9)   **Degemination**   **no Degemination**
        wijste (wijs + st + e)   wasster   (was + ster)
        Friese (Fries + s + e)[8]   verglaassel   (ver + glaas + sel)

Degemination across syllable boundaries occurs only before a suffix that does not contain a vowel, as also observed in the introduction to [*Woordenlijst 1872*]. If we postulate that Syllabification (see 5.7) is applied every time a suffix is added, the correct result is derived. After the addition of *-st* to *wijs*, this suffix is syllabified in the same syllable as the stem final *s* and Degemination is

---

[7] Some writers incorrectly apply the rule to suffixes with a vowel, e.g. **fietster* (*fiets+ster*) instead of *fietsster*, so it seems that they apply the rule across syllable boundaries as well.

[8] Before 1947 /s/ was written as <sch> in these words, e.g. *trotscht* (**trotschst*), *Friescht* (**Frieschst*) but the sequence <schst> occurred in *mutsenwaschster*. I will discuss deletion of <s> after <sch> below.

applicable: *wijsst* → *wijst*. Subsequent addition of the suffix *-e* creates a second syllable. In the case of *-ster*, the suffix forms a syllable on its own so Degemination is not applicable. The contrasts between verbal forms such as *verbrandde* (first person singular past tense of *verbranden*) and adjectives like *verbrande* (inflected form of *verbrand*) can be accounted for in a similar way, since the verbal forms are composed of *ver+brand+de*, whereas adjectives like *verbrande* have the structure *ver+brand+d+e*.[9]

A second complication of Degemination is the fact that <s> is not only deleted after <s> and (because of Spelling Devoicing) <z> but also after <x>, <sch>, <sh>, <sj> or <z> (in the case of the genitive suffix it is replaced by an apostrophe). Such cases only occur in names and non-native words:

(10)   a   logisch+ste   logisch**te**[10]
               complex+ste   complex**te**
         b   Strijbosch+s   Strijbosch**'**
               Alex+s   Alex**'**
               Heutz+s   Heutz**'**
               Bush+s   Bush**'**

These facts seem to indicate that it would be a mistake to account for facts such as *wijs-wijst*, etc. by a spelling rule: if we assume that the alternation is the effect of the reflection of the sound rule Degemination, we can immediately account for (10): all words end in /s/, so Degemination is applicable after the addition of a suffix beginning with /s/. However, facts such *verbran**dd**en* show that the sound process Degemination is not reflected in orthography.

The facts in (10) cannot be accounted for by orthographical Consonant Degemination either. Although <ss> is also replaced by <s'> or <s> when the first <s> is silent as illustrated in (11a), (11b) shows that combinations of another letter with <s> are *not* changed to <s> when <x> or <z> are silent:

(11)   a   Louis   [luwi]   Louis**'**   [luwis]
               bourgeois   [burzjwɑ]   bourgeoi**s**   [burzjwɑs]
         b   pince-nez   [pɛ̃sne]   pince-ne**zs**   [pɛ̃snes]
               bordeaux   [bɔrdo]   bordeau**xs**   [bɔrdos]

More examples are *entre-deuxs* and *cache-nezs*. These data show that the rule responsible for (10) is not the same as Consonant Degemination.

---

[9] For this analysis it is crucial that the past tense suffix is *-de* as proposed by Booij (1996) and not *-d+e* as proposed by Zonneveld (1980), since in the latter case we would not predict the contrast between verbs and adjectives.

[10] The examples under (10a) do not occur in [*Woordenlijst 1995*], but *logischte* occurs in [*Woordenlijst 1954*].

Observe that it is possible to formulate a descriptively adequate spelling rule for the facts in (10) on the basis of spelling only. We can denote <sch>, <sh> and <sj> by 'graphemes starting with the letter s', and we can also formulate a separate rule for <x>. The distinction between silent and non-silent letters can be made by referring to CV-structure if silent letters are not linked to CV-positions; the rules only apply to graphemes linked to a C-position.[11]

(12)   **Additional Degemination rules**

$s \rightarrow 0$ / <sC_0> _   (*logische*)

$s \rightarrow 0$ / <x> _   (*complexte*)

CONDITION: <sC_0> is linked to one C-position; <x> is linked to two C-   positions.

Since (12) accounts for the facts without referring to sound representations and without violating the Morphological Principle, I will prefer an account that includes rule (12) to one which implies that the alternations in (10) and (11) are the reflection of phonological Degemination.

5.4 Hyphenation

A spelling phenomenon that has not yet been discussed in the preceding chapters is Hyphenation. When words are too long to fit on a line, they are broken off and a hyphen is inserted after the first part. In Dutch, this cannot happen just anywhere, but only in specific positions. This phenomenon is described in different editions of the *Woordenlijst* and Wester (1985b). These accounts will be summarized here. I will show that Wester's claim that Hyphenation is conditioned by phonological syllables leads to incorrect predictions, and argue that Hyphenation is governed by orthography.

Hyphenation in [Woordenlijst 1954]
In the earliest prescriptive works such as Te Winkel (1863) Hyphenation was not discussed, but this topic was addressed later in Te Winkel (1884). Hyphenation instructions in [*Woordenlijst 1954*] are as follows (I give the formulation of 1954 which is more explicit, but essentially the same as the most recent version):[12]

---

[11] Note that this is the only spelling rule where the use of orthographical distinctive features (as proposed by Zonneveld) could lead to a modest simplification: we could define a class of sibilants after which <s> is deleted. I consider this to be insufficient reason for the extension of the formalism available for the formulation of autonomous spelling rules.

[12] In [*Woordenlijst 1995*], p. 28, the following new requirements were added: "Hyphenation may not leave a syllable of one separate letter at the end or beginning of a line. This also holds for words that are part of a compound or derivation. Not *a-drenaline* or *studi-o*; *mensa-pen* or *vide-oachtig*. [...] No

(13) **Hyphenation according to [*Woordenlijst 1954*]**
      Hyphenation is possible:
      1. between two adjacent vowels that do not form a diphthong: *be-amen, brij-achtig, bui-ig, draai-ing, kri-oelen, na-apen* etc.;
      2. before a word or stem that is part of a compound or derivational compound: *boom-vrucht, doorn-struik, eier-koek, hard-nekkig, heren-huis, in-heems, kerse-boom, kwaad-aardig, weer-spannig, ziels-be-droefd* etc.;
      3. after a prefix: *be-horen, er-kennen, ge-lag, her-ademen, on-gelukkig, ont-vangst, ver-zuimen, wan-beleid* etc.;
      4. before the suffixes *-aard* and *-achtig*, and before suffixes beginning in a consonant: *blood-aard, blauw-achtig, boom-pje, dek-sel, naai-ster, was-ster; gedwee-ste* etc.
      5. If hyphenation is not settled by one of the foregoing rules, the following positions are available:
          a. before one intervocalic consonant or *ch*: *be-ter, bre-kebeen, he-ren, la-chen, li-chaam, lo-pen*, etc.
          b. between two intervocalic consonants (*ng* counts as two consonants: *konin-gen* […], *gees-ten, har-ten, har-tig, paar-den; mees-ter; bes-te, mees-te; wijs-te*.
          c. in clusters of more than two consonants before the longest string of consonants that can occur at the beginning of a word: *amb-ten, art-sen, ek-ster, ern-stig, erw-ten, koort-sig; praktisch-te*.
      In the case of non-native words, there are special complications that merit a separate discussion. For monomorphemic words, or words that can be considered monomorphemic, the rules 1 and 5 generally apply as is also the case for members of compounds. In addition, obstruent and liquid, as well as *qu* (=*kw* or *k*) are both considered part of the following syllable (*a-pril, a-quaduct, cho-quant*). Between two vowels there is no hyphenation before or after *x* (*exa-men, exo-tisch*); *y* in words such as *royaal, relayeren* belongs to the preceding syllable *roy-aal*.[13]

---

unwell-formed spelling of the first part should arise. This implies: a. the first part should be a possible syllable-final sequence, not *naa-ste*, but *naas-te*. b. hyphenation may not give rise to another pronunciation than intended, i.e. not *reg-lement* but *re-glement*, not *pis-tool* but *pi-stool*, not *rec-lame* but *re-clame*. The second part must be pronounceable: not *am-bten* but *amb-ten*. The diminutive *-kje* is considered pronounceable: *harin-kje*."

[13] "Men mag afbreken: 1. tussen twee klinkers die onmiddellijk op elkaar volgen en geen tweeklank vormen: *be-amen, brij-achtig, bui-ig, draai-ing, kri-oelen, na-apen* enz.; 2. vóór een woord of stam, deel uitmakend van een samenstelling of samenstellende afleiding: *boom-vrucht, doorn-struik, eier-koek, hard-nekkig, heren-huis, in-heems, kerse-boom, kwaad-aardig, weer-spannig, ziels-bedroefd* enz.; 3. na een voorvoegsel *be-horen, er-kennen, ge-lag, her-ademen, on-gelukkig, ont-vangst, verzuimen, wan-beleid* enz.; 4. vóór de achtervoegsels *-aard* en *-achtig*, en vóór achtervoegsels die

Wester (1985b)

The hyphenation positions prescribed by (13) generally coincide with phonological syllable boundaries: [kri-ul]/*kri-oel*, [lɪ-xam]/*li-chaam*, [bɛ-tər]/*be-ter*, [hɑrtəx]/*har-tig*, [ɑmp-tə]/*amb-ten*, etc. For this reason Booij (1985, 1995) and Wester (1985b) claim that hyphens appear at boundaries of phonological syllables. The most detailed account of Hyphenation is provided by Wester (1985b), who proposes the following rule:

(14)   **Hyphenation**
         0 → - / [+seg]]σ _ σ[[+seg]

The claim that Hyphenation is conditioned by phonological syllables does not always lead to the correct spelling. We have already seen some examples in chapter 2 (**aa-ien*, **ee-uwen*, **vi-nger*) and chapter 3 (**grij-saard*, **reu-zachtig*). Incorrect predictions are also made for the facts in (15):

(15)	**Phonological syllables**	**Predicted hyphenation**	**Hyphenation**
a	[wa-rɑxtəx]	**w**a-r**achtig	w**aar-**achtig
	[ar-dɑpəl]	**aar-**d**appel	aar**d-**appel
	[hɛ-rɪ-nə-rə]	**he**r-r**inneren	he**r-**inneren
	[he-lɑl]	**h**e-l**al	h**eel-**al
b	[mɑrk-sɪst]	**Marxist	Ma**r-x**ist
	[ɛk-stra]	**extra	e**x-t**ra

In (15a) hyphenation positions coincide with morphological boundaries rather than phonological syllables. We would expect no hyphenation in (15b) since <x> encodes heterosyllabic consonants.[14]

---

met een medeklinker beginnen: *blood-aard, blauw-achtig, boom-pje, dek-sel, naai-ster, was-ster; gedwee-ste* enz. [...] 5. Als de afbreekplaats niet reeds door een der voorafgaande regels vaststaat gelden de volgende regels: a. Één tussenmedeklinker (ook ch) gaat naar het volgende woorddeel: *be-ter, bre-kebeen, he-ren, la-chen, li-chaam, lopen*, enz. [...] Bij het afbreken geldt *ng* voor twee medeklinkers. b. Van twee tussenmedeklinkers gaat de eerste naar het voorafgaande, de tweede naar het volgende woorddeel: *konin-gen* [...], *gees-ten, har-ten, har-tig, paar-den; mees-ter; bes-te, mees-te; wijs-te*. c. Van meer dan twee tussenmedeklinkers gaan er zoveel naar het volgende woorddeel als er aan het begin van een woord kunnen voorkomen: *amb-ten, art-sen, ek-ster, ern-stig, erw-ten, koort-sig; praktisch-te*. Bij de *bastaardwoorden* doen zich eigenaardige moeilijkheden voor, die een afzonderlijke bespreking noodzakelijk maken. Voor enkelvoudige of als zodanig te beschouwen woorden gelden in het algemeen de regels 1 en 5; zo ook voor de delen van een samenstelling. Daarbij worden occlusief + liquida, alsmede *qu* (=*kw* en *k*) beide bij het volgende woorddeel getrokken (*a-pril, a-quaduct, cho-quant*). Tuscsen twee klinkers wordt voor en na *x* niet afgebroken (*exa-men, exo-tisch*); *y* in woorden als *royaal, relayeren* gaat naar het voorafgaande woorddeel: *roy-aal*."

[14] Note that other languages where <x> also represents [ks] do not treat the letter <x> as expected on the basis of phonological syllable structure. In South African, a hyphen is inserted after <x>: *Tex-as, pirox-een*, [*Afrikaanse woordelys*], *p*.65, and in German before <x>: *fle-xibel, bo-xen*, see [*Duden*

Incorrect predictions are also made in the case of the cluster <st>. It is hard to determine the phonological syllable structure for words with intervocalic [st]; is [s] the last consonant of the preceding syllable, or the first consonant of the following one? De Schutter & Collier (1986) showed that listeners' judgements vary here.[15] Different analyses have been proposed. Trommelen (1983:128) proposes that [st] is an onset after obstruents, but not after sonorants: [ɛk-stər] versus [mɛs-tər], [ɣlɪns-tər]. Van der Hulst (1984:116–117) and Booij (1995:26–28) on the other hand always analyse [s] as part of the onset: [ɛk-stər] versus [me-stər], [ɣlɪns-stər]. In phonology either sonority or the number of syllable positions is the factor that determines whether [st] is split, but in orthography it is the distinction between vowel letters and consonant letters: *st* is split after a vowel digraph, but not after a consonant. Overview (16) gives the hyphenation patterns predicted by the syllable structure proposed by Trommelen (T) and Van der Hulst or Booij (V+B) and the actual pattern, respectively:

(16)   **The hyphenation of the cluster *st***

T:	ek-**st**er	mee**s**-**t**er	*	glin**s**-**t**er
V+B:	ek-**st**er	*	mee-**st**er	glin-**st**er
actual spelling:	ek-**st**er		mee**s**-**t**er	glin-**st**er

The overview in (16) shows that under either analysis phonological syllables incorrectly predict the hyphenation positions in some words.

A final indication that phonological syllables are not the conditioning factor for Hyphenation is the fact that words with the same phonological syllable structure but with a different morphological structure are sometimes hyphenated differently:

(17)

		pronunciation	**hyphenation**	**morphological structure**
	a	[vɛr-stə]	ver-**st**e	ver+ste
		[vɛr-stə]	ver**s**-**t**e	vers+ste
	b	[ma-tjə]	ma-**tj**e	ma+tje
		[ma-tjə]	maa**t**-**j**e	maat+je

[15] Although it is not difficult for native speakers to indicate the number of syllables in spoken words, it is sometimes harder to make out where one syllable ends and the following one begins, see for instance De Schutter & Collier (1986). One consonant may even be analysed as part of two consecutive syllables in words such as [jɑmər] (*jammer*). Schiller et al. (1997) demonstrated with a syllable reversal test that words with a similar syllable structure are not always treated the same way, and that syllabification must be considered the effect of preferences rather than strict rules, see Schiller (1997:172). Resyllabification may take place, depending on style and speech rate. As shown by hyphenation patterns, orthographic syllabification does not vary, and there is no evidence for ambisyllabic consonants or style-dependent resyllabification.

The facts in (15)-(17) show that Hyphenation can only be adequately described under the assumption that it is conditioned by orthographic syllables. Under this analysis Hyphenation is an autonomous spelling rule. I will therefore adopt Wester's Hyphenation rule, but claim that it applies to orthographic syllables. It is reformulated as (18) in which 'L' indicates an arbitrary letter and 'X' stands for either V or C:

(18)   **Hyphenation**

$$
\begin{array}{cc}
\text{X} & \text{X} \\
| & | \\
\end{array}
$$
$$0 \rightarrow \text{- / L }]_s \underline{} [_s \text{ L}$$

The computation of orthographic syllables will be discussed in 5.7.

5.5 Placement of diacritics

Within monomorphemic words and their derived or inflected forms Dutch spelling uses two types of diacritics to prevent an incorrect reading that could arise after the combination of morphemes (and in some cases morpheme-internally in non-native words), i.e., diaereses and apostrophes:

(19)   **spelling without diacritic and suggested reading**
         geind   *[ɣɛint]
         pas   *[pɑs]
         spelling with diacritic and intended reading
         geïnd   [ɣəɪnt]
         pa's   [pas]

In both cases, diacritics change a potentially ambiguous underlying spelling into a surface spelling that is not ambiguous with respect to the pronunciation it represents. I will now discuss the distribution of diaereses and apostrophes.

5.5.1 Diaeresis placement

A diaeresis is added to vowels (ä, ë, ï, ö, ü) to prevent incorrect digraph interpretations within monomorphemic words and their derived and inflected forms: *moeë* ([mu-ə], not [mo-e]), *smeuïg* ([smØɪx], not [smə-œyx]). This rule is only applied to non-hyphenated words (cf. *moeë* versus *moe-e*). A diaeresis is also inserted in compound numerals such as *tweeëntwintig*, 'twenty-two' (see [*Woordenlijst 1995*], p. 33–34, but not in derivations with *-achtig*, see

[*Woordenlijst 1995*], p. 31–33). I will discuss the rules for this phenomenon in [*Woordenlijst 1954*], Wester (1985b) and Zuidema et al. (1994) and propose some modifications of the latter approach in order to make it fit into the framework proposed here.[16]

[Woordenlijst 1954]
The distribution of diaereses is described as follows in the [*Woordenlijst 1954*], p. LXVI:

(20)   A diaeresis is added to the second of two adjacent vowels within a word to prevent a possible incorrect pronunciation that could be the result of a mechanical reading.[17]

Wester (1985b)
Wester points out that this rule predicts that we should write a diaeresis in all cases of two possible parsings. This is correct in the case of sequences of two letters. However, in the case of three or more consecutive vowel letters, there are potentially ambiguous words without a diaeresis like *uien*, *prieel* and *eieren*. According to Wester, this description can be improved by referring to the relation between letters and phonemes. It is necessary to know whether or not two letters represent one sound (*gedeisd* [ɣədeist] versus *deïst* [deɪst]). Sounds and spelling are related by the following set of (recursively) ordered rules ('V' abbreviates *a*, *e*, *i*, *o* and *u*; 'V$_i$V$_i$' abbreviates *aa*, *ee*, *oo* and *uu*):

(21)   V$_i$V$_i$ > oei > oe > ei > ui > eu > au > ou > ie > V [18]

The ordering in (21) states that where there is an ambiguous spelling, the parsings on the left have priority over those on the right, for instance *geuit* is parsed as *ge-uit* because <ui> has priority over <eu>. Whenever a reading is possible that is ranked higher than the reading that is intended, e.g. <ie> instead of <ee> in *knieën*, the diacritic is added. Wester thus explains why there is a diacritic in the words under (22a) but not in those under (22b):

---

[16] Another account is given in Kerstens (1981:39–41).

[17] "Het deelteken (trema) wordt geschreven op het tweede van twee opeenvolgende klinkertekens van eenzelfde woord in gevallen waarin uitsluitend het voorkómen van een bij mechanisch lezen mogelijke onjuiste uitspraak van het woord wordt beoogd."

[18] As observed by Zuidema et al., Wester (1985a-b) and (1987b) have slightly different rankings. For the present discussion this difference is irrelevant.

(22)   a **diaeresis**   b **no diaeresis**   **ranking**
               smeuïg   [smØəx]   geuit   [ɣəœyt]   ui > eu
               knieën   [knijə]   prieel [priɛl]   ee > ie

Wester's formulation, which uses CV-structure, is an improvement over the traditional description in [*Woordenlijst 1954*]. In Westers's view, placement of diaeresis is not governed by letters only, but also by the corresponding pronunciation, which is not possible in the framework developed here.

Zuidema et al. (1994)
Zuidema et al. (1994) argued that some of the predictions of (22) cannot be checked, since the relevant vowel sequences do not occur. Only the following rankings are really supported by the occurrence of examples:

(23)   trigraphs > digraphs > monographs
               ui > eu   (geuit, smeuïg)
               Ie > ee   (prieel, knieën)
               ui > ie   (uien, jezuïet)

In addition, they pointed out that (21) incorrectly predicts that there is always a default parsing of a given letter sequence (in which no diacritic is needed). In some cases, however, different parsings of the same sequence require a diaeresis:

(24)   **predicted by (23)**   **ranking**   **actual spelling**
               *reëel-*reeen*   ee > e   *reëel-reeën*
               *poëem-*moee*   ee > oe   *poëem-moeë*
               *geëist-*weeig*   ee > ei   *geëist-weeïg*

For this reason, Zuidema et al. (1994) formulated alternative rules:

(25)   **Diaeresis Placement (Zuidema et al. 1994)**
               a   Add a diaeresis to the second of a sequence of two vowel letters
                    that can be incorrectly interpreted as a grapheme.
               b   Add a diaeresis in a sequence of more than two vowels if the
                    sequence *ee* is ambiguous, or if *i* can be incorrectly combined
                    with a vowel in the preceding syllable.[19]

---

[19] "Als een *i* het begin is van een lettergreep èn een mechanische lezer zou menen van niet, dan moet er een trema op die *i*. Als een *ee* deel is van een klinkertekenreeks met een ambigue structuur, dan moet met trema's de structuur worden aangegeven. Als een reeks van twee monografen ten onrechte als een digraaf gelezen dreigt te worden, dan moet er een trema op de tweede monograaf."

These rules avoid the disadvantages of Wester's rules and do not refer to the phonemes encoded by letter sequences.[20] I will therefore follow Zuidema et al. and claim that rankings play no role for diaeresis placement.

Modifications of the rules of Zuidema et al. (1994)
In order to adopt the proposal of Zuidema et al., it is necessary to be able to predict which letter sequences form graphemes. In Wester's framework, where letter combinations are graphemes if they are linked to one position on the CV-tier, we could use CV-structure for this purpose. Since the version of CV-structure proposed in this study expresses the fact that some single letters count for two (see 2.4.3), it cannot be used to account for the fact that digraphs behave as an entity with respect to diaeresis. This seems to be a disadvantage of the formalism proposed here.

However, if a diaeresis clarifies syllable structure rather than the parsing of letters into graphemes, CV-structure is not needed for this phenomenon. The rules of Zuidema et al. can be reformulated as follows:

(26) **Diaeresis Placement**
    a. Add a diaeresis to the second of two vowel letters that can be interpreted to be in the same syllable, if this reading is not intended.
    b. Add a diaeresis to a syllable-initial vowel in a sequence of more than two vowels that contains *ee* or *i* when the position of the syllable boundary is ambiguous.

In addition to the rules of (26) we need a list of letter combinations that are allowed in a syllable: V_iV_i (as defined above); *ae, ie, oe, ai, ei, oi, ui, au, eu, ou.*

The reformulation in (26) implies that we no longer need to analyse letter sequences that occur without a diaeresis as graphemes. For instance, we need not analyse <aai>, <ooi> and <oei> as trigraphs to derive the correct spelling. I will therefore adopt the reformulated version of the rules of Zuidema et al. (1994) that refers to orthographic syllables. In the analysis given here, Diaeresis Placement is an autonomous spelling rule.

5.5.2 Apostrophe placement

The apostrophe is inserted before specific suffixes after single vowels to prevent a reading as a short vowel, for instance *pa's* ([pas], not [pɑs]), *baby'tje*

---

[20] The rules in [*Woordenlijst 1995*], p. 33 are based on the rules of Zuidema et al. (1994): "A diaeresis is used within simplex words, derivations or inflected forms to prevent two subsequent vowel letters from being read as one sound. This pertains to the following fourteen letter combinations: *aa, ae, ai, au, ee, ei, eu, oe, ie, oi, oo, ou, ui* and *uu*. [...] In strings of more than two vowel letters, a diaeresis is not added directly after *i*, and only *e* or *i* receive a diaeresis."

([bɛbitjə], not [bɛbɪcə]). A second function of apostrophes is to mark deletions, for instance the deletion of suffixes that consist of one letter and undergo Degemination (as in *Parijs'* from *Parijs+s*) or other deletions (as in *m'n* from *mijn*). Finally, apostrophes separate suffixes from acronyms or numerals, e.g. *A4'tje, tv's, BRT'er*. The rules are formulated as follows in [*Woordenlijst 1995*], p. 34–35:

(27)   a   An apostrophe is added to plural and genitive ending *-s* after words ending in *a*, *e*, *i*, *o*, *u* or *y*, preceded by a consonant letter or syllable boundary (*e* refers to /e/).
       b   If a name ends in a sibilant, the apostrophe replaces the genitive ending *-s*
       c   The apostrophe is used in derivations of numerals and acronyms.
       d   An apostrophe is used in diminutives of words ending in *y* preceded by a syllable boundary.[21]

Apostrophe placement is not further discussed in the literature, so the rule of the [*Woordenlijst 1995*] will be taken as a point of departure here. The formalism proposed in this study enables a simple definition of vowels after which the diacritic is added: single vowel letters linked to two V-positions:

(28)   **Apostrophe placement**

VV
\\/

       a   s → 's / V + _ #   e.g. *ma's*
       b   s → ' / <s> + _ #   e.g. *Louis' boek*
       c   s → ' / <x> / + _ #   e.g. *Marx' theorie*
       d   s → ' / <s…> / + _ #   e.g. *Bush' doctrine*
       e   tje → 'tje / y _   e.g. *baby'tje*

CONDITION: (28b-d) apply to the genitive suffix only, (28a) to the plural or genitive suffix only

Note that not all ambiguity is avoided by the use of apostrophes. In the first place, apostrophes are absent in some contexts where we would expect them, as illustrated in (29a). In (29b) Vowel Doubling (see 5.6.2) is applied instead:

---

[21] "De apostrof wordt gebruikt bij de meervouds-s van woorden die eindigen op *a, e, i, o, u* of *y*, voorafgegaan door een medeklinkerletter of lettergreepgrens (met de e is de /ee/ bedoeld) [...] De apostrof wordt gebruikt bij de tweedenaamvals-s van woorden die eindigen op *a, e, i, o, u* of *y*, voorafgegaan door een medeklinkerletter of lettergreepgrens (met de e is de /ee/ bedoeld) [...] De apostrof wordt gebruikt in verkleinvormen van woorden die eindigen op een *y* voorafgegaan door een medeklinkerletter [...] Als een naam eindigt op een sisklank, wordt de apostrof gebruikt in plaats van de tweedenaamvals-s."

(29)	**input**	**adjectival -*s***	**genitive -*s***
a	Parijs+s	een Parij**s** straatje	Parij**s'** bistro's
b	Venlo+s	een Venl**oos** straatje	Venl**o's** straten

The examples in (29) illustrate that derivational -*s* is treated differently than inflectional (genitive) -*s*: in the first case the suffix is attached to the words and Degemination is applied when applicable, in the latter case an apostrophe is added. This could be accounted for by restricting apostrophe placement to inflectional suffixes, which means that spelling rules must also be able to distinguish suffix types.

Secondly, no apostrophe is written when a plural or genitive suffix is added to a word ending in schwa. This implies that such words are ambiguous: [ɑnəs] is written as *Annes* that could also be interpreted as [ɑnɛs], cf. *succes*-[syksɛs]. Note, however, that the spelling with an apostrophe would also lead to ambiguity in this case: **Anne's* would suggest [ɑnes], cf. *ave's* ([ɑves]).

In one case, the apostrophe seems to have a wrong effect: the spelling *sjwa's* for [*sjwɑ] suggests a pronunciation with a long vowel, see for instance *ma's*-[mas]. Here we are probably dealing with a mistake; *sjwa* is the only word in which a single final vowel is pronounced as a short vowel. Except for these two cases, apostrophes counteract ambiguity.

5.6 Remaining alternations

In this section, I will discuss cases where letters alternate that were not yet discussed in earlier chapters.

5.6.1 Alternation of <i> and <ie>

In chapter 2, we have seen that the difference between <i> and <ie> sometimes encodes sound contrasts, e.g. *lip* ([lɪp]) versus *liep* ([lip]). However, there are also <i>-<ie> alternations that are independent of the pronunciation: in words ending in unstressed /i/, <ie> is replaced by <i> before a suffix with a vowel (*neurie-neuriën*). On the other hand, <e> is added to words ending in <i> before a consonant-initial suffix: (*ski-skietje*) (see [*Woordenlijst 1995*], p. 35). This does not hold for the plural or genitive suffix -*s* where an apostrophe is added, as already mentioned in 5.5: *ski's*. These alternations are purely letter-based, which can be illustrated by the fact that they only affect non-hyphenated words (*skiede* versus *ski-de*). Consequently, we expect that these alternations are regulated by autonomous spelling rules. However, chapter 4 revealed that alternations such as *neurie-neuriën* are accounted for by phoneme-to-grapheme conversion rules in

[*Woordenlijst 1954*]. Since this involved applying grapheme-to-phoneme conversion rules to domains larger than a single morpheme, the rules were split into two parts. Grapheme-to-phoneme conversion rules only account for morpheme internal effects (which result in a constant morpheme spelling), whereas the alternations across morpheme boundaries just mentioned are accounted for by autonomous spelling rules. These rules are repeated below:

(30)   Change \ie\ into <i> in an unstressed syllable before a vowel (*oliën*, *neuriën*)

(31)   Change morpheme final \i\ into <ie> before a consonant-initial native suffix (*taxiede*) except before the plural or genitive suffix *-s* where an apostrophe is added: *ski's*

Like the rules of 5.2, (31) prevents an incorrect interpretation of vowel letter sequences. Note that the environment of rule (30) is extended from vowel letters representing schwa to all vowel letters. This way, the rule also accounts for native words such as *miauwen*, *krioelen* and *leliaard*. This alternative description of the spelling of /i/ avoids application of phoneme-to-grapheme conversion rules across morpheme boundaries.

    However, with the reformulation of the rule, a new problem arises. Rule (30) seems to violate the restrictions of the hybrid two-level model. As observed above, the difference between words such as *ruziën* and *knieën* depends on the stress pattern of the spoken words. In chapter 4, we have seen that this spelling distinction is based on the incorrect generalization by Te Winkel that the distribution of *i* and *ie* is influenced by the stress patterns of words. However, although Te Winkel's generalization was mistaken, stress is now relevant for the application of rule (30). It is possible to ensure that autonomous spelling rules can distinguish *neuriën* and *knieën* without directly referring to the pronunciation, by encoding information about stress pattern in spelling (during phoneme-to-grapheme conversion). This is an ad hoc measure, which seems warranted since stress only plays a marginal role.

5.6.2 Vowel Doubling

Single vowel letters that represent long vowels are doubled before *-tje*: *ma-maatje*, *auto-autootje* (see [*Woordenlijst 1995*], p. 35):

(32)   When a word ends in *a*, *e*, *o* or *u* preceded by a consonant letter or

syllable boundary, the last letter is doubled [before a diminutive suffix][22]

The same holds for syllables which are closed by the addition of a suffix, e.g. *kano ~ kanoot*, *na ~ naast* and *sla ~ slaan* (for which no rule is formulated in [*Woordenlijst 1995*]). With the formalism proposed in this study it is possible to formulate the following simple Vowel Doubling rules, where V_i refers to *a*, *e*, *o* or *u*: [23]

(33)   a   VV   V V
           \\/ | |
           $V_i \rightarrow V_iV_i$ / _ + tje

       b   VV   V V
           \\/ | |
           $V_i \rightarrow V_iV_i$ / _ + C_1]$_s$

The rules in (33) are not applied to hyphenated words, as illustrated by the contrast between *maatje* and *ma-tje*, which shows that the rule is sensitive to hyphenation. In other words, autonomous spelling rules interact. Note that the letter *y* is treated differently from other single letters linked to two V-positions. Before *-tje* an apostrophe is added (*baby'tje*), while *y* remains unchanged before a consonant (*gerugbyd*).

5.6.3 Orthographic Diminutive Allomorphy

After the addition of a diminutive affix the spelling of words is sometimes changed, e.g. *cliché-clicheetje*; *depot-depootje*, *entredeux-entredeu(x)tje*, *parachute-parachuutje*, *diner-dineetje*, *pardessus-pardessuutje*, *chalet-chaletje*.[24] These alternations only apply to unhyphenated words, as

---

[22] "Als een woord eindigt op *a*, *e*, *o*, *u* of *y*, voorafgegaan door een medeklinkerletter of lettergreepgrens, wordt de laatste letter doorgaans verdubbeld. (Met de *e* is de /ee/ bedoeld). De *i* wordt in deze positie geschreven als *ie*."

[23] An alternative approach would be to consider such geminates to be caused by blocking of Vowel Degemination rather than doubling. However, in that case we would not expect that the presence of a vowel geminate interacts with hyphenation: the word *goochem* in which Vowel Degemination is blocked, is hyphenated as *goo-chem*, not as *go-chem, but *autootje* as *auto-tje* (*autoo-tje). This suggests that *auto-autootje* is treated the same way as *ski-skietje/ski-tje*. In case of alternations such as in *ga-gaat* and *sla-slaan* hyphenation behaviour cannot be used to decide whether geminates are caused by blocking of Vowel Degemination or Vowel Doubling, since these words are monosyllabic. However, if these geminates are the result of blocking of degemination, incorrect predictions are made since autonomous spelling rules are applied iteratively (to account for facts such as *wijste*): *gaa* $\rightarrow$ *ga* $\rightarrow$ *ga+t* $\rightarrow$ *gat. I therefore propose that these facts are also accounted for by Vowel Doubling.

[24] The alternation *parachute-parachuutje* is a spelling phenomenon if the <e> in parachute is mute; otherwise it is a sound alternation, cf. *brunette-brunetje*, *directoire/directoirtje*, *machine/machientje*

illustrated by the contrast between *depootje* and *depot-tje*, and they can be characterized as follows: exceptional spellings (diacritics, mute letters) are changed to more regular spellings before the diminutive suffix (see [*Woordenlijst 1995*], p. 27, 35). The alternation of <é> and <ee> is sufficiently regular to be formalized as a rule:

(34) **Orthographic Diminutive Allomorphy**
$$é \rightarrow ee / \_ + tje$$

The other alternations are idiosyncratic, cf. *depot-depootje/*depottje* but *crapaud-crapaudtje/*crapootje*. These alternations should be described by listing the written allomorphs in the lexicon.

5.6.4 Alternation of <ng> and <n>

Ng is changed into *n* before *k*: *koning-koninkje* (see [*Woordenlijst 1954*], p. XLIV; this alternation is not mentioned in [*Woordenlijst 1995*]). This rule is formulated as follows:

(35) Alternation of <ng> and <n>
$$g \rightarrow 0 / n \_ ( ]_S) k$$

A possible motivation for this rule could be that contrasts such as *bang-bank* suggested that *ngk* is an illegal letter sequence.

5.7 Orthographic Syllabification

In the previous section, we have seen that one of the properties that distinguish phoneme-to-grapheme conversion rules from autonomous spelling rules is that the latter do not refer to the pronunciation. In other words, conditions on the application of rules such as Apostrophe Placement, Diaeresis Placement, Hyphenation, Consonant Doubling and Vowel Degemination concern letters only. However, in earlier accounts of the Dutch spelling system by linguists such as Booij and Wester these rules are considered to be conditioned by structural properties of the pronunciation. For instance, we have seen that Wester claimed that rules such as Consonant Doubling and Vowel Degemination are conditioned by phonological syllables. However, although earlier accounts convincingly showed that spelling simpler and more insightful when they refer to syllables,

---

or *machinetje*, *pelerine/pelerientje*, *serafine/serafientje*, *sinecure/sinecuurtje*, *terrine/terrientje*, *karbonade-karbonaadje*.

they did not prove that these syllables are isomorphic to phonological syllables. Indeed, we have seen in 2.4.3 and 5.4 that this is not the case.[25]

To be able to use simple rules and at the same time account for mismatches between spoken and written words, I proposed in chapter 2 that spelling rules are conditioned by *orthographical* syllables. We have already seen that this assumption improves the description of the alternation of single letters and geminates. Since chapter 2 contains no proposal for the derivation of orthographical syllables, I will address this question here.

As mentioned in 2.4, I proposed that orthographic CV-structure is built on the basis of sound. These structures provide all information necessary to account for syllabification contrasts between words with the same or a similar (superficial) spelling. For instance, contrasts such as *hu-meur ~ me-nu-et, zoeven ~ zo-even* and *rag-lan ~ i-glo* follow from the different number of V-positions linked to the relevant vowels: The digraphs <eu> and <oe> are linked to two V-positions, whereas in <eu> and <oe> each vowel letter is linked to two V-positions. The <a> of *raglan* has one V-position, but the <i> of *iglo* has two. Similarly, words with intervocalic <ng> and <x> on the one hand and words with intervocalic <ch> on the other hand can be distinguished by the fact that <ng> and <x> are linked to two C-positions, while <ch> and other consonants only take up one C-position.

In other cases, hyphenation contrasts are not accounted for by the number of CV-positions or the difference between Cs and Vs, but by the letters they are linked to. For instance, the examples in (36) illustrate that sequences of two or three Cs are treated differently, depending on whether they dominate *mbt* or *kst, rn* or *bl* or *nts* or *nst:*

(36)   **amb-t**en   ɑmp-tə   **ek-st**er   ɛk-stər
        toe**r-n**ooi   tur-nɔj   dou-**bl**é   du-ble
        pa**nt-s**er   pɑnt-sər   gli**n-st**er ɣlɪn-stər

In this case, the similarity between phonological and orthographical syllables is not accounted for by the initial orthographical representations, but by the fact that letters are treated the same way by orthographical syllabification rules as the corresponding sounds are by the phonological syllabification rules. Stated differently, orthographical syllabification is often similar to its phonological counterpart (but crucially different in some respects).

The orthographical syllabification rules are given under (38): [26]

---

[25] In [*Woordenlijst 1954*], p. LIII there was variation in the hyphenation of words such as *bioscoop*: *bios.coop* (because of the pronunciation with a short [ɔ]) or *bio.scoop* (because of the morphological structure).

[26] In the absence of data which show that orthographic syllables have internal structure I will directly combine Cs and Vs into orthographic syllables.

(37)   **Orthographic Syllabification**
      Syllabification domains are formed by free morphemes, by suffixes that
      start with a consonant and contain a vowel, and by *-aard or -achtig*:
      heel.al, *gedwee.ste*, *laf.aard*, *wreed.aard*.
      Within syllabification domains, syllables are built on the basis of the CV-
      tier in the following way:
      a. Adjacent V-positions are heterosyllabic, provided that the first
          syllable has two V-positions: *kri.oel*, *ui.er* (**u.ier*).
      b. One intervocalic C-position belongs to the next syllable, unless it is
          associated to a vowel letter: *be.ter*, *la.chen* but *maai.en*, *coy.ote*.
      c. In the case of two or more intervocalic C-positions:
          • if the C-positions are linked to the combinations *bl*, *br*, *cl*, *cr*, *dr*,
            fl, *fr*, *gl*, *gr*, *kl*, *kr*, *kw*, *pl*, *pr*, *qu*, *tr*, *vr* and *str*, this cluster is
            parsed as an onset: *a.qua*, *di.ploma*, *oe.strogeen*, etc. unless this
            means that the preceding syllable ends in a single V-position, in
            that case one C-position goes to the preceding syllable: *as.trant*,
            rag.lan, *tef.lon*, etc., except in case of the digraph *qu*: *choqueren*,
            attaqueren. Remaining letters belong to the preceding syllable:
            ex.tra, *in.strument*. *uw* is no onset, hence *eeu.wen*.
          • if the C-positions are linked to *sc*, *sk*, *st* or *x*, they are split after
            vowel letters, but not after consonant letters: *fis.caal*, *bruus.keer*,
            gees.ten, *pas.ta*, *taxi*, *bauxiet* (the syllable boundary is between the
            two C-positions linked to *x*) versus *ob.sceen*, *ern.stig*, *in.stant*,
            mar.xist.
      d. Remaining Cs go to the preceding syllable: *amb.ten*, *plank.ton*.
      EXCEPTIONS: *naas.te*;   *bas.taard*,   *bo.gaard*,   *do.laard*,   *dros.saard*,
      grijn.zaard, *ho.vaar.dij*, *mos.taard*, *mut.saard*, *pon.jaard*, *stan.daard*,
      tab.baard, *vein.zaard*; *trots.kist.*

There are three differences between these syllabification instructions and those
that are suggested by the Hyphenation instructions in [*Woordenlijst 1954*].

   Firstly, in (37) the syllabification of *baaierd* and *eeuwen* does not imply the
postulation of trigraphs. Rather, the generalization is that vowels associated to a
C-position do not occur in syllable initial position.

   Secondly, clusters other than combinations of obstruent + liquid are treated
as <st>. This implies that the contrast between *bauxiet*, *ex.tra* and *Mar.xist* is
accounted for in the same way as the contrast between *mees.ter* and *glin.ster*. In
the case of the cluster *sk* or *sc* it is not clear whether it is parsed as an onset after
consonants. The contrast between *ob.scuur* and *bruus.keer* can also be accounted
for by the presence of the prefix *ob-*. The only case where *sk* or *sc* occurs after a
consonant that is not the last consonant of a prefix is *trotskist*, which is

syllabified as *trots.kist* in [*Woordenlijst 1995*]. I consider this to be a mistake.

Finally, although the instructions in [*Woordenlijst 1954*] suggest that all letter sequences that correspond to obstruent-liquid clusters may form onsets of (word-internal) syllables, this is not always correct: **a-tleet*, **mue-sli*, **sei-smisch*. Therefore it is necessary to list the possible onset clusters here.

In some cases, the application of the rules in (37) is complicated by the morphological make-up of words or by the fact that it is not clear whether a given vowel is long or short. These complications are discussed in appendix F.

5.8 Non-native words and autonomous spelling rules

In this section, I will argue that autonomous spelling rules are not sensitive to the native/non-native distinction (as already suggested by the fact that some non-native examples were already discussed in the previous sections).

In line with the general approach followed in 4.4, we can consider two options to describe non-native words. In the first, we follow the Principle of Etymology and state that the spelling of non-native words is given, so we do not apply spelling adaptation rules such as Consonant Doubling to derive their spelling. In the second, we apply autonomous spelling rules to all words irrespective of their origin. We now have option A' and B' for autonomous spelling rules:

Option A': Non-native words are not affected by the autonomous spelling rules for native words

Option B': Non-native words are subject to autonomous spelling rules for native words.

First observe that autonomous spelling rules apply regularly to most complex non-native words. In (38) there are examples of the rules disussed so far: (38)

Rule	native word	non-native word
Consonant Doubling	tonnen (ton+en)	ballonnen (ballon+en)
		duelleer (duel eer)
Vowel Degemination	banen (baan+en)	kanalen (kanaal+en)
		organisch (orgaan isch)
Consonant Degemination	iets vers (vers+s)	iets divers (divers+s)
ng/n-Alternation	koninkje (koning+kje)	saronkje (sarong+kje)
Diaeresis Placement	geïnd (ge+ind)	geüsurpeer (ge+usurpeer)
		reïteratie (re iteratie)
ie → i	neurien (neurie+en)	oliën (olie+en)

i → ie	-	taxiede (taxi+de)
Diminutive Allomorphy	-	cafeetje (café+tje)
Apostrophe Placement	la's (la+s)	corona's (corona+s)
Vowel Doubling	laatje (la+tje)	galaatje (gala+tje)
Spelling Devoicing	vreesde (vreez+de)	chineesd (chineez+de)

Observe that the autonomous spelling rules seem to apply to complex native and non-native words in the same way. There are only a few derived words that are exceptional. The rule that changes \i\ into <ie> has the following exceptions: *pi'tje, i'tje* (letter names), *skister* and *etuitje*. Vowel Degemination has exceptions such as *cartoonist*, *keeper*, *engineering*, *screenen*, *zoomen* where the absence of degemination signals an unusual pronunciation; <ee> and <oo> correspond to /i/ and /u/ instead of /e/ and /o/. Consonant Doubling has exceptions such as *cabaretier*, *snobisme*, *spioneer*. Except for words such as *screenen*, *zoomen* and *spioneer*, these words could have been adopted as complex words with an exceptional spelling. In the case of *cabaretier* this is the most likely option, since the relevant affixes are not productive in Dutch. Exceptional pairs such as *spion-spioneer* are discussed in Appendix H.

Many monomorphemic words also seem to have been derived by the regular application of autonomous spelling rules such as Vowel Degemination, Consonant Doubling and Diaeresis Placement, but there are also exceptions. In (39) there are examples of regular and irregular words: [27]

(39) **spelling rule**
       Vowel Degemination
       fol**i**o, j**u**ra   ~   shamp**oo** (*shamp**o**), baz**oo**ka (*baz**o**ka)
       Consonant Doubling
       le**mm**a, ba**ll**on   ~   ca**b**aret (*ca**bb**aret), na**s**i (*na**ss**i)
       Diaeresis Placement
       na**ï**ef, ru**ï**ne   ~   muse**u**m (*muse**ü**m), optici**e**n (*optici**ë**n)
       Spelling Devoicing
       cursie**f**, precie**s**   ~   wit**z** (*wit**s**), qui**z** (*qui**s**)

In order to find out whether the similarity between native and non-native words in (39) means that autonomous spelling rules apply to all words alike, I will take a closer look at one rule, namely Consonant Doubling, which is repeated below:

(40) **Consonant Doubling**
            |

---

[27] In some words we find the sequence *-ciën: auspiciën, anciënniteit* but: *ancien*. Similarly, *-uüm* does have a diaeresis: *vacuüm*, unlike *-eum* and *-eus* (and *-uus: obliquus*).

$$0 \rightarrow C_i \: / \: C_0 \: V \: \_ \:]_S \: [_S C_i$$

The distribution of consonant geminates and single consonants in non-native words is illustrated under (41). The words in (41a) seem to have undergone (41). Most words are of this type. The examples in (41b) are exceptions to (41) and the words in (41c) have geminates in a context that does not satisfy the conditions of (41):

(41) a **regular according to (40)**
        ante**nn**e, bro**cc**oli, ce**ll**o, gru**tt**o, ma**mm**oet, o**ff**icier, ra**bb**i, te**rr**ein, ze**pp**elin, eta**pp**e, ge**kk**o, ma**ss**a, etc.
      b **irregular (exceptional single consonant letters)**
        acce**l**ereren, ca**b**aret, na**s**i, impre**s**ario, spi**n(n)**aker, co**m**ité, i**m**age, etc.
      c **irregular (exceptional geminates)**
        to**ss**, ja**zz**, ye**ll**, **Ll**oyd, croi**ss**ant, para**ll**el, regi**ss**eer, concu**rr**eer, sa**ff**raan, a**bb**revieer, shu**tt**le, ma**ngg**is

Option A' implies that there is no doubling in non-native words so that (42b) is not irregular, while option B' implies that doubling is applied to non-native words and native words alike, so that (42a) is regular. Once again the description that violates the Principle of Etymology is most attractive, since it has fewer exceptions. Some idiosyncratic cases such as those in (42b-c) are discussed in Appendix F.

   Another reason to prefer option B' is that there are indications that words such as *nasi* are exceptional. As observed in 4.4, we expect no regularization when non-native words are considered to be inherently outside the Dutch spelling system. However, spelling changes that can be interpreted as the removal of exceptions to autonomous spelling rules have in fact occurred (since they were non-preferred spelling variants, *taffia* and *dessa* were abolished in 1995):

(42) mafia, rafia → maffia, raffia   tafia   → taffia/tafia
      desa   → desa/dessa   gramophone   → grammofoon
      albatross   → albatros   mandrill   → mandril

Another example of a spelling change caused by autonomous rules applied to non-native words concerns the introduction of diaeresis, for instance in the following examples from Van der Sijs (1996): *druides* (Latin) → *druïde*, *ruiner* (French) → *ruïneren*, *vacuum* (Latin) → *vacuüm*.

   The presence of 'spelling pronunciations' also supports option B'. Consider, for instance, words written with geminate consonants after an originally long vowel that often give rise to a pronunciation with a short vowel. On the other

hand, words without geminates are often pronounced with long vowels even when they are originally short or derived from a language without a length contrast:

(43)   **spelling   original pronunciation changed pronunciation**
        a   villa   [vila]   [vɪla]
        b   rotan   [rɔtɑn]   [rotɑn]
            Donald   [dɔnəlt]   [donɑlt]

This phenomenon suggests that native speakers assume that the spelling of non-native words is not inherently irregular, but forms an accurate representation of the pronunciation, even where what seems to be the result of a spelling rule actually is an etymological spelling.

Another indication that native speakers do not consider the spelling of non-native words to be inherently irregular is formed by the behaviour of words with exceptional geminates. We write inflected forms of a verb such as *stressen* as *strest*, *streste*, *gestrest* (not: **stresst*, **stresste*, **gestresst*), but a noun such as *stewardess* is not written as **stewardes*. This contrast could be explained as follows: if a word has a geminate in the canonical form where it satisfies the conditions of Consonant Doubling (which is the case with verbs: *stressen*), it is reinterpreted as the result of doubling, and the geminates are replaced by single consonants where gemination is not applicable (*strest*). However, in the canonical form of the noun (*stewardess*) it is obvious that the geminate is not the result of doubling (this can also explain *screenen-screen*, not **screnen*). The contrast between *stewardess* and *strest* thus also suggests that geminates are interpreted to be the result of Consonant Doubling.

A third indication that regular patterns are attributed to autonomous spelling rules is formed by the behaviour of the inflected forms of *cursief*, *tarief*, *abrikoos*, *abuis* and *turkoois*. These words were adopted from French, see Van der Sijs (1996). In the case of *cursief-cursieve*, derived from French *cursif-cursive*, the voiced fricative of the inflected form can be borrowed as well, but in *tarief-tarieven*, *abrikoos-abrikozen*, *abuis-abuizen* and *turkoois-turkooizen*, the <z> or <v> in the inflected forms cannot be derived from French *tarif* (*tarifer*), *abricots*, *abus* (*abusif*) and *turquoise*, and must be the effect of interpreting <s> and <f> to be the result of devoicing.

I conclude that autonomous spelling rules apply to all words and that spelling phenomena that can be interpreted as being the effect of autonomous spelling rules, are treated that way. There is a residue of exceptions, such as *bazooka* (without Degemination) and *opticien* (without Diaeresis), but these form a minority and do not prevent speakers of Dutch to treat regular non-native words the same way as native words. This implies that A' is not an accurate statement about Dutch spelling.

In this context it should be noted that the representation of vowel length in Dutch is fortunate. Firstly, consonant geminates ensure that the length difference is visible in *liter-bitter*. Secondly, the present system has its advantages with respect to the incorporation of loan words. Most loan words of Latin or Greek origin fit in the present system with little adaptation: only single letters in final syllables are sometimes adjusted, for instance in *bazaar* (*bazar*), *natuur* (*natura*), *school* (*schola*) and *fideel* (*fidelis*).

In 5.4, we have seen that in the case of phoneme-to-grapheme conversion rules the rejection of the Principle of Etymology is also supported by a separate non-native regularity. However, this does not hold for autonomous spelling rules. This does not imply that there are autonomous spelling rules that are only relevant to non-native words, e.g. <i>/<ie> Alternation and Orthographic Diminutive Allomorphy. The first rule, <i>/<ie> Alternation, changes <i> into <ie> in words such as *skietje* (from *ski*). The second rule changes <é> into <ee> in words such as *cafeetje* from *cafè*). However, these rules do not constitute counterexamples to the claim that autonomous spelling rules are valid for the whole lexicon, because they do not treat similar letter sequences in native and non-native words differently. Native words do not end in <i> or contain <é>, so these rules are not applicable. Wherever the rules for native words are relevant, they are applied. For instance, the only complex consonant cluster that occurs in native words, <st>, is syllabified in non-native words just as in native words, although <st> could be syllabified as an onset in words such as *poesta*, cf. *mees-ter ~ poes-ta*. It appears that autonomous spelling rules are not sensitive to the distinction between native and non-native words.

We can conclude that option B', a spelling system that is not in accordance with a strict interpretation of the Principle of Etymology, is to be preferred, since it gives the most restrictive description of the spelling facts, and since it is compatible with spelling changes. However, there is no evidence that suggests that autonomous spelling rules apply differently in native words and non-native words. In the approach chosen here, differences between the spelling of native words and non-native words are exclusively accounted for by phoneme-to-grapheme conversion rules, and not by autonomous spelling rules. In a way, the fact that only phoneme-to-grapheme conversion rules are sensitive to the distinction between native words and non-native words supports the postulation of two distinct types of rules.

5.9 Mutually exclusive properties of the two types of spelling rules

In 5.1, I formulated three generalizations about autonomous spelling rules on the basis of the cases encountered so far, i.e., Vowel Degemination, Consonant Doubling, Consonant Degemination and Spelling Devoicing: autonomous

spelling rules are insensitive to phonological context, they are sensitive to orthographical context, and they apply across morpheme boundaries. I will now examine whether these generalizations also hold for the other autonomous spelling rules discussed in this chapter.

Autonomous spelling rules are insensitive to phonological context
Sections 5.3–5.6 showed that the autonomous spelling rules enumerated in (3) can be formulated as generalizations about letter sequences, so they are indeed independent of the pronunciation. To be able to formulate these autonomous spelling rules, it is necessary that spelling representations indirectly refer to certain properties of the pronunciation. However, the new formalism is not merely a notational variant of an account in which all spelling rules refer to the pronunciation. It is crucially different in two respects.

The first difference is that the properties of sound representations that are (indirectly) relevant to the proper application of spelling rules are restricted. Autonomous spelling rules can only distinguish between letters that do or do not form a grapheme (*la-chen* versus *vin-ger*), between long and short vowels (*bijter* versus *bitter*), and between vowels representing schwa or /ɛ/ (*twijfelen* versus *bellen*). All other properties of the pronunciation are irrelevant to the application of these rules (with one exception, stress, that will be discussed below).

The second difference is that relevant properties of the pronunciation are encoded in spelling by a mapping procedure, so we can account for those cases where spelling properties differ from those expected on the basis of the (synchronic) sound representations. For instance, the fact that the letter combination <ng> is not treated as a grapheme although these letters encode the phoneme [ŋ] can be accounted for by the fact that the relevant phoneme is mapped onto two graphemes.

There is one case for which the new approach is inappropriate: the distinction between stressed and unstressed vowels is relevant to the distribution of <i> and <ie>, cf. *oliën* versus *genieën*, see 5.6.1. It should be noted, that the fact that stress plays a role here is based on an incorrect generalization by Te Winkel. The relevant spelling rule can only be formulated as an autonomous spelling rule if spelling representations contain information about stress.

Autonomous spelling rules are sensitive to orthographical context
We have already seen some indications that autonomous spelling rules are sensitive to orthographical context. Vowel Doubling, Replacement of <i> by <ie>, Apostrophe Placement and Diaeresis Placement are blocked by the presence of hyphens. Diaeresis Placement is sensitive to the effect of Vowel Degemination, as illustrated by the contrast between *prieel* and *priëlen*. Finally, the presence or absence of a diacritic depends on the spelling of a vowel as a single letter (<a>) or digraph (<oe>), cf. *pa ~ pa's* versus *moe ~ moes*.

Obviously, phoneme-to-grapheme conversion rules that have the sounds as input, cannot easily account for interactions of spelling phenomena or for contrasts that can only be characterized in terms of letter sequences, whereas rules that affect letter sequences can.

Autonomous spelling rules apply to domains larger than the morpheme
Autonomous spelling rules are not restricted to the morpheme domain. The Alternation of *i* and *ie*, Vowel Doubling and Orthographic Diminutive Allomorphy are even exclusively applied to complex words. The examples in (45) show that the rules would derive the incorrect spelling if they were restricted to morphemes:

(44) **Rule**   **structure**   **rules applied**   **rules applied**
                                               to morpheme   **to word**

Hyphenation	heks+en	*heks-en	hek-sen
Apostrophe	pa+s	*pas	pa's
Diaeresis	ski+en	*skien	skieën
i/ie	ski+de	*skide	skiede
Vowel Doubling	kano+tje	*kanotje	kanootje
Diminutive Allomorphy	depot+tje	*depottje	depootje
ng/n	koning+kje	*koningkje	koninkje

The fact that autonomous spelling rules apply to domains larger than the morpheme does not imply that these rules ignore all morphological information.

The rules mentioned in this study are restricted to monomorphemic and inflected words and derivations. They do not operate across the boundaries of compounds or word groups as illustrated in (45) (where '#' indicates a compound boundary):

(45)   **Role of morphological structure**

Consonant Doubling:	i**n**ademen (in#ademen)	bi**nn**en
Vowel Degemination:	w**aa**nidee (waan#idee)	gr**a**niet
<ng> → <n>	za**ngk**unst (zang#kunst)	koni**nk**je
Syllabification:	zoen**.**of.fer (zoen#offer)	zoe**.**nen

Diaereses are added to monomorphemic words and their inflected and derived forms only; at compound boundaries hyphens are inserted instead, as illustrated by the contrast between the inflected word *oliën* (*olie+en*) and the compound *olie-embargo* (*olie#embargo*).[28] Autonomous spelling rules are restricted to

---

[28] In all cases where diaeresis would be inserted in non-compounds, hyphens are (obligatorily) inserted at compound boundaries. In addition, hyphens are inserted to separate <ii> and <ij> in

prosodic words. However, when a suffix is attached to a compound or word group, spelling rules do look across the boundary that separates the compound or word group from the suffix *niet-roker* (**niet-rooker*), *fijndradig* (**fijndraadig*). Apparently, the depth of embedding is irrelevant to spelling rules. This is also suggested by the fact that right-branching and left-branching compounds are treated the same way, cf. [[zaal[voetbal]]en] ~ *zaalvoetba**ll**en* and [[[opblaas]bal]en] ~ *opblaasba**ll**en*. Another way in which morphological structure is relevant is that Vowel Doubling and Apostrophe Placement are only applied to the inflectional suffix *-s* as was illustrated in (28).

Autonomous spelling rules apply to native and non-native words alike
In the preceding section we have found yet another property common to autonomous spelling rules: they treat native and non-native words in the same way. If we compare these properties to those of phoneme-to-grapheme conversion rules, we see that both rule types have mutually exclusive properties:

(46)	**phoneme-to-grapheme conversion rules**	**autonomous spelling rules**
context	phonological	orthographical
domain	morpheme	word
native/non-native sensitive	yes	no

This implies that the two-level model is more restrictive than a model with one set of rules only: in a spelling model with only one set of rules, any property displayed by one rule must be allowed for all spelling rules, and one must stipulate which rules have which properties.

5.10 Concluding Remarks

It is an improvement to consider spelling alternations discussed in this chapter to be the result of autonomous spelling rules for the following four reasons.
Firstly, his approach accounts for the fact that the alternations are different than would be expected on the basis of the pronunciation, especially phonological syllable structure. Only autonomous spelling rules can derive the correct spelling and hyphenation patterns for words such as *baaierd*, *lafaard*, *snoodaard*, *grijsaard*, *zin-gen*, *mar-xist* and *glin-ster* as well *mees-ter*. Sound-based spelling rules would incorrectly derive **bajerd*, **laffaard*, **snodaard*, **grijzaard*, **zi-ngen*, **marxist* (no hyphenation possible) and either *glin-ster* and **mester* or

---

examples such as *anti-intellectueel* and *gummi-jas*. Hyphens may be optionally inserted at compound boundaries to clarify words such as *pret-ogen* (see [*Woordenlijst 1995*], p. 33), but they must not be used in derivations such as *chicheid*.

glins-ter and *meester*. Consequently, the approach proposed is more adequate.

Secondly, We can restrict the descriptive power of spelling rules, since phoneme-to-grapheme conversion rules and autonomous spelling rules have mutually exclusive properties. There is one alternation that seems to exceed the restrictions imposed on autonomous spelling rules, i.e., the stress-sensitive rule of replacement of <ie> by <i>. For this rule, stress information must be added to the spelling representation.

Thirdly, as mentioned in chapter 2, the use of autonomous spelling rules removes a number of counterexamples to the claim that spelling encodes the spelling of morphemes. The alternations discussed in this chapter concern domains larger than a single morpheme, as illustrated in (44).

Finally, pelling rules become simpler. Obviously, phoneme-to-grapheme conversion rules are less complex, since they need not account for all spelling phenomena, as shown in chapter 2. A spelling rule that accounts for a given spelling alternation is often simpler as an autonomous spelling rule than it would be as a phoneme-to-grapheme conversion rule, since it applies to easily defined sets of letters rather than to arbitrary sets of phonemes. For instance, Vowel Degemination applies to geminate vowel letters, Apostrophe Placement and Vowel Doubling apply to single vowel letters linked to two V-positions, Diaeresis Placement applies to heterosyllabic letter combinations that can occur in the same syllable. With autonomous spelling rules we can account for the fact that spelling alternations interact. For instance, Vowel Doubling and Apostrophe Placement are blocked by the presence of hyphens. It would be difficult to describe such phenomena by means of phoneme-to-grapheme conversion rules.

An apparent disadvantage of the use of autonomous spelling rules is the fact that it implies that we need spelling rules that resemble sound rules, such as Consonant Degemination, Spelling Devoicing and Syllabification. This duplication could suggest that a generalization is missed. This, however, is not the case, since autonomous spelling rules are only proposed for those spelling alternations that crucially differ from the sound alternations, so the introduction of autonomous spelling rules is well-motivated.

I conclude that Dutch spelling can be described more adequately with two rule types than with phoneme-to-grapheme conversion rules only. The findings of this chapter suggest that spelling is more than merely a code for sound representations. Although many of the autonomous spelling rules can be considered to be indirectly motivated by the need to provide a code for spoken words since they improve readability by counteracting ambiguity, other rules do not seem to serve this purpose. An overview is given in (47):

(47)   a   **Autonomous spelling rules that counteract ambiguity**
              Vowel Doubling, $i \rightarrow ie$, Diaeresis Placement, Apostrophe Placement,
              Consonant Doubling

b **Autonomous spelling rules that do not counteract ambiguity**
   ie → *i*, Vowel Degemination, Consonant Degemination, Spelling
   Devoicing, Alternation of *ng* and *n*, Diminutive Allomorphy

It appears that the application of the rules in (48b) cannot be indirectly motivated by the need to provide a readable code for spoken words. Apparently the rules in (48b) apply to satisfy orthographic well-formedness restrictions such as the ban on tautosyllabic geminates. I will refer to such rules as 'graphotactic rules'. Consequently, orthography is not merely a code for the pronunciation, but also a rule system of its own with independent well-formedness conditions.

Chapter 6

Conclusions and suggestions for further research

This study reports on an extensive and systematic investigation of Dutch spelling. I examined how the current spelling of Dutch words can be derived most adequately from the pronunciation. For that purpose, a set of rules that convert the sounds of words to their spelling was compiled and applied to the pronunciation of words by means of a computer programme. The words were a selection from the *CELEX* (*CEnter for LEXical information*) database. The starting point of the set of spelling rules was formed by traditional spelling rules, such as the rules in the Dutch spelling dictionaries, see [*Woordenlijst 1954*], [*Woordenlijst 1995*], as well as rules taken from previously published descriptive accounts. In addition, I took into account the four spelling principles that govern the application of these rules according to Te Winkel. These rules and principles were subsequently modified in order to improve their accuracy. The result is a description of the Dutch spelling system in the form of rules and exceptions. These rules can be used for the teaching of writing and for the improvement of the consistency of the spelling in dictionaries, as well as for language processing applications. The investigation has led to the following conclusions on Dutch orthography and suggestions for further research.

6.1 Conclusions

6.1.1 The optimal organization of spelling rules

The starting point of the investigation was a simple model for the derivation of spelling from sound representations based on observations in the literature:

(1) | abstract sound representations |

phoneme-to-grapheme conversion rules

↓
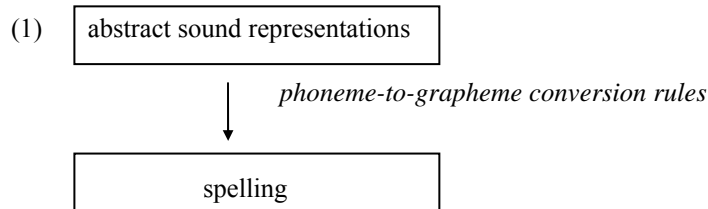
| spelling |

This model suggests that all (predictable) variation in the spelling of a sound is the reflection of variation in the pronunciation, or at least caused by phoneme-to-grapheme conversion rules that are conditioned by phonological context.

Closer examination of the data showed that not all spelling distinctions encode sound distinctions; alternations such as those of single and double letters are predictable, but on the basis of letters rather than sounds. For instance, *raam-ramen* and *laf-laffe* seem to be predictable on the basis of phonological syllables, but in some cases this assumption leads to incorrect results: **a-ien, *gochem, *laffaard, *pochchen*. Model (1) thus cannot optimally account for such purely letter-based alternations. For this reason, I supplemented model (1) with autonomous spelling rules, which do not directly refer to the pronunciation. This gives the following model of spelling computation:
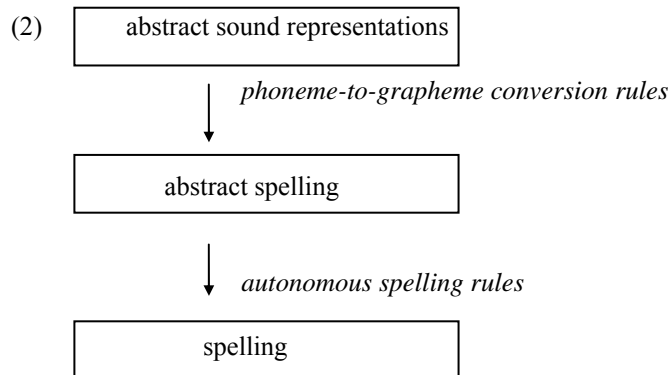
(2) | abstract sound representations |

phoneme-to-grapheme conversion rules

↓

| abstract spelling |

↓ *autonomous spelling rules*

| spelling |

Phoneme-to-grapheme conversion rules neutralize some distinctions that are relevant to the proper application of rules such as Vowel Degemination, e.g. the sounds /ɛ/ and /ə/ are both represented by <e>. I therefore assume that these distinctions are

encoded in underlying spelling representations. In line with current practice in phonology this was done by linking letters to an orthographical CV-tier.

The introduction of a second rule type enables us to formulate simpler phoneme-to-grapheme conversion rules, since they need not account for all types of spelling variation. This implies that the relation between underlying phonemes and underlying graphemes comes closer to an ideal one-to-one relation than was the case in earlier accounts, and that the role of context in phoneme-to-grapheme conversion rules is reduced to some historically motivated cases (at least as far as native words are concerned) such as the choice between <uw> and <w> in *wiek-kieuw*.

By describing some spelling alternations on the basis of letter sequences only, we get a simpler account that offers more insight, since it is easy to define which letters undergo the rules. The alternation in the spelling of long vowels, for instance, holds for an arbitrary subset of sounds, but for an easily defined set of graphemes, i.e., geminates. Consequently, describing these alternations as the result of autonomous spelling rules is simpler than it would be as the result of phoneme-to-grapheme conversion rules. With autonomous spelling rules we can also account for the fact that spelling alternations interact. For instance, Vowel Doubling and Apostrophe Placement are blocked by the presence of hyphens. Only one type of inconsistency remains, i.e. the contrast between *bourgeois* (*bourgeois+s*) where <s> is deleted after a silent letter and *pince-nezs* where it is not, see 5.3.

Another advantage of the use of autonomous spelling rules is that it allows for a restrictive spelling account. The two rule types have the following mutually exclusive properties:

	phoneme-to-grapheme conversion rules	**autonomous spelling rules**
context	phonological	orthographical
domain	morpheme	word
native/non-native sensitive	yes	no

In a spelling model that uses one set of rules, any property displayed by one rule must be allowed for all spelling rules. The new model restricts the aspects of the pronunciation that are (indirectly) relevant to the proper application of autonomous spelling rules: they can no longer refer to the pronunciation of words, only to those aspects of the pronunciation that are encoded in orthographical structure.

With the aid of model (2) the spelling of most native words is predictable from their pronunciation. An exception is formed by the competing spelling variants for the sounds /ɛi/ and /ɑu/, e.g. *wij-wei*, *boud-baud*. The default spelling of these

sounds in native words is <ij> and <ou>; words with <ei> and <au> are listed as exceptions. In addition, there are some individual exceptions such as *ambt*, *erwt*, *misschien*, **oir**, **th***uis*, see Appendix B. Another exception is formed by the spelling of words such as *viezer***ik***en*. The absence of doubling shows that <i> is treated as if it encodes a schwa here, see Appendix F.

Spelling encodes morphemes
Examination of complex words revealed that the spelling of written morphemes is often constant in related words. Various accounts for this phenomenon have been proposed, for instance that spelling aims at providing a uniform spelling of morphemes, or that spelling represents one underlying or intermediate level in the sound derivation. Both proposals have the disadvantage that they are not compatible with all types of spelling variation that occur in Dutch spelling. If it is assumed that spelling aims at uniformity, we cannot account for the fact that the application of autonomous spelling rules introduces spelling variation. If we assume that spelling represents an intermediate level in the derivation of the pronunciation we cannot account for the fact that the effect of the same rule, Nasal Assimilation, is visible within morphemes but not across morpheme boundaries, as illustrated by pairs such as *ra**m**p-i**n**pakken* (*in+pakken*). The fact that phoneme-to-grapheme conversion rules are restricted to morphemes, cf. *aaien* (*aai+en*) *~bajes*, is not accounted for.

For this reason, I conclude that the uniform spelling of morphemes can be best accounted for by the assumption that spelling encodes the abstract sound representation of morphemes. Thus, (2) can be made more precise:

(3) | sound representation of morphemes |

↓ *phoneme-to-grapheme conversion rules*

| spelling of morphemes |

↓ *autonomous spelling rules*

| spelling of words |
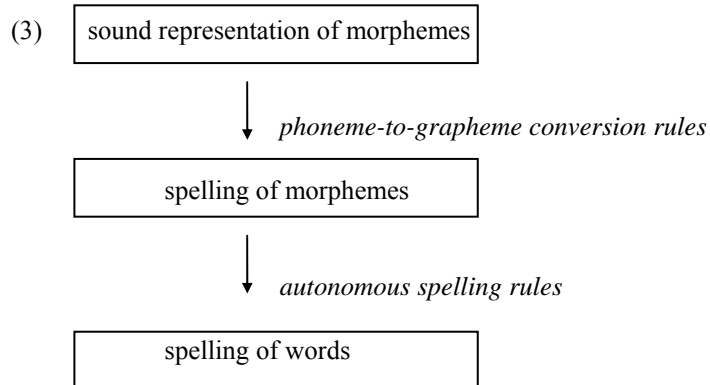
The assumption that phoneme-to-grapheme conversion rules are restricted to the morpheme domain allows the representation of morpheme-internal effects of sound rules and is compatible with variation introduced by autonomous spelling rules. It is also compatible with contrasts such as *ramp-inpakken*, while it still accounts for the uniform spelling of *aai-aaien* and *hond-honden*. It should be noted that combinations of non-native stems and non-native affixes do not form domains for spelling rules, cf. contrasts such as *markiezen-markizaat*.

Model (3) predicts that spelling abstracts from the effect of sound rules that operate across morpheme boundaries. However, this expectation does not always prove correct. For instance, the effect of Diminutive Allomorphy is visible in spelling. These facts can be accounted for by the assumption that spelling represents competing allomorphs, rather than one abstract underlying morpheme. Alternatively, we could account for these data by the Readability Requirement, since abstracting from these rules would result in a spelling that could be read incorrectly.

Apparent counterexamples against the claim that spelling encodes the pronunciation of morphemes, i.e., alternations such as *lieve-lief* and *haatte-gehaat*, could be given an alternative account by describing them by autonomous spelling rules. In some individual cases it is assumed that the effect of sound rules is lexicalized. This explains why we write [bɛltənɪs] and [lifələk] as *beeltenis* and *liefelijk* rather than as **beeldenis* and **lievelijk*. A remaining inconsistency is the fact that [draxələk] is written as *dra**g**elijk*, not **draa**ch**elijk* (see chapter 3).

Separate phoneme-to-grapheme conversion rules for non-native words
Another important conclusion about Dutch orthography is the following. The spelling of loan words is not always exceptional when compared with native words. In various words a new, hybrid spelling pattern has evolved, based on the combination of etymological spelling patterns and partial, context-dependent spelling adaptation. When we exploit these generalizations in a special set of phoneme-to-grapheme conversion rules for non-native words, the spelling of many loan words becomes predictable. In other words, phoneme-to-grapheme conversion involves two sets of rules; simple rules for native words and more complex rules for non-native words. There are few exceptional native words, but many non-native exceptions.

Results
The sets of spelling rules proposed in this study accurately predict the spelling of 85% of the 45,821 entries in the test lexicon. If we look at the 24,288 native words only, the rule set correctly predicts the spelling of 95% of the words. Of the incorrectly-spelled words, 13% are loan words that only show their origin by the

spelling and can therefore not be recognized as non-native by the criteria in appendix B, e.g. *cluster* and *bis*. Sixty-eight per cent of incorrect predictions consists of words in which /ɛi/ and /ɑu/ are written as *ei* and *au* instead of the proposed default spelling *ij* and *ou*, e.g. *weiger*, *klauter*. The remaining 19% of the incorrect predictions is formed by individual exceptional morphemes such as *erwt*, *oirbaar*, *-ik* etc.

The spelling of the 21,126 non-native words (i.e., formally recognizable loan words) is less predictable. However, we still derive the correct spelling for 73% of the words. Even though the spelling of many non-native words is idiosyncratic, the spelling of derived forms is largely predictable from the constituting parts. This may be accounted for by the fact that autonomous spelling rules do not distinguish between native and non-native words. Onl2% of the exceptional non-native words is caused by autonomous spelling rules, e.g. in *tonic* (no Consonant Doubling), *screenen*, *stress* (no Degemination), *fez* (no Spelling Devoicing), *museum* (no Diaeresis Placement), *etuitje*, *skister* (no replacement of *i* by *ie*):

Native words	Non-native words
95% regular	73% regular
5% irregular	27% irregular
• 68% words with ɛi/ɑu	• 98% due to phoneme-to-grapheme conversion rules
• 13% formally native loan words	• 2% due to autonomous spelling rules
• 19% other exception	

It should be noted that the description proposed a relatively small set of rules, see Appendices C, D and E. More accuracy could be obtained by increasing the number of rules or by storing the spelling of morphemes, but at the expense of insight.

6.1.2 Orthographic Principles

Traditionally, the relation between orthography on the one hand and phonology and morphology on the other is described by the Phonological Principle (spelling is a code for phonemes), the Morphological Principle (spelling provides a uniform spelling for morphemes as far as the pronunciation allows this, e.g. *branden-brand/*brant*) and the Etymological Principle (the choice between competing spellings is made on the basis of an earlier pronunciation, sometimes in the donor language, e.g. *wij-wei*, *therapeut*). After the examination of Dutch orthography, it is

possible to evaluate the extent to which these principles are accurate statements about the present spelling system. It appears that some generalizations about Dutch orthography have not yet been captured by these traditional principles, which must therefore be somewhat modified and supplemented by a new principle that governs the application of autonomous spelling rules.

First consider the **Phonological Principle** that states that spelling encodes phonemes. This principle leads to the expectation that the same phoneme is written in the same way unless deviations are induced by the Morphological or Etymological Principle. The present investigation shows that there are two types of spelling variation in native words that are not accounted for by one of these principles:

1. alternations such as the insertion of <w> after <ou> and <au> (*vrouw-hout*), and the alternation of <i> and <j> (*ja-aai*), see 2.5.
2. phenomena such as the alternation of single letters and geminates (*raam-ramen*, *pak-pakken*), see 2.4.

The first type of spelling variation does not really contradict the Phonological Principle: the spelling can still be described as a code for the pronunciation, but the necessary phoneme-to-grapheme conversion rules are context-sensitive rather than context-free (in 2.5 it was suggested that variation of the type *trouw-koud* may be related to the Morphological Principle in an indirect way). However, in 2.4.2 it was argued that variation of the last type, e.g. *raam-ramen* and *pak-pakken* cannot be described as a function of encoding the pronunciation. The variation is predictable, but on the basis of orthographical rather than phonological context. I proposed to account for these facts by autonomous spelling rules that apply to satisfy certain graphotactic conditions. The Phonological and Morphological Principle are sometimes violated by these alternations. This suggests that a higher-ranking principle forces these violations. However, Te Winkel did not formulate such a principle. I therefore propose a new orthographic principle, the **Graphotactic Principle**: the spelling of words must satisfy certain graphotactic conditions.

The **Morphological Principle** states that spelling aims at providing a uniform code for morphemes as long as the pronunciation allows this. However, we have seen above that this only holds for phoneme-to-grapheme conversion rules. Autonomous spelling rules introduce variation across morphemes, even when this does not serve to increase readability.

The **Etymological Principle** accounts for the presence of unpredictable spelling variation such as *rijk-reik*. Although these exceptional spellings may be explained by their history, this explanation is of no use for a synchronic description of Dutch spelling. In native words this principle thus functions as an exception mechanism. In loan words a strict interpretation of the Etymological Principle would suggest that

such words keep the spelling of the donor language and that their spelling is inherently irregular rather than derived by rules. However, we have seen that their spelling can in many cases be predicted by a separate set of spelling rules. This means that the accuracy of spelling rules can be improved by distinguishing two sublexicons, native and non-native words and applying separate rule sets to both sublexicons. The Etymological Principle is thus reformulated as follows:

a. The spelling system has some historically motivated exceptions.
b. Native and non-native words are subject to different sets of phoneme-to-grapheme conversion rules.

Finally, the role of the **Readability Requirement** must be extended. This requirement does not only govern the interaction of the Morphological Principle with the Phonological Principle, but also accounts for exceptions to the Graphotactic Principle. For instance, one of the conditions that spelling must satisfy is that there are no tautosyllabic geminates, but these do occur in a word such as *raam*. If degemination were applied to this word in order to satisfy the graphotactic constraint, we would get a spelling that would suggest *[rɑm].

6.1.3 The optimal transcription of lexical sound representations?

Chomsky (1970:4) assumes that 'conventional English orthography in its essentials appears to be a near-optimal system for representing the spoken language' (Chomsky 1970:4). This statement is motivated by the fact that both spelling and underlying sound representations abstract from predictable sound variation. A similar claim has been made for Dutch by Wester (1987). The present investigation of the spelling system of Dutch has shown this claim to be untenable. Although the spelling often resembles lexical representations, it cannot be maintained that spelling is an optimal representation thereof.

Firstly, spelling is not optimal since it lacks sufficient symbols to uniquely encode all phonemes. For instance, the contrast between full vowel and schwa and between /ŋ/ and /ŋɣ/ is neutralized in the written forms *bevel-nevel* (bəvɛl]-[nevəl]) and *slingeren-fingeren* ([slɪŋərə]-[fɪŋɡerə]). Even if there is a one-to-one relation between written morphemes and lexical representations after phoneme-to-grapheme conversion, this one-to-one relation is often undone by the application of autonomous spelling rules. This leads to variation in related words such as *raam-ramen*, *neurie-neuriën*, *ski-skietje*, *koning-koninkje* etc. Therefore written words do not provide an invariant code for lexical representations any more than phonological

realizations do.

Secondly, spelling does not always correspond to lexical representations, since it encodes the morpheme-internal effect of (predictable) sound rules. For instance, we write *ramp* rather than **ranp*. Unless a competing allomorph analysis is assumed, this also holds for sound rules that operate across morpheme boundaries, such as Diminutive Allomorphy.

Finally, in some cases the uniform spelling of morphemes in related words is not caused by the fact that it represents an abstract sound representation, but by the fact that it encodes a historical pronunciation. In these cases spelling deviates from the lexical representation that speakers use to synchronically derive the pronunciation as well as the spelling of words. Appendix G revealed serious problems caused by the postulation of abstract phonological representations to account for pairs such as *provoceer-provocatie* ([provoser]-[provokatsi]), *actie-actief* ([ɑksi]-[ɑktif]), *solutie-resolutie* ([solytsi]-[rezolytsi]). Conversely, in 2.5 we have seen that variation of the spelling in words such as *wei-wij* and *louw-lauw* cannot be accounted for by the assumption that there still is an underlying sound distinction that is neutralized by sound rules. Instead, the neutralization must be seen as a diachronic process, and the spelling variation is arbitrary from a synchronic point of view.

Thus, written words do not constitute a more accurate representation of lexical representations than spoken words.

6.1.4 Relevance for alphabetic spelling systems in general

This study deals exclusively with the spelling of Dutch. However, the findings also allow us to make some remarks regarding the nature of alphabetic spelling systems in general.

It has been shown here that Dutch orthography is best described by using autonomous spelling rules as well as phoneme-to-grapheme conversion rules. This implies that (alphabetic) spelling systems are not necessarily strictly derivative. At least in the case of Dutch, spelling also is a linguistic module in its own right. Like phonological rules, spelling rules apply to satisfy certain well-formedness conditions. Other languages also have spelling alternations that do not encode sound differences, e.g. English *orthography-orthographies*. The use of autonomous spelling rules may improve the description of these phenomena.

The existence of autonomous spelling rules also has implications for the evaluation of the complexity of a spelling system. If spelling is considered strictly secondary to speech, its complexity consists of two factors. The first factor is the extent to which sound-spelling relations are predictable, and the complexity of the

phoneme-to-grapheme conversion rules. The second factor is formed by the distance between spoken words and the level of phonological representation that is encoded in written forms. A spelling system that abstracts from sound alternations is more complex than one that encodes less abstract levels of sound representation, see for instance Frost (1992) (the preceding section showed that 'deep' orthographies may not reflect lexical representations but historical pronunciations). However, the introduction of autonomous spelling rules brings in a third factor: the number and complexity of these rules contributes to the complexity of spelling systems. Even a spelling system with few and simple phoneme-to-grapheme conversion rules and a small distance between spoken forms and the level encoded in spelling may be complicated by intricate autonomous spelling rules.

Finally, the fact that there are two sets of phoneme-to-grapheme conversion rules in Dutch, one for native and one for non-native words, suggests that there is yet another factor that contributes to the complexity of a spelling system: the extent to which the lexicon is divided into different sets of words that are subject to different rules.

6.2 Suggestions for further research

This study presents the first systematic and explicit account of the spelling of Dutch words. Some questions have not been answered yet, because they were outside the scope of the present investigation or because they require more research. During the investigation some questions with respect to phonology and morphology were raised that deserve to be examined in more detail.

Further spelling research
Firstly, it remains to be established whether the spelling rules formulated here, which are generalizations on the basis of the existing vocabulary, are productively used by language users and have psychological reality. This could be tested by experiments in which speakers are audibly presented with new words, names or nonsense words and asked to write them down. By comparing their spelling with the spelling predicted by the rules it is possible to find out if the rules proposed here are actually used. The description of Dutch spelling given here is a good starting point for such experiments since it is more explicit and complete.

Secondly, the spelling rules given here still need to be supplemented by a description of phenomena that were outside the scope of the present investigation, i.e., the spelling of compounds and word groups, and aspects of spelling that encode

certain syntactic or semantic aspects of language, such as the use of capitals and the placement of spaces.

Thirdly, the intricate interaction between Vowel Degemination and Consonant Doubling suggests that a description in the framework of Optimality Theory, see for instance Prince & Smolensky (1993), may increase insight into these phenomena. In this framework attention is shifted from the rules themselves to the constraints that are satisfied by the application of these rules. Constraints may be violated in order to satisfy higher-ranking constraints. Such an analysis may account for the fact that geminates within orthographic syllables are forbidden but do occur in words such as *raam* anyway, since the alternative **ram* would violate a higher-ranking Readability Requirement. Furthermore, by formalizing spelling principles as ordered constraints we may evaluate the optimality of the present spelling system, and predict future spelling changes.

Another question raised by the account of writing presented here is whether the findings of this study can shed new light on the formulation of letter-to-sound conversion rules. Should reading also be considered a two-step process in which the surface spelling is converted into the lexical representation before it is related to sounds? Does the distinction between native and non-native words play a role?

Finally, it would be interesting to compare Dutch orthography with the spelling of other languages, particularly to examine whether the description of the spelling of other languages also necessitates the postulation of a spelling model with autonomous spelling rules as well as phoneme-to-grapheme conversion rules. A comparison of spelling descriptions may pinpoint flaws of individual descriptions and eventually lead to a better understanding of alphabetical spelling systems.

Further phonological and morphological research
This study has also drawn attention to some areas of phonology where further research would be welcome.

The first question that is raised is whether all types of variation in the realization of loan words should be considered the effect of the synchronic application of sound rules. Examples where this is not self-evident are formed by rules that result in absolute neutralizations such as final devoicing in words such as *wodka* ([wɔtka]) and glide insertion in words such as *contractueel* (kɔntrɑktyel]). In the literature it is assumed that the difference between spelling and sound realization is caused by neutralization rules that operate in certain contexts only. However, it has not yet been systematically investigated whether speakers productively use these rules or whether the neutralization should be viewed as the result of a historical rule. Another case where we might not be dealing with productive rules is formed by the

neutralization of the length contrast in pairs such as *agressie-aggregaat* ([ɑɣrɛsi ~ ɑɣrəɣat]). The difficulty in determining the underlying quality of the first vowels in these pairs might suggest that there is no length contrast outside stressed syllables. There are only a few words where the spelling suggests that there is such a contrast, and the spelling of the relevant words may be idiosyncratic rather than the representation of a sound contrast.

A second topic that deserves to be further investigated is the distinction between native and non-native words. The criteria used here can possibly be further refined, and the role of this distinction in phonology and morphology can be further investigated.

A final question raised in this study is whether morpholexical variation is derived from abstract morpheme representations by sound rules. An alternative would be to assume that competing allomorphs are listed in the lexicon. This question may not receive the same answer for native and non-native morphemes. With respect to non-native morphology it would be useful to establish that diachronically related words are still felt as related by speakers, and whether judgements of all speakers agree in this respect, or whether there is variation in speakers' judgements, possibly related to the size of the individual vocabulary and knowledge of the spelling of words. After all, we have seen in Appendix G that certain spelling conventions ensure that the spelling of diachronically related words is more similar than their pronunciation.

Bibliography

Allen, J., M.S. Hunnicutt & D. Klatt (1987). *From Text to Speech: the MITalk System.* Cambridge, Mass.: MIT press.

[*Afrikaanse woordelys*]. *Afrikaanse woordelys en spelreëls in opdrag van de Suid Afrikaanse Akademie vir Wetenskap en Kuns verzorgd door de Taalkommissie.* Kaapstad: Tafelberg-Uitgewers Beperk, 1991.

Aronoff, M. (1992). Segmentalism in linguistics: The alphabetic basis of phonological theory. *The Linguistics of Literacy*, edited by P. Downing, S.D. Lima & M. Noonan, 69–82. Amsterdam-Philadelphia: John Benjamins.

Assink, E.M.H. (1983). *Leerprocessen bij het spellen: aanzet voor de verbetering van de werkwoordsdidaktiek.* Doctoral dissertation, University of Utrecht.

Back, O. (1995). Die deutsche Orthographie im europäischen Vergleich. *LernSprache Deutsch* **3**, 39–60.

Battus (1981). *Opperlandse taal- & letterkunde.* Amsterdam: Querido.

Bennis, H., A. Neijt & A. van Santen (1991). *De Groene Spelling.* Amsterdam: Bert Bakker.

Berendsen, E., S. Langeweg & H. van Leeuwen (1986). Computational phonology: merged not mixed. *Proceedings of the International Conference on Computational Linguistics '86*, 612–614.

Berendsen, E. (1988). De relatie tussen spelling en uitspaak. *Ter sprake: spraak als betekenisvol geluid in 36 thematische hoofdstukken*, edited by M.P.R. van den Broecke, 359–366. Dordrecht: Foris.

Berendsen, E., J. Lammens & H. van Leeuwen (1989). Van tekst naar foneemrepresentatie. *Informatie* **31**, 559–636.

Beurden, L. van (1987). Playing Level with Dutch Morphology. *Linguistics in the Netherlands 1987*, edited by F. Beukema & P. Coopmans, 21-30. Dordrecht: Foris.

Bezooijen, R. van & M. Gerritsen (1994). De uitspraak van uitheemse woorden in het Standaard-Nederlands: een verkennende studie. *De Nieuwe Taalgids* **87**, 145–160.

Booij, G.E. (1977). *Dutch Morphology. A Study of Word Formation in Generative Grammar.* Dordrecht: Foris.

Booij, G.E. (1981). *Generatieve fonologie van het Nederlands.* Utrecht-Antwerpen:

Het Spectrum.

Booij, G.E. (1984). Syllabestructuur en verkleinwoordsvorming in het Nederlands. *GLOT* **7**, 207–226.

Booij, G.E. (1985). Spelling als weergave van taal. *Visies op spelling*, edited by G. Verhoeven & E. Assink, 47–55. Groningen: Wolters-Noordhoff.

Booij, G.E. (1987). ʻThe Reflection of Linguistic Structure in Dutch Spelling. *Orthography & Phonology*, edited by P.A. Luelsdorff, 215–224. Amsterdam-Philadelphia: John Benjamins.

Booij, G.E. (1991). Spellingsysteem en fonologisch systeem. *Taalleerproblemen in het voortgezet onderwijs*, 27-36. Dordrecht: ICG.

Booij, G.E. (1995). *The Phonology of Dutch.* Oxford: Clarendon Press.

Booij, G.E., C. Hamans, G. Verhoeven, F. Balk & C.H. van Minnen (1979). *Spelling.* Groningen: Wolters-Noordhoff.

Booij, G.E. & J. Rubach (1987). Postcyclic versus Postlexcial Rules in Lexical Phonology. *Linguistic Inquiry* **18**, 1–44.

Booij, G.E. & A. van Santen (1995). *Morfologie. De woordstructuur van het Nederlands.* Amsterdam: Amsterdam University Press.

Bree, C. van (1987). *Historische Grammatica van het Nederlands.* Dordrecht: Foris.

Burnage, G. (1990). *Celex: a Guide for Users.* Nijmegen: Centre for Lexical Information.

[*CELEX 1990*]. *CELEX Dutch database.* Release N31, Computer software. Nijmegen: Centre for Lexical Information.

Carney, E. (1994). *A Survey of English Spelling.* London-New York: Routledge.

Chomsky, N. & M. Halle (1968). *The Sound Pattern of English.* New York: Harper and Row.

Chomsky, N. (1970). Phonology and Reading. *Basic studies on Reading*, edited by H. Levin & J.D. Williams, 3–18. New York: Basic Books.

Clements, G.N. & S.J. Keyser (1983). *CV Phonology. A Generative Theory of the Syllable.* Cambridge, Mass.: MIT Press.

Cohen, A., C.L. Ebeling, F. Fokkema & A.G.F. van Holk (1978). *Fonologie van het Nederlands en het Fries.* Den Haag: Martinus Nijhoff.

Cohen, A. & A. Kraak (1972). *Spellen is spellen is spellen. Een verkenning van de spellingproblematiek.* Den Haag: Martinus Nijhoff.

Coninck, R.H.B. de (1970). *Groot uitspraakwoordenboek van de Nederlandse taal.* Antwerpen: Nederlandsche boekhandel.

Damsteegt, B.C. (1971). Enkel of dubbel medeklinkerteken? *Taal en Tongval* **23**, 80–87.

Derwing, B.L. (1992). Orthographic aspects of linguistic competence. *The*

Linguistics of Literacy, edited by P. Downing, S.D. Lima & M. Noonan, 193–
210. Amsterdam-Philadelphia: John Benjamins.

Dibbets, G.R.W. (1983). Lochenen: belachchelijk? Argumentatie in een
ha(ch)chelijk spellingprobleem. *Gramma* **7**, 67–79.

Dougherty, D. & A. Robbins (1997). *SED & AWK, 2nd Edition.* Sebastopol:
O'Reilley & Ass.

Donaldson, B.C. (1983). *Dutch. A Linguistic History of Holland and Belgium.*
Leiden: Martinus Nijhoff.

[*Duden 1991*]. Rechtschreibung der deutschen Sprache. 20., völlig neu bearb. und
erw. Aufl. Mannheim etc.: Dudenverlag 1991.

[*Duden 1996*]. Rechtschreibung der deutschen Sprache. 21., völlig neu bearb. und
erw. Aufl. Mannheim etc.: Dudenverlag 1996.

Eisenberg, P. (1989). Die Schreibsilbe in Deutschen. *Schriftsystem und
Orthographie*, edited by P. Eisenberg & H. Günther, 57–84. Tübingen: Max
Niemeyer Verlag.

Frost, R. (1992). Orthography and phonology: The psychological reality of
orthographic depth. *The Linguistics of Literacy*, edited by P. Downing, S.D. Lima
& M. Noonan, 255–274. Amsterdam-Philadelphia: John Benjamins.

Geerts, G. (1994). Besluitvorming in het jongste spellingdebat. *De Spellingcommis-
sie aan het woord*, edited by A. Neijt, I. Roggema & J. Zuidema, 59–72. Den
Haag: SDU Koninginnegracht.

Geerts, G., E. Assink, P. van den Craen, J. de Rooij, G. de Schutter, G. Verhoeven &
W. Duthoy (1988). *Rapport van de Werkgroep ad hoc Spelling* [*Voorzetten* **20**].
's-Gravenhage: Stichting Bibliographia Neerlandica.

Geerts, G., J. van den Broeck & A. Verdoodt (1977). Successes and Failures in
Dutch Spelling Reform. *Advances in the Creation and Revision of Writing
Systems*, edited by J.A. Fishman, 179–245. The Hague-Paris: Mouton.

Geerts, G. & Heestermans, H. (1995). *Van Dale groot woordenboek der
Nederlandse taal. Twaalfde druk in de nieuwe spelling.* Utrecht-Antwerpen: Van
Dale lexicografie.

Gelb, I.J. (1963). *A Study of writing: the Foundations of Grammatology.* Chicago:
University of Chicago Press.

Giegerich, H.J (1992). *English Phonology: an Introduction.* Cambridge: Cambridge
University Press.

Gysseling, M. (1960). *Toponymisch woordenboek van België, Nederland,
Luxemburg, Noord Frankrijk en West Duitsland (voor 1226).* Brussel: Belgisch
Interuniversitair Centrum voor Neerlandistiek.

Grimes, J.E. & R.G. Gordon Design of New Orthographies. *Orthography, reading*

and dyslexia, edited by J.F. Kavanagh & R.L. Venezky, 57–76. Baltimore: University Park Press.

Günther, H. (1987). Phonological Recoding in the Reading Process. *Orthography and Phonology*, edited by P.A. Luelsdorff, 151–169. Amsterdam-Philadelphia: John Benjamins.

Günther, H. (1988). *Schriftliche Sprache. Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen.* Tübingen: Max Niemeyer Verlag.

Gussenhoven, C. (1980). 'Review' of Trommelen and Zonneveld (1979). *Gramma* **4**, 174–183.

Gussenhoven, C. (1992). Illustrations of the IPA: Dutch. *Journal of the International Phonetic Association* **22**, 45–47.

Gussenhoven, C. & R.H. Bremmers (1983). Voiced Fricatives in Dutch. Sources and Present-day Usage. *North-Western European Language Evolution* **2**, 1–22.

Haas, W. (1970). *Phono-graphic Translation.* Manchester: Manchester University Press.

Haas, W. de & M. Trommelen (1993). *Morfologisch handboek van het Nederlands: een overzicht van de woordvorming.* Den Haag: SDU.

Haeringen, C.B. van (1962). 'Spelling pronunciations' in het Nederlands. *Neerlandica, verspreide opstellen.* 37–60. Den Haag: Daamen.

Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M.C. van den Toorn (1997). *Algemene Nederlandse spraakkunst.* Groningen-Deurne: Martinus Nijhoff/Wolters Plantyn.

Heemskerk, J.S. (1989). Morphological Parsing and Lexical Morphology. *Linguistics in the Netherlands 1989*, edited by H. Bennis & A. Kemenade, 61–70. Dordrecht: Foris.

Heemskerk, J.S. & V.J. van Heuven (1993). MORPA: a Morpheme Lexicon-Based Morphological Parser. *Analysis and Synthesis of Speech, Strategic Research towards High-Quality Text-to-speech Generation*, edited by V.J. van Heuven & L.C.W. Pols, 67–85. Berlin: Mouton-De Gruyter.

Heynderickx, P. & J. van Marle (1994). Over het hybride karakter van -isch: Op de grens van inheems en uitheems. *Spektator* **23**, 229-239.

Heuven, V.J. van (1978). *Spelling en lezen. Hoe tragisch zijn de werkwoordsvormen?* Assen: Van Gorcum.

Heuven, V.J. van (1980). Aspects of Dutch Orthography and Reading. *Orthography, Reading and Dyslexia,* edited by J.F. Kavanagh & R.L. Venezky, 57–76. Baltimore: University Park Press.

Heuven, V.J. van (1994). Bastaards en andere vreemde woorden. *De Spellingcommissie aan het woord*, edited by A. Neijt, I. Roggema & J. Zuidema,

47–58. Den Haag: SDU Koninginnegracht.

Heuven, V.J. van & A. Hoos (1991). Hiatus Deletion, Phonological Rule or Phonetic Coarticulation? *Linguistics in the Netherlands 1991*, edited by F. Drijkoningen & A. van Kemenade, 61–70. Amsterdam: John Benjamins.

Heuven, V.J. van, A.H. Neijt & M. Hijzelendoorn (1994). Automatische indeling van Nederlandse woorden op basis van etymologische filters. *Spektator* **23**, 279-291.

Heuven, V.J. van & A.C.M. Rietveld (1997). *Algemene fonetiek*. Bussum: Coutinho.

Horbach-Kleijnen, R. (1988). *Hardnekkige spellingfouten. een taalkundige analyse*. Amsterdam: Swets & Zeitlinger.

Horck, A. van, H. Kempff, T. van Opstal & J. Renkema (1986). *Inventarisatie spellingvarianten. Een vooronderzoek naar de varianten in de Woordenlijst der Nederlandse Taal*, Internal report KUB, Tilburg.

Hulst, H.G. van der (1984). *Syllable Structure and Stress in Dutch.* Dordrecht: Foris.

Kager, R. (1989). *Stress and Destressing in English and Dutch.* Dordrecht: Foris.

Kager, R. & Zonneveld, W. (1986). Schwa, Syllables and Extrametricality in Dutch. *Linguistic Review* **5**, 197-221.

Kempen, G. (1994). De mythe van het woordbeeld. Spellingherziening taalpsychologisch doorgelicht. *Spektator* **23**, 292–301.

Kempff, H., T. van Opstal & J. Renkema (1988). *Inventarisatie spellingvarianten (vervolg). Een vooronderzoek naar de varianten in de Woordenlijst der Nederlandse Taal*, Internal report KUB, Tilburg.

Kerstens, J. (1981). Abstracte Spelling. *De Nieuwe Taalgids* **74**, 29–44.

Kiparsky, P. (1981). *Lexical Morphology and Phonology.* Cambridge, Mass.: MIT Press.

Klima, E.S. (1972). How Alphabets may reflect Language. *Language by Ear and by Eye: the Relationship between Speech and Reading*, edited by J.F. Kavanagh & I.G. Mattingly, 57–80. Cambridge, Mass.: MIT Press.

Kollewijn, R.A. (1916). *Opstellen over spelling en verbuiging.* Groningen*:* Wolters.

Kooij, J.G. (1983). Morphological alternations and phonological rules: the case of Dutch intervocalic fricatives. *Linguistics in the Netherlands 1983*, edited by H. Bennis & W.U.S. van Lessen Kloeke, 133–138. Dordrecht: Foris.

Kraak, A. (1996). De alfabetische bril. *Taalvariaties: toonzettingen en modulaties op een thema*, edited by R. van Hout & J. Kruijsen, 173–186. Dordrecht: Foris.

Kruisinga, E. (1951). *Het Nederlands van nu.* Amsterdam-Antwerpen: Wereldbibliotheek.

Luijn-Hindriks, T. van (1992). *Spelling, spraak en fonemische analyse: klankfouten van brugklasleerlingen.* Doctoral dissertation, University of Utrecht.

Mohanan, K.P. (1986). *The Theory of Lexical Phonology.* Dordrecht: Reidel.

Morais, J., P. Bertelson, L. Cary & J. Allegria (1986). Literacy Training and Speech Segmentation. *Cognition* **24**, 45–64.

Neijt, A. (1991). *Universele fonologie, een inleiding in de klankleer.* Dordrecht: ICG.

Neijt, A. (1994). Van orthografie naar ortografie. *De Spellingcommissie aan het woord*, edited by A. Neijt, I. Roggema & J. Zuidema, 59–72. SDU Koninginnegracht, Den Haag.

Neijt, A. & A.M. Nunn (1997). The recent history of Dutch Orthography. Problems solvedreated. *Leuven contributions in linguistics and philology* **86**, 1–26.

Neijt, A. & R. Riemersma (1995). *De nieuwe Spellinggids van de Nederlandse Taal.* Utrecht: Het Spectrum/Van Dale lexicografie.

Neijt, A. & J. Zuidema (1994a). Als kiviet naar de woordenlijst. *Spektator* **23**, 260–268.

Neijt, A. & J. Zuidema (1994b). *Spellingdossier.* Vol. 1, *Spellingrapport*. [*Voorzetten* **44**]. 's-Gravenhage: Stichting Bibliographia Neerlandica,.

Nijen Twilhaar, J. (1994). Genus en fonologisch gedrag van nomina op schwa. *Dialectfonologie*, edited by G.E. Booij & J. van Marle, 83–100. Amsterdam: P.J. Meertens-Instituut.

Nooteboom, S.G. (1972). *Production and Perception of Vowel Duration: a Study of Durational Properties of Vowels in Dutch.* Doctoral dissertation, University of Utrecht.

Nunn, A.M. (1992). Een modulair model voor spelling en fonologie. *Forum der letteren* **4**, 258–274.

Nunn, A.M. (1995). The spelling of long vowels in Dutch: one system or two? *Proceedings of the CLS opening Academic Year '95/'96*, edited by M. Biemans and M. Woutersen, 57–73. Tilburg: Centre for Language studies.

Nunn, A.M. & V.J. van Heuven (1993). MORPHON: Lexicon-Based Text-to-Phoneme Conversion Rules. *Analysis and Synthesis of Speech, Strategic Research Towards High-Quality Text-to-Speech Generation*, edited by V.J. van Heuven & L.C.W. Pols, 97–109. Berlin: Mouton-De Gruyter.

Nunn, A.M. & A. Neijt (1996). Dutch orthography: a near-optimal phonological transcription? *Linguistics in the Netherlands 1996*, edited by C. Cremers & M. den Dikken, 195–206. Amsterdam: John Benjamins.

Nunn A.M. & A. Neijt (1997). Rules and exceptions in the spelling of loan words in Dutch. *Linguistics in the Netherlands 1997*, edited by J. Coerts & H. de Hoop, 147–157. Amsterdam: John Benjamins.

Ohala, J.J. (1992). The costs and benefits of phonological analysis. *The linguistics of*

literacy, edited by P. Downing, S.D. Lima & M. Noonan, 211–237. Amsterdam-Philadelphia: John Benjamins.

Oostendorp, M. van (1995). *Vowel quality and syllable projection.* Doctoral dissertation, University of Tilburg.

Prince, A. & P. Smolensky (1993). *Optimality: Constraint Interaction in Generative Grammar.* Technical Report#2 of the Rutgers Center for Cognitive Science, Rutgers University.

[*Rapport 1967*]. *Rapport van de Nederlands-Belgische commissie voor de spelling van de bastaardwoorden.* 's-Gravenhage: Staatsuitgeverij, 1967.

[*Rapport 1969*]. *Eindvoorstellen van de Nederlands-Belgische commissie voor de spelling van de bastaardwoorden.* 's-Gravenhage: Staatsuitgeverij, 1969.

Reker, S. & N. Streekstra (1988). Abbreviërende morfologie. *Tabu* **18**, 195–211.

Reitsma, P. (1983). *Phonemic and graphemic codes in learning to read.* Doctoral dissertation, University of Amsterdam.

Rijnsoever, R.J. van (1988). *Spelling en leren lezen: schijnbewegingen van orthografische effecten in de ontwikkeling van de leesvaardigheid.* Doctoral dissertation, University of Nijmegen.

Saciuk, B. (1970). *Lexical strata in generative phonology (with examples from Ibero-Romanic).* Ann Arbor: University Microfilms.

Schiller, N.O., A.S. Meyer & W.J.M. Levelt (1997). The syllabic structure of Spoken Words: Evidence from the Syllabification of Intervocalic Consonants. *Language and Speech* **14**, 103–140.

Schutter, G. & R. Collier (1986). Intuïtieve syllabifiëring in het Nederlands. *De Nieuwe Taalgids* **79**, 441–452.

Schane, S.A. (1977). Rule Breaking in English Spelling, a study of final e. *Studies in Descriptive and Historical Linguistics, Festschrift for Winfred P. Lehman*, edited by P.J. Hopper, 217–234. Amsterdam-Philadelphia: John Benjamins.

Siegenbeek, M. (1805a). *Verhandeling over de spelling der Nederduitsche taal, ter bevordering van eenparigheid in dezelve.* Uitgegeven in naam en op last van het Staats-bewind der Bataafsche republiek, Amsterdam: J. Allart.

Siegenbeek, M. (1805b). *Woordenboek voor de Nederduitsche spelling.* Amsterdam: J. Allart.

Sijs, N. van der (1996). *Leenwoordenboek: de invloed van andere talen op het Nederlands.* Den Haag-Antwerpen: SDU/Standaard.

Taeldeman, J. (1994). De beeldvorming van het Nederlands. *De Spellingcommissie aan het woord*, edited by A. Neijt, I. Roggema & J. Zuidema, 33–46. Den Haag: SDU Koninginnegracht.

Trommelen, M. (1983a). *The syllable in Dutch with special reference to diminutive*

formation. Dordrecht: Foris.

Trommelen, M. (1983b). Phonological Properties of the Dutch Velar Nasal. *Sound structures: studies for Antonie Cohen*, edited by M. van den Broecke, V. van Heuven et al., 259–269. Dordrecht: Foris.

Trommelen, M. & Zonneveld, W. (1982). *Inleiding in de generatieve fonologie.* Muiderberg: Coutinho.

Trommelen, M. & Zonneveld, W. (1989). *Klemtoon en Metrische Fonologie.* Muiderberg*:* Coutinho.

Velde, I. van der (1956). *De tragedie van de werkwoordsvormen.* Doctoral dissertation, University of Groningen.

Venezky, R.L. (1970). *The structure of English Orthography.* The Hague-Paris: Mouton.

Verhoeven, G. (1985). *De strategieën van de speller.* Groningen: Wolters-Noordhoff.

Vieregge, W.H. (1985). *Transcriptie van spraak.* Dordrecht: Foris.

Wester, J. (1984). SF: contouren van een toegepaste fonologie. *Nieuwe Taalgids* **11**, 30–43.

Wester, J. (1985a). Language Technology as Linguistics: A Phonological Case Study of Dutch Spelling. *Linguistics in the Netherlands 1985*, edited by H. Bennis & A. van Kemenade, 205–212. Dordrecht: Foris.

Wester, J. (1985b). Autonome Spelling en Toegepaste Fonologie, of: naar een generatieve spellingtheorie. *Gramma* **9**, 173–196.

Wester, J. (1987). Proeve van een generatieve spelling: de representatie van fricatieven. *Gramma* **11**, 59–82.

Wester, J. (1989). De Tao van taal, of: Chanel no. 5 en het Geval Nederlands. J. Wester, G. Krol & C. Crombach, *Gaat het Nederlands teloor? Drie essays.* Vianen: De Haan.

Wijk, N. van (1939). *Phonologie, een hoofdstuk uit de structurele taalwetenschap.* 's-Gravenhage: Nijhoff.

Winkel, L.A. te (1860). *De Nederlandsche Spelling onder beknopte regels gebragt.* Tweede, verbeterde en vermeerde uitgave. Leiden: D. Noothoven van Goor.

Winkel, L.A. te (1863). *Ontwerp der Spelling.* Leiden: D. Noothoven van Goor.

Winkel, L.A. te (1865). *De Grondbeginselen der Nederlandsche Spelling, Regeling der spelling voor het Woordenboek der Nederlandsche Taal.* Leiden: D. Noothoven van Goor.

Winkel, L.A. te (1884). *De grondbeginselen der Nederlandsche spelling, regeling der spelling voor het woordenboek der Nederlandsche taal*. Den Haag: Van Cleef.

[*Woordenlijst 1866*]. Vries, M. de & L.A. te Winkel *Woordenlijst voor de Spelling der Nederlandse Taal, met aanwijzing van de geslachten der naamwoorden en de vervoeging der werkwoorden,* 's-Gravenhage-Leiden-Arnhem: Thieme, 1866.

[*Woordenlijst 1914*]. Vries, M. de & L.A. te Winkel *Woordenlijst voor de Spelling der Nederlandse Taal, met aanwijzing van de geslachten der naamwoorden en de vervoeging der werkwoorden.* Leiden-'s-Gravenhage: Nijhoff/Sijthoff, 1914.

[*Woordenlijst 1954*]. *Woordenlijst van de Nederlandse Taal, samengesteld in opdracht van de Nederlandse en de Belgische regering.* Den Haag: SDU uitgeverij, 1954.

[*Woordenlijst 1995*]. *Woordenlijst Nederlandse taal. Samengesteld door het Instituut voor Nederlandse lexicologie in opdracht van de Nederlandse Taalunie.* Den Haag-Antwerpen: SDU Uitgevers/Standaard Uitgeverij, 1995.

Zonneveld, W. (1978). *A formal theory of exceptions in generative phonology.* Dordrecht: Foris.

Zonneveld, W. (1980). Autonome Spelling, *De Nieuwe Taalgids* **73**, 518–537.

Zonneveld, W. (1982). The descriptive power of the Dutch theme-vowel. *Spektator* **11**, 342–365.

Zonneveld, W. (1983). Lexical and Phonological Properties of Dutch Voicing Assimilation. *Sound structures: studies for Antonie Cohen.* edited by M. van den Broecke, V. van Heuven et al., 59–110. Dordrecht: Foris.

Zonneveld, W. (1993). Schwa, Superheavies, Stress and Syllables in Dutch. *Linguistic Review* **10**, 59–110.

Zonneveld, R. van (1983). *Affix-grammatika. Een onderzoek naar woordvorming in het Nederlands.* Doctoral dissertation, University of Groningen.

Zuidema, J. (1988). *Efficient spellingonderwijs: een leer- en expertmodel voor het spellen.* Doctoral dissertation, University of Utrecht.

Zuidema, J. (1994). Alles behalve de letters en dan toch de puntjes op de i. *De Spellingcommissie aan het woord*, edited by A. Neijt, I. Roggema & J. Zuidema, 19–32. Den Haag: SDU Koninginnegracht.

Zuidema, J., A. Neijt & J. Weber (1994). Hiërarchieën op de knieën. *Spektator* **23**, 137–163.

Zuidema, J., J. Weber, A. Nunn, A. Neijt & V. van Heuven (1994). *Groen zonder bladeren. Specificaties voor een elektronische versie van de nieuwe Woordenlijst in opdracht van de Nederlandse Taalunie*. Unpublished report, Nederlandse Taalunie, Den Haag.

Zwaardemaker, H.C. & L.P.H. Eykman (1928). *Leerboek der phonetiek: inzonderheid met betrekking tot het standaard-Nederlandsch.* Haarlem: Bohn.

Appendix A
The lexicon and implementation of rules

In order to obtain an accurate formal description of the Dutch spelling system, the spelling rules were formalized and applied to the sound representation of words. The derived spelling and the actual spelling were compared., and the spelling rules were then modified where necessary. This appendix describes the lexicon used for this purpose, its source, the selection of words, the modification of the information provided for each entry, and the implementation of the spelling rules.

A.1 The lexicon

The lexicon used here is composed of data from the CELEX (*Center for lexical information*) database, see *CELEX* (1990).

Selection of entries.
Since the investigation was restricted to the word domain, I selected all lemmas except for compounds, i.e. monomorphemic words, derivations, lexicalized complex words and words of which the morphological structure was marked as undefined (some of which were in fact compounds, e.g. *hou#vast*, *citrus#pers*). In order to keep the lexicon as small and manageable as possible, not all inflected forms were included in the lexicon. However, inflected forms are necessary to be able to test the accuracy of spelling rules that apply exclusively to inflected words, e.g. morpholexical spelling rules or apostrophe placement. Therefore, examples of all types of inflection (plural and diminutive form of nouns, and inflected forms of verbs, adjectives and adverbs) with all types of spelling (words with a regular spelling, words ending in silent letters etc.) were selected.

- Words marked as names were initially selected as well, but excluded later on, since they often had no phonetic representation or an inaccurate one. Another problem was that they proved idiosyncratic not only with respect to spelling (names often have a historical spelling, e.g. *Kraemer* [krɑmər], *Coole*-[kolə], see [*Woordenlijst 1954*], p. XXXIX) but also with respect to morphological

behaviour. Native names are often combined with non-native affixes, e.g. *Schultink+iaans*. This made it difficult to classify them as native or non-native.

- Words with the same pronunciation and spelling were considered one lemma even when they had a different meaning or category. For instance, the noun *behouden* ('to keep') and the adjective *behouden* ('safe') form one entry, and the same holds for the nouns *inning* (id.) and *inning* ('collection'). On the other hand, spelling variants (*stampij-stampei*) and homographs (*bedelen* [bədɛlə] ~ [bedələ], 'to dole out' ~ 'to beg') were considered separate lemmas. Since the lexicon was created before the most recent spelling reform, non-preferred spelling variants of words were initially included (except for variants ending in -*izeren* and -*izatie* that CELEX gives for all words in -*iseren* and -*isatie* were omitted). This way a lexicon of 49189 words was created.
- For every selected entry, the spelling and the phonetic transcription was taken, together with other types of information, i.e. spelling with syllable boundaries, category and the CELEX identification number.

Modification of information provided by CELEX.

The information provided by the CELEX database was modified in several ways.

Adaptation of the spelling

The words in the CELEX database were given in the spelling from before the spelling reform of 1995, so their spelling had to be adapted. Spelling changes that involve the linking morpheme in compounds were hardly necessary since the test lexicon contained few compounds. More important was the abolishment of most non-preferred spelling variants, which reduced the number of entries to 45821 words. In a few cases the preferred and non-preferred variant were interchanged (*produkt* (also *product*) → *product*) or the spelling of a word was changed (*lambrizering* → *lambrisering*). For an overview of the reforms, see Appendix H. Not all remaining words have an official spelling, since some words do not occur in [*Woordenlijst 1995*]. In that case the spelling of Geerts & Heestermans (1995) was chosen. There were 652 words that occurred in neither dictionary, for instance *verspoelen*, *gelieg* and *Mokummer*. Such words are very infrequent, totally predictable, or incorrect; the official spelling of the last word should be *Mokumer*.

Phonological representations

CELEX does not provide a phonological transcription for all words. The transcriptions that are available do not always abstract from all predictable sound processes (e.g. *sleeën* and *uier* were represented as /slejən/, /œyjər/ with predictable

inserted glides rather than as /sleən/, /œyər/). In other cases the phonological transcriptions represented the effect of unpredictable processes. For instance, CELEX transcribes *opzichter* as /ɔpziər/ instead of /ɔpzɪxtər/ and *sterfelijk* as /stɛrvələk/ instead of /stɛrfələk/ with the lexicalized effect of an unpredictable devoicing process. I therefore decided to not to use these representations, but to compute them by a method developed in a previous project, see Nunn & Weber (1993), Zuidema et al. (1994). This method involved taking the phonetic representation as a starting point and then reversing the effect of all predictable phonological rules. Most sound alternations are not accompanied by spelling alternations, so the effect of these rules could be undone by comparing the spelling of a word with its pronunciation ([hɔnt] → /hɔnd/ because of *hond*). Some sound alternations coincide with spelling alternations, i.e. morpholexical rules ([bompjə]-*boompje*), final devoicing of *v* and *z* ([kis]-*kies*, [braf]-*braaf*) and in some cases degemination ([ɣəbrɑnt]-*gebrand*). The effect of morpholexical rules was not undone, since it is represented in orthography (see chapter 3), but the effect of degemination and devoicing had to be reversed by hand. The morpheme-internal or lexicalized effect of obligatory sound rules was not undone either, e.g. /ambɛld/, /ɑtfokat/ instead of /anbɛld/, /ɑdvokat/. The phonetic representations that form the basis of the abstract sound representations in the test lexicon were not always accurate and had to be modified. In some cases, errors manifest themselves since they lead to an incorrect spelling. For instance, the words *judaïsme*, *visite* and *havikachtig* ([jydaɪsmə], [vizitə] and [havəkɑxtəx]) were transcribed as *[visitə], *[jydawɪsmə] and *[havikɑxtəx] so the spelling rules derived *judawisme*, *vicite* and *haviekachtig*. Evident errors of this kind were detected and corrected. Incorrect pronunciations that do not lead to an incorrect spelling, for instance and [plyxə] instead of [plysj] (*pluche*) or [zevəntəx] instead of [sevəntəx] (*zeventig*) cannot be traced this way, so it is possible that there still are some incorrect sound representations in the lexicon.

In order to automatically undo the effect of sound rules, it was crucial to be able to refer to sounds and their spelling, e.g. '*d* pronounced as [t]'. Therefore phonemes and graphemes had to be related. For this purpose, I used an existing program developed at the Phonetic Laboratory of the Leiden University. This program takes the (syllabified) spelling and pronunciation of words (syllabified and provided with markers for primary stress) and makes a combined representation. As an example, the word *bank* (/bɑŋk/) is the following representation:

    ;';b;a;n,ng;k;

This representation should be read as follows: ;a; means <a> pronounced as /ɑ/, ;n,ng; means <n> pronounced as /ŋ/, etc. ';' marks the division between grapheme

phoneme pairs and ',' the division between graphemes and phonemes. In order to distinguish vowel letters corresponding to zero, one or two V-positions I used the following notational conventions: ;@; (e.g. *de* /də/), ;e; (e.g. *hemd* /hɛmd/) ;ee; (*chinees* /sjinez/) and ;e,ee; (e.g. *mechanisch-*/mexanis/, *velair* /velɛːr/). The spelling in the combined representation was changed into an abstract spelling representation by reversing the effect of autonomous spelling rules. Consequently, the combined representation can be converted to an abstract spelling representation (by deletion of the symbols representing sounds) or to an abstract pronunciation (by deletion of the symbols representing spelling). The first representation can be used to test the effect of all spelling rules, the second to test the effect of autonomous spelling rules only.

The computer-readable code used in CELEX to represent the pronunciation, was slightly altered, so that phonemes are represented by their default spelling where possible:

(1) **Sounds**   **Notation**
    p b t d k f v s z x ɣ h m n ŋ l r w j   p b t d k f v s z ch g h m n ng l r w j
    ɪ ɛ ɔ ʏ ɑ i ʏ u e ∅ o a ɛi œy ɑu ə   i e o u a ie uu oe ee eu oo aa ei ui au @

Morphological boundaries
Since spelling rules are sensitive to the morphological structure of words, the abstract sound representations were provided with markers for boundaries of affixes: '%' for native suffixes, '^' for non-native suffixes, '#' for compound and prefix boundaries, '$' for the boundaries of the prefixes *ge-*, *be-*, *ver-*, *her-* and *ont-*, and '+' for the boundaries of non-native prefixes (*coördineren*). These markers were added automatically by affix-stripping; compound boundaries were inserted by hand. Incorrect markers were discovered when they resulted in incorrect spelling predictions.

Labels
Two types of labels were added. The first label marks words as 'native', 'non-native' or 'mixed' (combining native and non-native morphemes in derivations and frozen compounds), and was computed by formalizing the criteria for native words given in Appendix B and applying them to all words of the test lexicon. In the case of derivations, the criteria were applied to the stem only, in the case of compounds to both stems. The criteria with respect to phonemes, syllable structure and prosodic words were implemented the same way as spelling rules, see below. Subsequently, it was examined if the plurals of monosyllabic nouns marked native satisfy the criteria imposed on inflected forms. Fifty per cent of the words in the test lexicon was

native, 49% non-native and 1% mixed. The second label marked words as irregular and was assigned to words whose spelling cannot be correctly predicted by the definitive version of the spelling rules.

A.2 Spelling rules

As observed in the introduction, it is in principle possible to formulate rules that give a 100% correct prediction of spelling, but this implies formulating rules that are valid for one word only, or using information not available to speakers. A more insightful description must strike a balance between accuracy and generality of the rules. While formulating spelling rules, I followed the following principles:

- No rules or subrules are formulated that are valid for one morpheme only except for productive affixes such as *-heid*. When a rule would affect less than five words, the relevant examples are listed as exceptions.
- Rules only use information available to literate speakers of Dutch. For instance, the rules do not refer to historical or dialectal pronunciations, the etymological origin of words or opaque morphological structure.

If rules that satisfy these requirements did not derive the correct spelling for a given word, even after they were improved as much as possible, the spelling of this words was considered irregular.

The spelling rules were implemented by means of the non-interactive editor SED (part of the UNIX operating system, also available as a DOS-program), see Dougherty & Robbins (1997). This editor takes its commands from a command file and applies them to an input file to create an output file. The rules in the Appendices C-E were translated into the required format. This way, four command files were created: two files with phoneme-to-grapheme conversion rules for native and non-native words, respectively, a file with orthographic syllabification rules and a file with remaining autonomous spelling rules. The exact format of the command files is not relevant and outside the scope of this appendix; it suffices to know that the commands are comparable to context-sensitive rewriting rules and thus sufficiently expressive to implement the spelling rules listed in Appendices C-E. By applying the command files to the abstract sound representations and comparing their output to the spelling as listed in the lexicon, the accuracy of the rules could be evaluated.

Appendix B
Classification criteria for native and
non-native words

This appendix gives the criteria with which the lexicon was divided into native and non-native words. For reasons given in chapter 4, I will not further divide non-native words into hybrid and foreign words.

In chapter 4, we saw that Te Winkel (1863) and Neijt & Zuidema (1993) distinguish native and non-native words by the extent to which they are adapted to the Dutch linguistic system. Van Heuven et al. (1994) showed that it is possible to predict the etymological origin of most words from these formal linguistic properties. Since Van Heuven et al. (1994) have the most complete set of criteria, their criteria will be chosen as the starting point. Instead of merely repeating their restrictions, it will be argued that the restrictions on native words can be seen as tightening of the restrictions on Dutch words. For this purpose, the literature on phoneme inventories, syllable structure, words and selection of inflection of Dutch words will be reviewed here.

Following Van Heuven et al., I will define the native lexicon by describing what is possible in native words rather than listing properties that mark words as non-native. The latter approach is more efficient, but it runs the risk of incorrectly accepting a word as native: properties of loan words are inherently unrestricted. We can define the properties of the loan words that are adapted so far, but there is no guarantee that a new loan words will not exhibit another property. The classification pertains to morphemes rather than to words, since complex words can combine native and non-native elements, see also Te Winkel (1863:70), Saciuk (1970:26), and since properties of complex words are much less restricted. For instance, native words are characterized as having one full vowel only, but this clearly does not hold for complex words such as *kleuterschool* (*kleuter* + *school*) or **luisteraar** (*luister* + *-aar*). However, complex non-native words such as *collaboreer* will be treated as one morpheme.

Although one of the properties that marks words as non-native is their ability to combine with non-native derivational affixes, see for instance Booij (1977:131–9) and Neijt & Zuidema (1994), Van Heuven et al. (1994) do not classify words as non-native when they fulfil all criteria for native words but are related to a word with a non-native affix. I will adopt this approach for the following reasons.

Firstly, it is not always easy to establish which words are to be considered

related. Relating words often involves morpholexical alternations. Some of the alternations are accounted for by allomorphy rules in the literature, e.g. insertion or deletion of schwa (*troebel-troebleer*, *code-codeer*) alternation of schwa with a full vowel (*eth**e**r-eth**e**risch*). However, other words cannot be related by rules of any generality, e.g. *feest-festiviteit*, *klant-klandizie*, *klerk-clericaal*, *mater-maternalistisch*, *vox-vocaal*, *cirkel-circulair*, *kaart-cartotheek*, *mars-marcheer*. In other cases the semantic relations are not clear: *file-fileer*. In these cases, both words are diachronically related to a common base, but the monomorphemic word, has been much more adapted, as a result of which it is formally native: *cirkel* and *circulair* are both related to *circulus*, *regel* and *reguleer* to *regula*. The adapted form of the word cannot be combined with non-native affixes: **cirkelatie*, **klerkaal*, so it will not be considered non-native, and in some cases the adapted form exists alongside the original form with a different meaning, e.g. *straat* ('street') and *stratum* (id.).

   Secondly, even when we only consider pairs that are related by regular word formation and sound rules, the criterion is not unproblematic, since some native morphemes are exceptionally combined with non-native affixes such as *bibberatie*, *kolderiek*, *wasserette*, *kretologie*, *prullaria*, *deskundoloog*, *wettisch*, *hollanditis*, *aanstelleritis*. The affixes that most frequently combine with native morphemes are *-iteit*, *-eer* and *-age*: *flauwiteit*, *stommiteit*, *kleineer*, etc., *pluimage*, *dijkage*, *lekkage*, *kijvage*, *timmerage* etc. Another complication is the fact that *-ist*, *-ier* and *-ement* combine with native (*fluitist*, *tuinier*, *dreigement*) as well as non-native words (*botanist*, *marinier*, *changement*).

   Finally, not all words marked as non-native by this criterion actually have a non-native spelling. In many cases the spelling of the shorter word does not reveal that it is non-native, e.g. *faam-fameus*, *dupe-dupeer*, *sekte-sektarisch*, *filter-filtreer*, *taks-taxeer*. In some cases the shorter form is adapted while the longer one kept the etymological spelling: *akte-acteer*, *kader-encadreer*, *soepel-souplesse*, *lier-lyrisch* etc. An exhaustive list of such words that have a non-native spelling is the following:

(1)   base-baseer   charme-charmant   claus-clausaal
         code-coderen   column-columnist   crypt(e)-cryptisch
         fase-faseer   farce-farceer   frase-fraseer
         friet/frites-frituur   pose-poseer   tact-tactisch
         truc-trucage   type-typeer

The facts in (1) suggest that the presence of a related non-native word does not prevent spelling adaptation of the base words. It appears that as far as spelling is

concerned the mere presence of a related non-native word does not mark words as non-native. Words such as *code* will be considered exceptionally non-native words (we will see below that there are more formally native words with a non-native spelling, e.g. *laser* and *piste*; words such as *code* will be treated the same way).

In this appendix, I will discuss restrictions imposed on native phonemes, syllables, words, stress patterns, inflection, respectively and subsequently list words incorrectly classified as native or non-native.

B.1 Native phonemes

In their discussion of Dutch phonemes, authors such as Van Wijk (1939:39), Cohen et al. (1978:25, 34), Booij (1981:19, 1995:6–7) and Trommelen & Zonneveld (1982:17, 22) observe that certain phonemes only occur in loan words. All these authors consider the long vowels of *rose* ([rɔːzə]), *freule* ([frØːlə]) and *serre* ([sɛːrə]) foreign phonemes. The other foreign phonemes are mentioned in some of the studies only: for instance, the nasalized vowels of *fin*, *bon* and *sans* in Cohen et al. (1978), Booij (1981, 1995). The fact that some foreign phonemes are omitted from phoneme inventories (which could imply that they are considered too marginal even to be listed as a foreign phoneme) underlines that it is better to list the native phonemes and consider all remaining phonemes foreign phonemes. Native phonemes were already listed in chapter 2, and repeated here:

(2)   p b t d k f v s z x ɣ h m n ŋ l r w j ɪ ɛ ɔ ɣ ɑ i ʏ u e Ø o a ɛi œy ɑu ə

All morphemes containing another sound are considered to be non-native.

B.2 Native words

The first author to formulate restrictions on the form of native (underived) words (after Te Winkel) is Zonneveld (1993): "polysyllabic words, or more precisely, words with more than one full vowel, count as 'foreign'." Following Te Winkel, I will consider this observation to be accounted for by the fact that native Dutch words are monosyllabic, unless they contain semi-affixes, of which the vowel mostly is a schwa.

Support for the affix-like status of syllables with schwa is the well-known observation by Kager & Zonneveld (1986) that clusters before a syllable with a

schwa behave as if they were word final. For instance, unlike clusters preceding a full vowel, clusters preceding the schwa always have decreasing sonority: *halter*, **hatler* (cf. *atleet*) and contain word-final appendices: *feest-heester* (cf. **feestiviteit*). Kager & Zonneveld accounted for this distribution by assuming that the sequences *-er* and *-el* (the schwappendix) can be freely added at word edges. They observed that there are some exceptions such as *ritme*, *ordner* and *franje*, but fail to relate this to the fact that these words are loan words. Zonneveld (1993:80) first recognized that this generalization only holds for native words.

A further indication that syllables with a schwa originally were suffixes is the fact that they predict the behaviour of the whole word. For instance, Nijen Twilhaar (1994) showed that words ending in schwa select the article *de* (*de aarde*, *de keuze* etc.). Unlike words with a full vowel in the final syllable (e.g. *huis-huizen*, *wee-weeën*), words ending in a schwallable (also) select the plural *-s* (*keuze-keuzes/keuzen*, *kelder-kelders*, *keuken-keukens*, *wortel-wortels*). In some cases the affix status of such sequences is not so clear. However, prefix-like sequences such as *ge-*, *be-*, *ver-* and *-ont* and suffix-like sequences such as *-e*, *-er(d)*, *-en(d)*, *-el(d)*, *-em*, *-ik*, *-erik*, *-ens*, *-ent*, *-es*, *-uw(e)*, *-ig*, *-elig*, *-ing*, *-aar(d)* or *-(e)lijk* will be allowed to occur in native words. Van Heuven et al. also mention *je-*, *me-*, *de-*, *-eren*, *-ige*, *-ers*, *-and*, *-og*, and *-ond*. However, *je-* only appears in *jenever*, *jenoffel* (of French origin) and *jelui* (a compound) and *te-*, *me-* and *de-* in lexicalized phrases. *-eren*, *-ige* and *-ers* are only present in inflected morphemes. The syllables *-ond*, *-og*, and *-and* are listed exclusively for *avond*, *oorlog*, *hertog* and *vijand*. These words are considered (native) pseudo-compounds here.

B.3 Native syllables

Syllables are defined both by universal principles, such as the Sonority Sequencing Generalization (henceforth Sonority) and language-specific restrictions on the number of positions in rhyme, nucleus and onset and the possible combinations of segments (see Trommelen 1983b, Booij 1995:24–25). Dutch rhymes consist of a nucleus of one position followed by a coda of one or two consonants. The rhyme must have at least two, maximally three positions. At word edges an appendix of one or two dental consonants: (/d/, /t/, /z/, /s/, /st/ or /ts/) can be added (in some words the appendix is not filled by a dental obstruent but by a dental sonorant: *hoorn*, *voorn*, *deern* etc., cf. Kager (1989:206), or by another consonant: *twaalf* and the inflected verb forms *verwierf*, *zwierf*, *stierf*, *hielp*, *wierp*).

Differences between syllables in native and non-native words are expected to

follow from the language specific co-occurrence constraints. However, such differences are not systematically listed in the literature, since the native/non-native distinction has received even less systematic attention with respect to syllable structure than with respect to phoneme inventories. Sometimes sequences that only occur in loan words are excluded, e.g. in Van Wijk (1939). In other cases 'foreign' clusters are allowed, e.g. /sk/ in Cohen et al. (1978), Booij (1981:211), Trommelen & Zonneveld (1982:14) and Trommelen (1983:112).

Zonneveld (1993:82) is the first to formulate separate co-occurrence restrictions for native and non-native words. For our purpose, we need only consider the syllable template for native words (in fact, in view of the unrestricted properties of loan words, we can only define syllable templates for native words and possibly hybrid words; not for non-native words). Zonneveld observes that sequences such as /skr/ and /skl/ only occur in loan words (e.g. *scribent*, *sclerose*), but he does not formulate the complete set of segmental conditions on native syllable templates, so this remains to be done.

The easiest way to define native syllables is to take the Dutch syllable as a starting point and formulate stricter restrictions that can only be satisfied by the syllables of native words. As mentioned above, most linguists already exclude some properties of loan words, but since they have no clear definition of native words, other loan word properties are still allowed. The basis of the native constraints used in this study are those of Booij (1995:24–29, 33–42). At first sight Booij's syllable constraints do not form a good starting point since they are defined for (possibly complex) words, while we need to define morphemes. However, the conditions imposed on native words imply that all syllables behave as if they were morpheme-final, and the difference between possible final syllables in words and morphemes is only formed by the fact that morphemes cannot have appendices with more than two dental consonants; there are no monomorphemic counterparts to words such as [bərum**tst**] (*be+roemd+st*). When we only allow appendices with two dental consonants since they occur in monomorphemic native words such as [ɣirst] (*gierst*), we derive the conditions for morphemes. Thus, the following revised Appendix Constraint holds for morphemes:

(3)   **Appendix Constraint**
          an appendix consists of one or two dental obstruents

I will now propose additional co-occurrence constraints on onsets, codas and rhymes that can only be satisfied by native morphemes.

Rhyme constraints

- The diphthong /ɑu/ is only followed by a dental obstruent in the rhymes of native words: *kous*, *stout*, etc., so Booij's contraint that diphthongs can not be followed by /r/ must be tightened for /ɑu/ (Van Heuven et al. claimed that the only cluster that can follow /ɛi/ is /st/, but words such as *grijns* and *eind* show that this is too restrictive).

Onset constraints

- The sounds /x/, /sk/, /sf/, /sw/ and /fj/ do not appear in the onsets of native words, but only in loan words such as *cholerisch*, *sferisch*, *skelet*, *swastika* and *fjord*, see Te Winkel (1860:20). However, /s/ must be allowed in combination with /x/ to allow for words such as *schreeuw* and *schat*.
- Booij excludes /fn/ and /ɣn/, but these clusters will be allowed here as they appear in native words like *fnuiken* and *gnuiven*.

Coda constraints

- Booij allows /s/ to violate Sonority in coda position because of words like *wesp*, *astma* and *grotesk*. However, /sk/ does not occur in native words, so only /sp/ and /st/ will be allowed. The clusters /rx/ and /rf/ do not occur in native words either (only in non-native words such as *morfine* and *monarch*), and /lf/ only in *elf*, so we can conclude that liquids are only followed by voiced fricatives (or an appendix). *Elf* will be considered an exception, as well as some words in which liquids followed by non-obstruents or voiceless fricatives: *murw, ernst, wormt, garnt, olms*; *burcht, elft, helft, knurft, schurft* and *herfst*.
- There are no native words in which /b/ follows a long vowel (*Swaab* and *kuub* are loan words), so we can extend Booij's ban on /b/ in the second coda position to the first coda position. Van Heuven et al. claim that /lx/ and /rx/ are allowed in native words, but I have found no examples of these clusters. On the other hand, they excluded /f/ and /ft/ after long vowels and /kt/ and /pt/, but these sequences occur in words that do not show non-native properties: *poef, ooft, dokter, naakt, sekt, stipt, nipt, prompt, inkt, markt*, and *abt*.

Survey (4) lists the additional requirements to distinguish native words:

(4) **Additional co-coccurrence constraints for native words**
   Rhyme constraints
   The diphthong /ɑu/ can only be followed by dental consonants.
   Onset constraints
   The sound /x/ is only allowed after /s/.
   After /s/ only (voiceless) labiodental or velar fricatives, and (voiceless) bilabial
      or dental plosives are allowed. (/sp/ can be followed by /l/ and /r/, the other
      clusters only by /r/).
   Obstruent-nasal clusters are excluded except for /sm/, /sn/, /kn/, /ɣn/ and /fn/.
   The sound /j/ is preceded by voiceless alveolar or velar fricatives; /w/ is preceded
      by alveolar consonants except for /s/, or by a velar plosive.
      Coda constraints
   Liquids are followed by voiced fricatives only (or an appendix).
   /b/ only occurs in a non-branching coda after a non-branching nucleus.

Thus the co-occurrence constraints for Dutch words are changed in those for native
morphemes with a small number of revisions and additions. Native morphemes must
satisfy the conditions imposed on syllable structure of possible Dutch words, as well
as the additional requirements under (3) and (4).

B.4 Native stress patterns

Stress of native Dutch words should be on the full vowel of the stem. In most words
this requirement is vacuous as most semi-suffixes and semi-prefixes contain a schwa
and are therefore stressless (this also holds for vowels in *-uw*, *-ing* and *-ig*, see
Appendix F). This condition thus only marks words as non-native when they have
stress on the sequences that could otherwise be analysed as semi-affixes such as *-ing*
or *-aar* in *sering* and *velaar*.

B.5 Native inflection

Native words can also be characterized by their choice of inflectional suffixes.
Native plural formation always selects the suffixes *-s* or *-en* (or *-deren* (*kinderen*),
-eren (*eieren*), *-ien* (*koeien*) or *-nen* (*redenen*)), see De Haas & Trommelen
(1993:157–163). The affix *-s* can only be chosen in native words when the last
syllable contains a schwa or a long vowel (not /i/ or /e/): *pa's*, *do's*, *roes*, see Van

Heuven et al. (the words *zoon*, *kok*, *knecht*, *maat*, *broer*, *oom*, *voorn*, *deern*, *doorn* and *hoorn* have an exceptional plural ending in -*s*). All other affixes (that sometimes replace part of the word to which they attach, e.g. *museum-musea*), are non-native. This criterion marks words with a plural in -*s* such as the following as non-native: *bat*, *bowl*, *mail*, *poule*, *rock*, *quote*, *shake*, *vue*.

B.6 Exceptions

With the conditions given above, we can divide the lexicon into native and non-native words, but there are some exceptions.

Non-native words marked as native
The following 146 words betray their foreign origin by their spelling only:

(5) axel, bacon, barrel, base, bi, blazer, bloc (en bloc), blouse, box, bühne, buste, button, camber, camel, camper, captain, ceder, ceel, cent, center, ces, chintz, chose, cider, cijfer, cijns, cirkel, cis, citer, clipper, close, cluster, coaster, con, counsel, counter, cracker, cri, cum, curve, cutter, cypers, cyste, daim, double, dyne, eau, ether, face, fade, fake, faun, fax, fine, flou, flox, flux, force, fox, fyle, hetze, hymen, krill, label, laser, lire, lux, lynx, mauser, mauve, mille, mine, mise, mix, naft, nix, öre, oxer, pace, pact, panel, pax, pickles, piste, place, plannen, praxis, pre, prise, qua, quaker, queeste, quine, quitte, raider, rance, re, record, roux, sext, shag, shave, shelter, shovel, shuttle, si, sic, sick, sine, slangspencer, sponsor, stance, stayer, sticker, stilton, stoned, straight, stress, stuff, tackle, tandem, taupe, teckel, thermen, these, thyrs, titel, titer, toast (also *toost*), toss, tracer, trailer, trainen, tripel (with the variant *triple* not listed in [*Woordenlijst 1995*]), ut, vice, vide, voute, vox, whist, witz, wow, yam, yes, yoghurt

These words together with those in (1) will be marked as exceptionally non-native. All other words with a non-native spelling are correctly marked as such by the criteria.

Native words marked as non-native
There are some words that are treated as native by phonological and morphological rules but marked as non-native by the criteria give here. This can be explained by the morphological structure of the words: the criteria are meant to apply to morphemes, so if a native word is complex, it can exceed these properties.

The first type of words that is really derived are pseudo-compounds or derived

words of which at least one of the morphemes is no longer recognizable. Some examples are *aalmoes*, *ambacht*, *antwoord*, *arbeid*, *argwaan*, *armoe(de)*, *assepoester*, *avond*, *ballast*, *balorig*, *eega(de)*, *eerbied*, *eiland*, *elkaar*, *ellende*, *gramstorig*, *havezaat*, *hertog*, *juffrouw*, *jullie*, *kleinood*, *leewater*, *lichaam*, *meineed*, *misschien*, *(n)iemand*, *ooievaar*, *oordeel*, *ootmoed*, *sieraad*, *twaalf*, *vennoot*, *vijand*, *wanneer* and *zieltoog*. Then there are lexicalized phrases like *meneer*, *mevrouw*, *mejuffrouw*, (from *mijn heer* etc.); *tegelijk*, *tehuis*, *dewelke*, *dezelfde*, *jelui* (from *te huis* etc.). Reduplicative words or rhyming compounds as described for instance in Battus (1981:157–160), De Haas & Trommelen (1993) often also exceed the possibilities of native words, but they can be considered compounds. Some examples are the following: *wirwar*, *hoteldebotel*, *wissewasje*.

Another group of exceptional words consists of abbreviations and acronyms like *ohaën*, *havo*, *aow'er*, *abc*, *raio* (that sometimes behave as non-native words, e.g. *buma-bumatisch*); interjections like *hola* and *kiekeboe* or onomatopoeia like *miauw*, *blèren*, *kievit*, *iaën* and *tsjilpen.*

All these words will be considered (complex) native words. The words mentioned above that have a property that exceeds the restrictions on native words, but are treated as native all the same will also be considered to be exceptions: *broer*, *burcht*, *deern*, *doorn*, *elf(t)*, *ernst*, *garnt*, *hauk*, *havik*, *helft*, *herfst*, *hoorn*, *knecht*, *knurft*, *kok*, *maat*, *murw*, *olms*, *oom*, *pauk*, *schurft*, *vijand*, *voorn*, *wormt*, *zoon*.

Words that contain native and non-native morphemes
Finally, since compounds can be composed of native and non-native words, a third label 'mixed' was introduced for words such as *citruspers* (non-native *citrus* and native *pers*) and *rekenmachine* (native *reken* and non-native *machine*). As remarked above, the combination of native stems with non-native affixes is marked. Such words, e.g. *kretologie*, *flauwiteit* were also labelled 'mixed'. Fifty per cent of the words in the lexicon was native, 49% non-native and 1% mixed.

Conclusion

In this appendix the set of rules was presented that defines native words. Only a small group of words remains that is native according to the criteria, but has a idiosyncratic spelling. Most native words that are incorrectly classified as non-native turn out to be complex. When the fact that they are not morphemes is taken into account, they are correctly accepted as native words. The extent of adaptation to the Dutch linguistic system is a good predictor of non-native spelling behaviour.

Appendix C
Phoneme-to-grapheme conversion rules for native words

This appendix gives an overview of phoneme-to-grapheme conversion rules for native words with their exceptions. For every rule, a representative example is provided between parentheses. Furthermore, native exceptions are exhaustively listed, but exceptional words that are formally native but etymologically loan words are not given here, see Appendix B. Spelling variants are separated by slashes. In order to show that no further generalizations can be made about the spelling of /ɛi/ and /ɑu/, homophones are added between parentheses, e.g. *nauw* (*nou*) for 'narrow' and 'now'. Vowel letters representing short full vowels are associated to one V-position; all vowel letters representing long vowels or diphthongs are associated to two V-positions, but this is only indicated for <i> and <ie>. Vowel letters which correspond to schwa are not associated to a V-position. Consonant monographs are linked to one C-position, except for <x>, which is linked to two C-positions.

(1)　**Rule for /ɪ/**

```
        V
        |
/ɪ/ → i
```
　　　　　　　　　　　　　　　　　　　　　　　(*pit*)

(2)　**Rule for /e/**
　　　/ɛ/ → e　　　　　　　　　　　　　　　　　(*wen*)
　　　EXCEPTIONS: *hè.*

(3)　**Rule for /ɔ/**
　　　/ɔ/ → o　　　　　　　　　　　　　　　　　(*mot*)
　　　EXCEPTIONS: *goh*, *joh.*

(4)　**Rule for /ʏ/**
　　　/ʏ/ → u　　　　　　　　　　　　　　　　　(*punt*)
　　　EXCEPTIONS: *puh*, *löss.*

(5) **Rule for /ɑ/**
/ɑ/ → a   (*ham*)
EXCEPTIONS: *bah.*

(6) **Rule for /i/**
       vv
       | |
/i/ → ie   (*bier*)
EXCEPTIONS: *bijzonder*, *ki(e)vit.*

(7) **Rule for /y/**
/y/ → uu   (*vuur*)

(8) **Rule for /u/**
/u/ → oe   (*roem*)

(9) **Rule for /e/**
/e/ → ee   (*week*)
EXCEPTIONS: *hé*, *heir-*, etc.

(10) **Rule for /∅/**
/∅/ → eu   (*beuk*)

(11) **Rule for /o/**
/o/ → oo   (*hoop*)
EXCEPTIONS: *zo'n*, *oir.*

(12) **Rule for /a/**
/a/ → aa   (*jaar*)

(13) **Rule for /ɛi/**
a /ɛi/ → ei in *-heid*, *-(er)lei* and in the words listed below
b /ɛi/ → ij elsewhere   (*tijd*)

WORDS WRITTEN WITH <EI>

arbeid, bei (*bij*), *beiaard, beide, beidjes/beitjes* (*bijtjes*), *beiden, beitel, beits, bimbambeieren, blei* (*blij*), *blein, brei-brij, breidel, brein* (*brijn*), *dein, deinzen, deis, dreigen, dreinen, dweil, ei* (*ij*), *eiber, eidereend, eigen, eik-ijk, eiker-ijker, eiland,*

eind, eins or *einze, einsel, eis* (*ijs*), *eit, feil, feit* (*fijt*), *geheim, gei* (*gij*), *geil* (*gijl*), *gein* (*gijn*), *geiser, geit, gerei, gevlij* (*gevlei*), *gewei, glei* (*glij*), *gleis, gneis, grein* (*grijn*), *hei* (*hij*), *heibei, heibel, heide, heiden, heil, heiland, heim(wee), hein, heinde, heister, heitje, ei,* (*ij, y*), *jein, kei, keil(der), keilen, keizer, klei, klein, koddebeier, lei* (*lij*), *leiden* (*lijden*), *leis* (*lijs*), *leist* (*lijst*), *meid-mijd(en), meiler, meineed, mij* (*mei*), *neigen* (*nijgen*), *peiger, peil* (*pijl*), *peins, peis, plein, pleister, pleit, pleite, poppedeintje, prei* (*prij*), *reiger* (*rijger*), *reik* (*rijk*), *reilen, rein* (*rijn*), *reis* (*rijs*), *rei* (*rij*), *schei, schei* (*schij*), *scheiden, scheil, schreien, seider, sein, seis* (*sijs*), *sintveitsdans, sjeik* (*sjijk*), *slei, sprei, spreiden, steiger* (*stijger*), *steil* (*stijl*), *teil, teisteren, treil, trein, treiteren, vei, veil, veil* (*vijl*), *veilig, veins, vleien* (*vlijen*), *wei* (*wij*), *weid(en)* (*wijd*), *weifelen, weigeren, weinig, weit, weit-wijt(en), wijd(e)-weide, weiden* (*wijden*), *zeik, zeil* (*zijl*), *zeis, zwei.*

(14)   **Rule for /œy/**

        /œy/→ui   (***ruik**er*)

(15)   **Rules for /ɑu/**

                     VVC

                     | | |

     a   /ɑu/   → auw / _ {V, +} in the morphemes listed below

                     VVC

                     | | |

     b   /ɑu/   → ouw / _ {V, +}   (***rouw***)

     c   /ɑu/   → au / _ C in some morphemes   (*kl**au**ter*)

     d   /ɑu/   → ou> / _ C elsewhere   (*k**ou**d*)

          EXCEPTIONS: *au   jou, kou, nou, wou, zou.*

          WORDS WRITTEN WITH <AU(W)>

          bietebauw, klauwen, knauw, krauw, krauwel, lauwer, miauwen, pauw, prauw, snauw, wauwel, (wenk)brauw, wiewauwen; bauwen (*bouwen*), *blauwen* (*blouwen*), *dauw* (*douw*), *flauw* (*flouw*), *gauw* (*gouw*), *grauw* (*grouw*), *hauw* (*houw*), *kauw* (*kouw/kou*), *lauw* (*louw*), *mauw* (*mouw*), *nauw* (*nou*), *rauw* (*rouw*), *schauw* (*schouw*), *wauw* (*wouw/wou*); *rauwdouwer/rouwdouwer* (both variants allowed). *au, klauter, lauden, mauser, paus, pauze, rauzen/rausjen, saus.*

(16)   **Rules for /ə/**

     a   /ə/ → ij   in *-lijk*   (*mogel**ij**k*)

     b   /ə/ → i   in *-ig*   (*gezell**i**g*)

     c   /ə/ → e   elsewhere   (*ad**e**r, hann**e**s*)

          EXCEPTIONS: *z'n, m'n, d'r, 'r, 't, 'm, 'n;* ***een, he**t, loeres/loeris,*

stennis/stennes, *jottum*, *dattum*, *lorum*, *krankjorum*, *uh*, *eh*; *dikwijls*.

(17) **Rule for /p/**
/p/ → p (***p**as*)
EXCEPTIONS: *ab**t***, *bi**b**s/bi**p**s*.

(18) **Rule for /b/**
/b/ → b (***b**al*)

(19) **Rules for /t/**
a /tsj/→ tj (***tj**onge*)
b /t/ → t (***t**aart*)
EXCEPTIONS: A: ***tj**irpen/**tsj**irpen*, ***tj**ilpen/**tsj**ilpen*, *ha**ts**jie*. B: *sin**d**s*, *loo**d**s*,
*va**d**sig*, *fon**d**s*, *gi**d**s*; *iemand**d***, *niemand**d***, *-er**d***, *Helmond**d**, Simpelvel**d***,
*Lelysta**d**, Willemsta**d***; ***th**ee*, *lu**th**er*, *(al)**th**ans*, ***th**uis*.

(20) **Rule for /d/**
/d/ → d (***d**eur*)

(21) **Rule for /k/**
/k/ → k (***k**ort*)
EXCEPTIONS: *we**ck**en*, *bo**ck**bier*, *te**k**kel/te**ck**el*.

(22) **Rule for /f/**
/f/ → f (***f**iets*)
EXCEPTIONS: ***v**eertig*, ***v**ijftig*, ***v**eter*.

(23) **Rule for /v/**
/v/ → v (***v**eel*)

(24) **Rule for /s/**
/s/ → s (***s**uf*)
EXCEPTIONS: ***z**estig*, ***z**eventig*; *Bos**ch***, *dros**ch**ke*, *sst* (an interjection
not listed in [*Woordenlijst 1995*]), *mi**ss**chien*, *janklaa**ss**en*.

(25) **Rule for /z/**
/z/ → z (***z**aak*)
EXCEPTIONS: *knie**s**oor*; *poe**s**en/poe**z**en*, *on**s**en/on**z**en*, *die**s**el*, *potverdo**s**ie*.

(26)   **Rule for /x/**
                    C
                    /\
          /x/ → ch   (la***ch***)
          EXCEPTIONS: *aa**gt**, oo**gst**, no**g**, geneu**gt(e)**, ste**gg**el, genoe**g**elijk, ho**g**elijk,*
          *kla**g**elijk, va**g**elijk, wal**g**elijk, zor**g**elijk, ontie**g**elijk, getui**g**enis, heu**g**enis.*

(27)   **Rule for /h/**
          /h/→   h   (***h***aat)

(28)   **Rule for /ɣ/**
          /ɣ/ → g   (***g***eld)

(29)   **Rule for /m/**
          /m/→ m   (***m***ooi)
          EXCEPTIONS: *aa**n**beeld* (variant of *aambeeld*).

(30)   **Rule for /n/**
          /n/ → n   (***n***oord)

(31)   **Rules for /ŋ/**
       a /ŋ/ → n / _ k   (da***n***k)
                  CC
                  ||
       b /ŋ/ → ng   elsewhere   (ba***ng***)

(32)   **Rule for /l/**
          /l/ → l   (***l***us)

(33)   **Rule for /r/**
          /r/ → r   (***r***iem)

(34)   **Rule for /w/**
       a /w/→ w / y _ +   (du***w***)
                  CC
                  ||
          /w/→ uw / V _ +   (ee***uw***)

b   /w/ → w   elsewhere   (**w**aar)
EXCEPTIONS: a: *leewieken, wiewauwen*   b: *krieuwel.*

(35)   **Rules for /j/**
                    C
                    |
a   /j/ → i / _ +   (aa**i**)
b   /j/ → j   elsewhere   (**j**aar, ba**j**es)
EXCEPTIONS: A: *poeier, ooievaar, kooiker, lichterlaaie*; b: *verstajem.*

FURTHER REMARKS
Exceptions with silent letters are *am**b**t, er**w**t, het, likmeves**t**je, staakt-het-vuren*. There is one native word (an onomatopoeia) in which the sound /ɛː/ occurs: [blɛːrə]. This word is written as *bl**è**ren*.

Appendix D
Phoneme-to-grapheme conversion rules for non-native words

This appendix gives an overview of sound to-letter conversion rules for non-native words with their exceptions. Rules for sound combinations are listed under the first sound. The overview ends with some rules for specific semi-affixes. Exceptions are exhaustively listed when there are five or less examples; in other cases some examples are given followed by *etc.* Vowel letters representing short vowels are linked to one V-position; vowel letters and vowel digraphs representing long vowels are linked to two V-positions.

Most foreign phonemes do not warrant the postulation of spelling rules, because they do not occur in the transcription provided by CELEX, or only sporadically, or because their spelling is too idiosyncratic. For this reason, I will only formulate rules for /g/ and some semi-affixes which contain foreign phonemes. All other words which contain these sounds are considered to be exceptional.

NOTATION:

'+'	morpheme boundary (does not include non-native sufixes)
V, FV	Vowel, Full vowel (all vowels but schwa)
LV, SV	Long vowel, short vowel
SON, OBS	Sonorant consonant, obstruent
ISUF	-i (*-ie*) and non-native suffixes beginning with /i/ -io (*-io*), -iɔn (*-ion*), -iʏs (*-ius*), -iʏm (*-ium*), -iel (*-ieel*), -ier (*ieer*), -ial- (*iaal, -ial-*), -iar- (*iaar, -iar-*), -iat- (*-iaat, iat-*), -iɛːr (*-air*), -ɪ∅r (*-ieur*), -ɪ∅s (*-ieus*).

(1) **Rules for /ɪ/**

/ɪ/ → i   (*flamingo*)

EXCEPTIONS: *apocalyps, crypt, lymf*, etc.; *(body)building, business, seigneur, sovereign*; *ukelele*.

(2) **Rules for /ɛ/**

/e/ → e   (*pamflet*)

EXCEPTIONS: *back, racket*, etc.; *heavy, ready*, etc.; *maison, relais*, etc.;
quaestor, quaestuur, quaestrix; *arrêt, crêpe*; *fraîcheur, maîtresse*; *crayon,
relayeren*; *ceintuur, peigné*; *knäckebröd, ländler*; *calèche, crèche*.

(3) **Rules for /ɔ/**

/ɔ/ → o   (*karton*)

EXCEPTIONS: *brainwashing, squadron, squash, stopwatch, swap, warrant(y)*;
poinsettia; *remouldband*; *ångström-eenheid*; *smörgåsbord*.

(4) **Rules for /ʏ/**

/ʏ/ → u   (**u**ltra)

EXCEPTIONS: *röntgen, smörgåsbord, smörrebröd*; *double, trouble*;
floodlight; *company, government, money*.

(5) **Rules for /ɑ/**

/ɑ/ → a   (**a**tleet)

EXCEPTIONS: *caissière, douairière*. In combination with j: *design, sliding,
trial*, etc.; *bypass, byte*, etc.; *taekwondo, maestro*; *balalaïka, maïs, maïzena*;
edelweiss, pfeiffer, poltergeist; *weihnachtsstol*; *diehard, tie-break*; *eye*.

(6) **Rules for /i/**

a /i/ → ie / _ C₀ +   (*dimensie*)

b /i/ → i   elsewhere   (*bolide*)

EXCEPTIONS TO THE SUBRULES:

a *maxi-, peri-*, etc. (prefixes), *taxi, macaroni*, etc.; *carbid, codicil*, etc.

b *dixieland, faliekant, iezegrim, mediene, retrieval, sowieso, spielerei,
spielmacher, vlieseline*; *flottielje, gienje, trielje*.

a/b *design, would-be*, etc.; *hyena, sherry*, etc.; *hockey, keyboard*, etc.;
gruyère; *sirih*; *parenchijm*.

(7) **Rules for /y/**

a /y/ → uu / _ C₀ +   (*azuur*)

b /y/ → u elsewhere   (*nuance*)

EXCEPTIONS TO THE SUBRULES:

a *jus, luxe, recul*, etc.   a/b *bruuskeren*; *fondue, revue*, etc.;
muesli/müsli; *hybris*; *rücksichtslos, überhaupt, Übermensch*.

(8)   **Rules for /u/**
   a   /u/ → ou / _ …∅ {r, z} (ə)   (*fou*r*eur*)
   b   /u/ → ou / _ …y r (ə)   (*cou*t*ure*)
   c   /u/ → ou / _ … i n (ə)   (*mou*s*seline*)
   d   /u/ → ou / _ … e   (*bou*c*lé*)
   e   /u/ → ou / r _ {l, t}   (*rou*t*e*)
   f   /u/ → ou / k _ {p, r}   (*cou*p)
   g   /u/ → ou in words with foreign phonemes   (*rou*g*e*)
   h   /u/ → ou / _ j   (*bou*i*llon*)
   i   /u/ → ou / _ … sj   (*tou*c*heren*)
        CONDITION: sj is not preceded by t
   j   /u/ → oe   elsewhere   (*koe*l*ie*)
   EXCEPTIONS TO THE SUBRULES:
   g   *foe*r*age(ur)*, *foe*r*gon*, *mango*e*stan*, *mango*e*ste*, *moe*z*jiek*
   i   *pu*s*h*.
   j   *sou*p*lesse*, (*camou*f*lage*), etc.; *antepenu*l*tima*, *ju*n*ta*, etc.;
        *manoeu*v*reren*; *grapefr*ui*t*, *wets*ui*t*; *cashe*w*noot*, *cr*e*w*, *scre*w*driver*,
        -*vie*w, etc.; *bl*ue*s*; *sc*oo*ter*, *vo*o*doo*, etc.; *m*o*ve*, *rem*o*ver*, *two*-*seater*.

(9)   **Rules for /e/**
   a   /e/ → ee / _ C$_0$ +   (*id*ee)
   b   /e/ → e   elsewhere   (*he*l*aas*)
   EXCEPTIONS TO THE SUBRULES:
   a   *acn*e, *benzo*ë, *cantabil*e, *facsimil*e, etc.; *cabar*e*tier*, *doss*i*er*, etc.; *fid*e*s*;
   *beign*e*t*, *fil*e*t*; *foy*e*r*, *touch*e*r*, etc.; *cache*-*n*e*z*, *rend*e*z*-*vous*, *pince*-*n*e*z*,
   *laiss*e*z*-*passer*; *pied*-*à*-*terre*, *souspi*e*d*.
   a/b *coup*é, *déd*ai*n*, etc.; *br*ea*k*, *st*ea*k*, etc.; *ess*ay, *pl*ay*boy*, etc.; *l*ae*sie*,
   *r*e*ggae*, etc.; *cr*a*zy*, *esc*a*pe*, etc.; *cl*ai*m*, *cockt*ai*l*, etc.; *cruz*ei*ro*, *mad*ei*ra*,
   *p*ei*gnoir*; *k*ey*nesiaans*, *surv*ey; *salonf*ä*hig*; *sansev*ie*ria*.

(10)   **Rules for /∅/**
   /∅/ → eu   (*lik*eu*r*)
   EXCEPTIONS:
   *am*oe*be*, *fo*e*tus*, oe*cumene* etc; *fr*ö*belen*, *knäckebr*ö*d*, etc.; *f*öh*n*.

(11) **Rules for /o/**

   a   /o/ → oo / _ C$_0$ +   (*ik**oo**n*)

   b   /o/ → o   elsewhere   (*ar**o**ma*)

EXCEPTIONS TO THE SUBRULES:

   a   *dep**o**t*, *palet**o**t*, etc.; *tourned**o**s*, *(a)prop**o**s*; *ten**o**r*, *rücksichtl**o**s*.

a/b *a**u**bade*, *crap**au**d*, etc.; *bur**eau***, *niv**eau***, etc.; *appr**oa**ch*, *g**oa**l*, etc.;
    *fell**ow***, *kn**ow**how*, etc.; *bord**eau**x*, *lits-jum**eau**x*; *mistlet**oe***, *t**oe**clip*;
    ***oh**m*; *behavi**ou**risme*, *rig**ou**reus*, *s**ou**l*.

(12) **Rules for /a/**

   a   /a/ → aa / _ C$_0$ +   (*escul**aa**p*)

   b   /a/ → a   elsewhere   (***a**zijn*)

EXCEPTIONS TO THE SUBRULES:

     a   *(milli)b**a**r*, *bl**a**gue*, *boulev**a**rd*, *branc**a**rd*, *ec**a**rt*, *eg**a**rd*, *flamb**a**rd*,
       *foul**a**rd*, *jacqu**a**rd*, *isl**a**m*, *k**a**n*, *kor**a**n*; *hang**a(a)**r*, *samow**a(a)**r*

     a/b   *vis-à-vis*, *voil**à***, etc.; *betj**ah***, *fell**ah***, *j**ah**veh*, *m**ah**dist*, *m**ah**jong*,
       *sj**ah***, *y**ah**tzee*.

(13) **Rules for /ɛi/**

    /ɛi/ → ij   (*tap**ij**t*)

EXCEPTIONS: *kal**ei**doscoop*, *r**ei**ne-claude*, *terr**ei**n*, etc.; *n**y**lon*; *erlenm**ey**er*.

(14) **Rules for /œy/**

/œy/ → ui   (*plav**ui**s*)

EXCEPTIONS: *d**eu**terium*, *d**eu**teron*, *farmac**eu**tisch*, *f**eu**illeton*,
*hermen**eu**tisch*, *n**eu**tr-*, *therap**eu**t*, *z**eu**gma*; *trompe-l'**oei**l*.

(15) **Rules for /ɑu/**

   a   /ɑu/   → auw / _ +   (*kabelj**auw***)

   b   /ɑu/   → ou / _ n [+cor −son]   (*s**ou**nd*)

   c   /ɑu/   → au / _ elsewhere   (*l**au**rier*)

EXCEPTIONS TO THE SUBRULES:

   a   *ken**au***; *cac**ao***, *curaç**ao***; *flamb**ou**w*, *karb**ou**w*, *kart**ou**w*, *kers**ou**w*,
    *land**ou**w*; *know-h**ow***, *high-br**ow***.   b   *a**u**straal*, *cl**au**str-*, *exh**au**st-*,

   c   *verb**ou**wereerd*, *h**ou**weel*, *h**ou**witser*; *kab**ou**ter*, *l**ou**dspeaker*, *sc**ou**t*;
    *cl**ow**n*, *d**ow**n*, etc.

(16)   **Rules for /ə/**
/ə/ → e   (*amandel*)
EXCEPTIONS: *achenebbisj, elixir, service, sheriff, spirit; recital, privacy,*
package, etc.; *bacon, counselor;*
glamour; rock-'n-roll; belvédère, dégénéré, ingénue, procédé, revérence;
gênant; yoghurt, column, tilbury; single, shuffle, etc.; *wire, umpire.*

(17)   **Rules for /p/**
/p/ → p   (*kompas*)

(18)   **Rules for /b/**
/b/ → b   (*framboos*)

(19)   **Rules for /t/**
a   /(t)sj/ → ti / LV _ ISUF   (*spatieus*)
b   /(t)sj/ → ti / SON _ ISUF   (*conventioneel*)
c   /tsj/   → tch / LV _   (*match*)
d   /tsj/   → ch elsewhere   (*speech, chester*)
e   /ts/   → t / LV _ ISUF   (*spatie*)
f   /ts/   → t / SON _ ISUF   (*conventie*)
g   /ts/   → ts elsewhere   (*tsaar, strapats*)
i   /t/   → t elsewhere   (*elegant*)
EXCEPTIONS TO THE SUBRULES:
d   *slivovitsj, tsarevitsj, katsjoe, gletsjer.*
e/f *staatsie, intelligentsia.*
g   *grazioso, scherzo, terzine, zarzuela,* etc.; *beneficium, -cento* (e.g.
    seicento), sedecimo, silicium, etc.; *forsythia.*
i   *cadmium, ordner, podsol, wadjan(g), wodka,* etc.; *quitte, settler, watt;*
    katholiek, litho, thyrs, etc.; *budget, gadget, midget, pidgin.*

(20)   **Rules for /d/**
a   /dzj/   → j   (*jeep*)
b   /d/ → d   (*idioot*)
EXCEPTIONS TO THE SUBRULES:
a   *dzjinn, hadzj(i); gentleman, gipsy, gin, lounge, teenager; hodgkin,*
    wedgwood; djahé, djakke, djaksa, djami, djati, djeroek, djimat, hadj(i);
    leggiero; deuce; deweysysteem.

(21)   **Rules for /k/**

a   /ks/   → ct / _ ISUF   (*functie*)
b   /ks/   → cc / a _ [−back]   (*accent*)
c   /ks/   → x elsewhere   (*max*)
d   /kw/   → kw / _ …{-un, -ɛin, -ir, oj, aj}   (*kwartier*)
     CONDITION: not in -*ulier/-urier*
e   /kw/   → qu elsewhere   (*quasi*)
f   /k/   → k / _ (….k) +   (*kaduuk*)
g   /k/   → k / _ (….k) [−back]   (*spektakel*)
h   /k/   → k / _ …{-un, -ɛin, -ir, oj, aj}   (*kalkoen*)
i   /k/   → c in words with foreign phonemes   (*courgette*)
j   /k/   → k / _ … œy   (*kazuifel*)
k   /k/   → c elsewhere   (*cadeau*)
     EXCEPTIONS TO THE SUBRULES:
     a/b *axiaal*, *axio-*, *fluxie*, *flexie*, *taxi* etc.; *axillair*.
     c   *djaksa*, *(d)oksaal*, *feniks*, *lariks*, *seks*, *tekst*; *fuchsia*, *fuchsine*;
          eczeem, *eczema*; *occident*, *succes*, *vaccin*; *slacks*; *aerobics*; *cokes*,
          cornflakes; e*xcel-*, e*xcentri-*, e*xcepti-*, e*xcerperen*, e*xcerpt*, e*xces*
          e*xciteren*; e*xcessief*. d   *aquamarijn*, *quadriljoen*
     e   *acquisitie*; *biscuit*, *cuisine*, *circuit*; *coiffure*; *qwerty*; *kwadraat*,
          kwartaal, *taekwondo*, etc.
     f   *gimmick*, *shock*, etc.; *loc*, *plastic*, *tonic*, *truc*, etc.; *cheque*, etc.;
          high-tech.
     g   *maquette*, *equipe*, etc.; *centaur*, *ceramiek*, *cyst-* (*cystoscoop*),
          ecce-homo, *rovescio*, *sacerdotaal*, *scep-*; *sofisticeren*, -*coccen*, *cai-*
          (*precair*, *caisson*), *handicap*; *rocheren*, *zucchetti*, *zucchini*.
     h   *benedictijn*, *sacristein*; *brancardier*, *cavalier*; *octrooi*.
     i   (*elektricien*), *katalyse*
     k   *acquit*, *jacquet*, *jacquard*; *chianti*, *chorus*, *chroniqueur*, *christ-*
          (*christologie*); *quarantaine*, etc.; *picknick*, etc.;
          (different spellings of /k/ within a word: *ekwinoctiaal*, *elektrificatie*,
          elektronica, *kinetica*, *kosmetica*, *kwalificatie*, *macrokosmos*,
          microkosmos, *skepticus*).

(22)   **Rules for /f/**
          /f/ → f   (*forens*)
          EXCEPTIONS: *dauphin*, *sophisticated*; *sheriff*; *kevlar*, *sovchoze*, *sovjet*;
          bath, *rebirthing*.

(23)   **Rules for /v/**
          /v/ → v   (*v*iool)

(24)   **Rules for /s/**
     a   /sj/ → ch / OBS _ er   (mar*ch*eer)
     b   /sj/ → ti / SON _ ISUF   (op*ti*oneel)
     c   /sj/ → ci / SON _ ISUF   (commer*ci*eel)
     d   /sj/ → ci / LV _ ISUF   (spe*ci*aal)
     e   /sj/ → si / SV _ ISUF   (mis*si*onair)
     f   /sj/ → ch elsewhere   (*ch*arge)
     g   /s/ → t / OBS _ ISUF   (op*t*ie)
     h   /s/ → c / LV _ [−back]   (spe*c*ie)
     i   /s/ → s elsewhere   (per*s*oon, mis*s*ie)
          EXCEPTIONS TO THE SUBRULES:
     a   appre*ci*ëren, asso*ci*ëren, benefi*ci*ëren, denun*ci*ëren, inji*ci*ëren,
          offi*ci*ëren
     d   pen*si*oen, pen*si*on.
     f   kit*sch*, **sch**lager; fa*sc*isme; an*sj*ovis; arti*sj*ok, bol*sj*e, ha*sj*(ie*sj*),
          ka*sj*mier, men*sj*ewiek, pa*sj*a; derwi*sj*, feti*sj*, koo*sj*er/kou*sj*er,
          me*sj*oche/me*sj*okke, rik*sj*a; ram*sj*, achenebbi*sj*; **sj**abrak, **sj**iiet, **sj**wa,
          etc.; fini*sh*, ka*sh*ba, **sh**ampoo, etc.; **s**pielerei, **s**pielmacher, **s**tuka;
          is*s*ue, tis*s*ue, **s**uit; ru*ch*e, plu*ch*e.
     g   autop*s*ie, eplilep*s*ie, etc.
     h   ar*s*een, kor*s*et, para*s*iet, por*s*elein, etc.; ab*sc*is, di*sc*ipel, ob*sc*een, etc.
     i   ra*c*e, servi*c*e, etc.; *c*onnaisseur, moe*s*son, pali*s*sade, etc; pas*s*, tople*s*s,
          etc.; haus*s*e, mous*s*e, sauci*s*se, sui*s*se; *c*ourse, (re)lea*s*e, respon*s*e,
          suspen*s*e; er*s*atz, hert*z*, wit*z*; bos**sch**age; aper*ç*u, fa*ç*ade, gar*ç*on,
          ma*ç*on, re*ç*u, rempla*ç*ant; mis*t*letoe; ac*c*esoire.

(25)   **Rules for /z/**

    a   /zj/ → g / _ er   (*chargeer*)
    b   /zj/ → si / _ ISUF   (*visionair*)
    c   /zj/ → j / _ [+back]   (*jus*)
    d   /zj/ → g elsewhere   (*genre*)
    e   /z/ → s / LV _ V   (*fusie*)
    f   /z/ → s / SON _ V   (*mensa*)
    g   /z/ → z / _ … …{-un, -ɛin, -ir, oj, aj}   (*azijn*)
    h   /z/ → z / _ +   (*precieze*)
    i   /z/ → z / _ in words with foreign phonemes (*mayonaise*)
    j   /z/ → z elsewhere   (*zodiak*)

    EXCEPTIONS TO THE SUBRULES:

    a   *privilegiëren, solfegiëren.*
    c   *dejeuner, jeu, oranjerie, sujet.*
    d   *prodigieus, refugié, stagiair; bourgeois, flageolet; gendarm, orangeade, sergeant; audio-visual; moezjiek.*
    e   *markizaat, oekaze, pauze, poëzie,* etc.
    f   *bamzaaien, barzoi, beëlzebub, benzine, benzoë, bremze, enzym, influenza, rapunzel, sperzieboon, terzet.*
    g   *fuselier, thesaurier, visioen.*
    h   *confuus, divers, diffuus, excuus, exquis, infuus, refuus* (the preceding words have inflected forms with <s>: *confuse,* etc.), *leguminosen.*
    i   *roze.*   j   *sowieso; acacia, coulisse.*

(26)   **Rules for /x/**

    /x/ → ch   (*echo*)
    EXCEPTIONS: *baghera, junta*

(27)   **Rules for /ɣ/**

    /ɣ/ → g   (*spagaat*)
    EXCEPTIONS: *sorghum, spaghetti*

(28)   **Rules for /h/**

    /h/ → h   (*heraut*)

(29)   **Rules for /m/**

    /m/ → m   (*eminent*)
    EXCEPTIONS: *schlamm*

(30)   **Rules for /n/**
a   /nj/→ gn / FV… _ …FV   (*signaal*)
b   /nj/→ gn / FV…FV…_   (*bourgogne*)
c   /nj/→ nj elsewhere   (*franje*)
d   /n/ → n   (*concept*)
EXCEPTIONS TO THE SUBRULES:
   a/b *banjo*, *benjamin*, *njonja*; *kampanje*, *karonje*, *kastanje*, *kokinje*,
       oranje.
   c   *doña*, *señorita*; *pinyin*; *piranha*; *besogne*.
   d   *sennhut*, *dzjinn*; *bonhomme*, *bonhomie*.

(31)   **Rules for /ŋ/**
a   /ŋ/ → n / _ k   (*franco*)
b   /ŋ/ → ng elsewhere   (*pisang*)

(32)   **Rules for /l/**
a   /lj/ → lj   (*biljart*)
b   /l/ → l   (*bolero*)
EXCEPTIONS TO THE SUBRULES:
a   *milieu.*   b   **Lloyd**; *bulldozer*, *grill*, etc.

(33)   **Rules for /r/**
/r/ → r   (*caramel*)

(34)   **Rules for /w/**
a   /w/→ w / y _ +   (*baljuw*)
b   /w/→ u w / V _ +   (*flambeeuw*)
c   /wa/, /wɑ/ → oi / _ {r, s}   (*trottoir*)
d   /w/→ w elsewhere   (*warande*)
EXCEPTIONS TO THE SUBRULES:
c   *croissant*, *exploiteren*, etc.; **one-man-show**; *foyer*, *voyeur*.
d   *conduite*, *etui*, *gratuit*, *nuit*, *suisse*, *(pour)suite*. N.B [w] is written after
back vowels in: *fluweel*, *fluwijn*, *houweel*, *houwitser*, *juweel*, *kauwoerde*,
klauwier, *saguweer/sagoweer*, *samowa(a)r*, *stouwage*, *struweel*, *stuwadoor*,
stuwage, *truweel*, *verbouwereerd*.

(35)   **Rules for /j/**

    a   /j/ → i / _ +   (*lawaai*)

    b   /j/ → i / _ ISUF   (*principieel*)

    c   /j/ → ll / {εi, œy, ɑ, u} _ V   (*surveilleer, bouillon*)

    d   /j/ → ll / {i, ɪ} _ {o, a, ɔ} +   (*carrillon*)

        /j/ → ll / ɪ _ +   (*vanille*)

        CONDITION: i is stressed

    e   /j/ → j elsewhere   (*junior*)

EXCEPTIONS TO THE SUBRULES:

    e   *yahtzee*, *yoghurt*, etc.; *attenoj*, *goj*; *boy*, *corduroy*, *tinneroy*; *iatrosoof*, *ion*; *ayatollah*, *coyote*, *maya*, *riyal*; *(barbe)cue*; *new*, *steward(ess)*; *computer*, *executive*, etc.; *beauty*; *feature*, *mixture*, *picture*, *pressure*.

        N.B. [j] is written after front vowels in *ejaculaat*, *ejaculatie*, *ejaculeren*, *ejecteur*, *ejector*, *pejoratief*, *plebejer*, *plebejisch*, *plejaden*, *pyjama*, *riyal*and the words in (35c-d).

(36)   **Rules for /g/**

    a   /g/ → gu / _ [−back]   (*sanguine*)

    b   /g/ → g elsewhere   (*goal*)

EXCEPTIONS TO THE SUBRULES:

    a   *boogie-woogie*, *gearing*, *gimmick*, *goalgetter*, *manggis(tan)*, *vigeur*.

    b   *bodyguard*, *guano*, *jaguar*, *linguaal*; *egg.*

(37)   **Rules for specific (semi-)suffixes**

a   /sjɛ̃/   → cien / _ +   (*opticien*)
b   /liːzə/ → lyse / _ +   (*analyse*)
c   /jɛː/   → iè / _ {m, rə}   (*première*)
d   /ɛːr/   → air (ə) / _ +   (*precair, stagiaire*)
e   /ɛːzə/   → aise / _ +   (*malaise*)
f   /ɛːro/   → aëro / _ +   (*aëronaut*)
g   /je/   → ier / _ +   (*premier*)
h   /ədzj/   → age / _ +   (*manage*)
i   /tɛit/   → teit / _ +   (*activiteit*)
j   /tɛk/   → theek / _ +   (*bibliotheek*)
k   /is/   → isch / _ +   (*thermisch*)

CONDITION: only holds for unstressed /is/
EXCEPTIONS TO THE SUBRULES:

a   *auspiciën*.
d   *camembert*, *colbert*, *couvert*, *dessert*, *expert*, *flobert*; *parterre*,
    pied-à-terre, *revers*, *serre*; *care*, *share*, *square*, *-ware*.

(38)   **Rules for specific (semi-)prefixes**

a	/sɪl/	→ syl / + _ V	(***syl***labe)
b	/sɪn/	→ syn / + _ C	(***syn***taxis)
		CONDITION: C is not /ɣ/ or /j/	
c	/sɪm/	→ sym / + _ {p, b, V}	(***sym***bool)
d	/psix/	→ psych / + _	(***psych***oloog)
e	/fiz/	→ fys / + _	(***fys***iek)
f	/hip/	→ hyp / + _ {o, ər}	(***hyper***actief)
g	/poli/	→ poly / + _	(***poly***theen)
h	/œy/	→ eu / + _	(***eu***thanasie)
i	/ap/	→ ab / + _ C	(***ab***stract)
		CONDITION: C is not /t/	
j	/ɔp/	→ ob / + _ C	(***ob***sceen)
		CONDITION: C is not /t/	
k	/sʏp/	→ sub / + _	(***sub***straat)
l	/ɑt/	→ ad / + _ C	(***ad***ministratie)
		CONDITION: C is not /t/ or /l/	
n	/tema/	→ thema / + _	(***thema***)
n	/terap/	→ therap / + _	(***therap***ie)
o	/teo/	→ theo / + _	(***theo***loog)
p	/ɔrto/	→ ortho / + _	(***orth***ografie)
q	/pat/	→ path / _ {o, i} {s, k}	(***path***os)

EXCEPTIONS TO THE SUBRULES:

b/c ***scin***ti- (e.g. ***scin***tillatie); ***sim***pel/***sim***pl-, ***cim***baal.

e   ad***vies***, con***fis***erie, ***fies***elemie.

g   ***poli***chinel, ***poli***o, ***poli***(kliniek), ***poli***tie(k), ***poli***tic-, ***poli***toer.

i/j ***ap***sis, ***op***soloog.   l   ***at***jar, ***at***mosfeer.

Appendix E
Overview of autonomous spelling rules

This appendix gives an overview of autonomous spelling rules with their exceptions. The rules are conditioned by orthographical context and apply to words. For every rule, a representative example is provided between parentheses.

NOTATION:　S = (orthographic) syllable　　L = Letter　　V = vowel letter
　　　　　　　C = consonant letter　　　　+ = morpheme boundary

(1) **Orthographic Syllabification**

Within syllabification domains (free morphemes, suffixes that start with a consonant and contain a vowel, and *-aard or -achtig*), syllables are built on the basis of the CV-tier in the following way:

a. Adjacent V-positions are heterosyllabic, provided that the first syllable has two V-positions: *kri.oel*, *ui.er* (**u.ier*).

b. One intervocalic C-position belongs to the next syllable, unless it is associated to a vowel letter: *be.ter*, *la.chen* but *maai.en*, *coy.ote*.

c. In the case of two or more intervocalic C-positions:

- if the C-positions are linked to the combinations *bl*, *br*, *cl*, *cr*, *dr*, *fl*, *fr*, *gl*, *gr*, *kl*, *kr*, *kw*, *pl*, *pr*, *qu*, *tr*, *vr* and *str*, this cluster is parsed as an onset: *a.qua*, *di.ploma*, *oe.strogeen*, etc. unless this means that the preceding syllable ends in a single V-position, in that case one C-position goes to the preceding syllable: *as.trant*, *rag.lan*, *tef.lon*, etc, except in case of the digraph *qu*: *choqueren*, *attaqueren*. Remaining letters belong to the preceding syllable: *ex.tra*, *in.strument*. *uw* is no onset, hence *eeu.wen*.
- if the C-positions are linked to *sc*, *sk*, *st* or *x*, they are split after vowel letters, but not after consonant letters: *fis.caal*, *bruus.keer*, *gees.ten*, *pas.ta*, *taxi*, *bauxiet* (the syllable boundary is between the two C-positions linked to *x*) versus *ob.sceen*, *ern.stig*, *in.stant*, *mar.xist*.

d. Remaining Cs go to the preceding syllable: *amb.ten*, *plank.ton*.
　　EXCEPTIONS: *naas.te*; *bas.taard*, *bo.gaard*, *do.laard*, *dros.saard*, *grijn.zaard*, *ho.vaar.dij*, *mos.taard*, *mut.saard*, *pon.jaard*, *stan.daard*, *tab.baard*,. *vein.zaard*; *trots.kist*

(2)   **Vowel Degemination**

$$V \qquad C$$
$$| \qquad |$$

$V_i \rightarrow 0 / V_i \_]_S [_S C_i$ (*later*)

CONDITION: Rule (2) does not apply to morpheme-final *ee*; it also applies to *uu* before *w*

EXCEPTIONS:native: *deemoed, leebraken, leewieken, eega(de), meekrap*
non-native: *teevee, tseetseevlieg, shampoo; weekend, cheeta* etc., *voodoo, bazooka*, etc. [1]

(3)   **Consonant Doubling**

$$V$$
$$|$$

$0 \quad \rightarrow C_i / [_{C_0} V \_]_S [_S C_i$ (*kommer*)

EXCEPTIONS: non-native: *condition, visual, body, cabaret(ière), caravan, chaperonne, column, cover, creditcard, desa, emplacement, everglaze, finish, colour, giveaway, hovercraft, impresario, intercity, officer, limerick, limit, living, manager, mêleren, military, mocassin, money, monocle, nasi, necessaire, novelty, omelet, penalty, polaroid, pony, rataplan, recovery, referee, robinsonade, sheriff, sovereign, sovjetoloog, spirit(ual), tomahawk; an+organisch, sub+altern* and other words with non-native prefixes.

(4)   **Consonant Degemination**

  a $[_S ...C_iC_i]_S \rightarrow C_i$ (*verbrand*)
  b $s \rightarrow 0 / \backslash s...\backslash \_$ (*logischte*)
  c $s \rightarrow 0 / x \_$ (*complexte*)
     CONDITIONS:
- in (4b) and (4c) the relevant letters are associated to at least one C-position
- In order to derive the contrast between words such as *fietste* and *fietsster*, syllabification and degemination are applied every time a suffix is added.[2]

---

[1] According to [*Woordenlijst 1954*], p. LII *bes-te* (**be-ste*), *mees-te* (**mee-ste*, cf. *gedwee-ste*) and *naas-te* (**na-ste*) are exceptional. In case of the first two examples, this may be attributed to the fact that we are dealing with irregular comparatives which are not composed of stem + *st,* cf. *goed-best, meer-meest,* but *na-naast* seems regular.

[2] Geminates such as those in *llano, Lloyd* and *watt, yell* where degemination does not work can be distinguished from cases such as **gehaatt* since they occur within morphemes and are linked to one C-

(5)   **Spelling Devoicing**

a $\backslash v\backslash \to$ <f> / _ C_0]$_S$ (*huis*, *leef*)

b $\backslash z\backslash \to$ <s> / _ C_0]$_S$

EXCEPTIONS TO THE SUBRULES:

a: *berlitzmethode, cache-nez, chintz, ersatz, fez, hertz, jazz, kolchoz, laissez-passer, pince-nez, quiz, razzia, rendez-vous, rez-de-chaussee, showbizz, slivovitz, weltschmerz, witz; mazda, samizdat.*

b: *leitmotiv; sovchoz(e), sovjet, kevlar, pravda.*

(6)   **Prevocalic E-deletion**

$\backslash e\backslash \to 0$ / _ + V   (*race-racen*)

CONDITION:

e is not linked to a V-position; V is part of a suffix

(7)   **Alternations of *i* and *ie***

a $\backslash ie\backslash \to$ <i> / _ V   (*olie-oliën*)

CONDITION:

only in unstressed syllables

b $\backslash i\backslash \to$ <ie> / _ + C   (*taxi-taxiede*)

CONDITION: C is part of a suffix

EXCEPTIONS TO THE SUBRULES:

b: *skister, etuitje, i'tje, pi'tje*

(8)   **Vowel doubling**

a   VV

    \/

$V_i \to V_i V_i$ / _ + tje   (*la-laatje*)

b   VV

    \/

$V_i \to V_i V_i$ / _ + C_1]$_S$ (*ga-gaat*)

CONDITIONS:

- V_i is *a, e, o* or *u*
- Only applied to the adjectival suffix *-s*, not the plural or genitive suffix *-s* (cf. een *Venloos straatje* versus *Venlo's straten*)
- (8a) is only applied in absence of hyphens

(9)   **Diminutive Allomorphy**

---

position.

\é\ → <ee> / _ + tje   (*cliché-clich**ee**tje*)

CONDITION: Only applied in absence of hyphen

LIST OF DIMINUTIVE ALLOMORPHS WITH SPECIAL SPELLING:

Deletion of silent letters: *colbert/colbertje, passe-partout/passe-partoutje, plafon(d)/plafonnetje,*
entremets/entremetje,   *entre-deux/entre-deu(x)tje*   (but: *cache-nez/cache-neztje,*
crapaud/crapaudtje, frame/framepje, pince-nez/pince-neztje, biscuit-biscuittje). Deletion of silent
letters and doubling**:** *dejeuner/dejeuneetje, diner/dineetje, souper/soupeetje, depot/depootje,*
jabot/jabootje,   matelot/matelootje,   paletot/paletootje,   pierrot/pierrootje,   tricot/tricootje,
pardessus/pardessuutje. *i/ie*-alternation: *souvenir-souveniertje.*

(10)   **Alternation of *ng* and *n***

\g\ → 0 / n _ ( ]$_S$) k   (*hari**nk**je*)

(11)   **Diaeresis placement**   (*ge**ë**ist*)

a. insert a diaeresis in the sequences V_iV_i (except for ii); *ae, ie, oe*; *ai, ei, oi,*
ui; *au, eu, ou* when the letters form part of different syllables

b. insert a diaeresis in the sequences *oee, iee, eee, eei, oei, eui, oui, aai, ooi,*
aie, eie, oie, uie when *ee* or *i* and the preceding vowel can be read
tautosyllabicly when this reading is not intended

EXCEPTIONS:-*eum, -eus, -ei* but: *vacuüm, -cien/*ciën* (*opticien, ancien*)
but: *auspiciën/*auspicien, anciënniteit *ancienniteit*

(12)   **Apostrophe placement**

$$\overset{\text{VV}}{\underset{\vee}{}}$$

a   \s\   → <'s> / V + _   (*ma-ma's*)
b   \s\   → <'> / s + _   (*Louis' boek*)
c   \s\   → <'> / x / + _   (*Marx' theorie*)
d   \s\   → <'> / s... / + _   (*Bush' doctrine*)
e   \tje\   → <'tje> / y + _   (*baby'tje*)
f   0   → <'> / _ +   after letter words and numerals   (*TBS'er, A4'tje*)

CONDITIONS: (12a) is applied to plural and genitive suffix *-s*; (12b-d) are
only applied to genitive suffix *-s*

EXCEPTIONS: Next to the plural forms *eega's*, *la's*, *ra's*, *tra's* and *vla's*,
we also find *eega**as**, la**as**, ra**as**, tra**as*** and *vl**aas**.*

(13)   **Hyphenation**

0   → - / L]_S _ _S[L   (*las*-*ter*)

CONDITION:
Hyphenation may not occur after an initial syllable which consists of one
letter only, or before a final syllable which consists of one letter only

Ordering of the rules

In most cases we derive the correct result by applying autonomous spelling rules
without stipulations about ordering. Some rules are intrinsically ordered, for
instance, syllable sensitive rules can be only applied after Orthographic
Syllabification. At first sight, it seems that Consonant Doubling must be ordered
before Vowel Degemination, in order to prevent doubling from affecting
degeminated consonants: *haamer* → *hamer* → **hammer*. However, the CV-structure
of *hamer* already prevents doubling from applying: Consonant Doubling only affects
consonants after a vowel letter linked to one V-position.

   Another type of rules which may require ordering are Apostrophe Placement and
Vowel Doubling. We could order Apostrophe Placement before Vowel Doubling to
account for a word such as *Venlo's*, but the same result could be achieved by giving
the rules complementary contexts. Similarly, the fact that addition of diacritics,
Vowel Doubling, replacement of *i* by *ie* and Diminutive Allomorphy are blocked by
the presence of hyphens may be accounted for by rule ordering or complementary
contexts.

   Finally, if Consonant Doubling is ordered after Spelling Devoicing, we need not
consider words such as *puzzel* exceptions. Again the same result may be arrived at
by not applying Spelling Devoicing after single vowel letters linked to one V-
position. I conclude that no extrinsic ordering of autonomous spelling rules is
needed.

Appendix F
Single or double consonant letters

Introduction

In this appendix, some complications with respect to orthographic Consonant Doubling are discussed. It is not the doubling rule itself which causes problems, but the complications concern the putative underlying representation of the vowels in question. It is sometimes difficult to establish whether we are dealing with a full vowel or schwa, or with a long or short vowel. The quality of the vowel determines the underlying spelling (structure) on the basis of which doubling is applied, so difficulty in establishing the quality of vowels indirectly has the effect that we cannot determine whether or not to apply doubling. Section F.1 examines contrasts such as *kennissen ~ monniken*, *Lochemer ~ Doetinchemmer* and *Dokkumer ~ Hilversummer*, while section F.2 focuses on complications with the choice of single or geminate consonant letters in non-native words such as *agressie ~ aggregaat*.

F.1 **SINGLE OR DOUBLE CONSONANT LETTERS AFTER LETTERS PRONOUNCED AS SCHWA**

This section will focus on one specific area in which Dutch orthography seems inconsistent, namely contrasts such as *kennissen* versus *monniken*, *Lochemer.* versus *Doetinchemmer* and *Dokkumer* versus *Hilversummer*.

To account for the absence of Consonant Doubling after letters pronounced as schwa, I proposed that schwa is converted to a vowel letter without a V-position, see chapter 2. This proposal leads to the correct spelling for most words, e.g. in words such as [twɛifələ]-*twijfelen*, and [hɑnəsə]-*hannesen*. Words such as [zɪnəɣə]-*zinnige*, [bɪbərəɣə]- *bibberige*, which only differ from the former examples by the fact that the schwa is written as <i> can be accounted for in the same way. In other words pronounced with schwa there is doubling, however, e.g. [kɛnɪsə]-*kennissen*. This suggests that we are not dealing with an underlying schwa in these cases. Indeed, the relevant vowel can also be pronounced as a full vowel, and if it follows a schwa,

only a realization with a full vowel is possible. *Geschiedenissen*, for instance, is pronounced as [ɣəsxidənɪsə], not as *[ɣəsxidənəsə]. This can be explained by avoidance of sequences of schwas. The assumption of an underlying full vowel seems to solve the apparent inconsistency between *bezigen* and *kennissen*. However, there are more endings which exhibit this vacillation in the pronunciation, and the spelling of this group is not consistent. Firstly, the <k> of the ending *-ik* which also vacillates is not doubled:

(1)   a   [kɛnəsə]/[kɛnɪsə]-kenni**ss**en   [ɣəsxidənɪsə]-geschiedeni**ss**en
           [dozəsə]/[dozɪsə]-dosi**ss**en   [metropolɪsə]-metropoli**ss**en
       b   [mɔnəkə]/[mɔnɪkə]-monni**k**en   [vizərɪkə]-viezeri**k**en

The spelling of part of the data in (1) is irregular, whether we postulate underlying full vowels or schwa. A further problem is that the sequences *-em* and *-um*, which show a similar vacillation in pronunciation, also have a variable spelling:

(2)   a   [dɔkəmər]-Dokku**m**er   b   [hɪlvərsʏmər]-Hilversu**mm**er
           [lɔxəmər]-Loche**m**er   [dutɪŋxɛmər]-Doetinche**mm**er

Stated otherwise, pairs with a comparable pronunciation in (1a), (1b) and (2) are written in three different ways: with geminates or single letters in both cases, or with a variable spelling which corresponds to the variation in the pronunciation.

    In this section the facts in (1) and (2) will be examined. It will be argued that the words in (1) have an underlying full vowel, which implies the spelling in (1b) is irregular, but that the spelling variation in (2) does not constitute an inconsistency, since the words in (2a) have an underlying schwa. I will first discuss the phonological analysis of *-ig*, *-ik* and *-nis*; *-em* and *-um*; and *-is*, *-um* and *-us* in loan words, respectively, and subsequently examine the implications for orthography.

Phonological analysis of -ig, -ik and -nis
To account for the vacillation of schwa and full vowels in (1), we can either postulate an underlying schwa and a raising rule, see Trommelen (1983b), or underlying full vowels and a reduction rule, see Zonneveld (1993). I will choose the latter option, since the alternative account by Trommelen uses a rule, schwa raising before velars, which incorrectly predicts that vacillation also occurs in the case of *-lijk* and *-ig*:[1]

---

[1] Trommelen (1983b) postulated an underlying schwa in *-ig* and *-ik* based on two arguments: (i) *-ig* and *-ik*

(3)   bez[ə]g/*bez[ɪ]g   zeurder[ə]g/*zeurder[ɪ]g
       vrol[ə]k/*vrol[ɪ]k   bangel[ə]k/*bangel[ɪ]k

The difference between (1) and (3) warrants different lexical representations. The suffixes *-ig* and *-lijk* have an underlying schwa, whereas *-ik* and *-nis* have a full vowel which can be reduced. Zonneveld (1993:95) reformulates the Dutch Arab-rule, see Kager (1989:294), in such a way that it accounts for the facts in (3). Unstressed post-tonic short vowels are reduced to schwa when the vowel in the immediately preceding syllable is short. The Dutch Arab-rule accounts for the contrast between the final vowels in *havik* (no reduction), *monnik* and *kennis* (reduction possible). This analysis also enables us to account for the difference with *-ig*. Since this ending has an underlying schwa a realization with a full vowel is not possible in *zuinig* or *gezellig,* irrespective of the length of the preceding vowel.

It can be concluded that *-ig* has an underlying schwa, whereas *-ik* and *-nis* have an underlying full vowel. [2] The reduction in words such as *monnik* and *kennis* will be

form a class with *-ing*, and (ii) *-ing* has an underlying schwa. The latter assumption is motivated by the fact that words with the ending *-ing* select the diminutive ending *-tje* instead of *-etje* (*koning*etje*) that normally occurs after short vowels followed by sonorants (cf. *dinget je*). The assumption that *-ig* and *-ik* behave like *-ing* is based on the fact that all three suffixes can be pronounced with a full vowel or schwa, which can be accounted for by a rule that raises schwa before velars (e.g. /konəŋ/ → [konɪŋ]). Trommelen finds support for this rule in the following quotation of Kruisinga: "in most words we write [the weakly stressed or unclear vowel as] *e*, e.g. in *bode*, *vreugde*, *schade* [...]. Before *k*, *g* and *ng* (representation of one sound) we write *i*, as in *havik*, *monnik*, *menig*, *zalig; woning*" [in de meeste woorden schrijven we [de zwakbeklemtoonde of onduidelijke klinker als] *e*, als in *bode*, *vreugde*, *schade* [...] Vóór de *k, g* en *ng* (teken voor één klank) schrijven we *i*, zo in *havik*, *monnik*, *menig*, *zalig; woning; [...],* see Kruisinga (1951:18–19). However, neither argument is conclusive. Firstly, selection of *-tje* does not always warrant the postulation of underlying schwa: there are (non-native) words with full vowels that behave the same way as words ending in *-ing:* they select *-tje* instead of (or as well as) the expected *-etje*, e.g. *professortje*, *pelgrimpje* or *pelgrimmetje*. Secondly, the evidence that Trommelen gave to treat the endings the same way was based on the remark of Kruisinga, but this concerned orthography, not phonology. Orthography can not be used to support a phonological analysis in this way, since the spelling can reflect the pronunciation of earlier stages of the language.

[2] For this analysis it is crucial that we do not follow Van Oostendorp (1995) who extends the full vowel analysis to the ending *-ig*. The first type of evidence he cites is the fact that words ending in schwa normally select the agentive suffixes *-aar*, e.g. in *wandelaar* ([wɑndəlar]), but *-ig* and *-ik* select *-er*: *punniker*, *beziger*. Secondly, there is no linking morpheme [ə] after schwa (*heuvelərug*, *kleuterəschool*), but words ending in *-ik*, *-nis* and *-ig* sometimes have [ə] (spelled *e* or *en*): *monnikəwerk, heiligəleven*. These facts are not convincing, however. Firstly, *-aar* follows /l, n, r/ only: *molenaar*, *twijfelaar*, *piekeraar* (Booij & van Santen 1995:132) versus *bezemer* (Booij 1995:73), so the fact that *-ig* selects *-er* does not necessarily indicate that it contains a full vowel. Secondly, some words ending in a syllable containing schwa have a linking morpheme, e.g. *goederətrein, engeləhaar*. Finally, *-ig* is always realized as schwa. I

accounted for by the Arab-rule, which is assumed to be optional.[3]

Phonological analysis of -em and -um
Now consider *-em* and *-um* which occur alternatively with single and geminate consonants in inflected forms. The contrast between pairs such as *Dokumer* and *Hilversummer* is only inconsistent under the assumption that the same suffix with a full vowel occurs in both words, but we will see that there is no evidence for this assumption. Some examples from Haeseryn et al. (1997) are given below:

(4)   a   Arnhem, Doetinchem, Edegem, Erembodegem, Hattem, Lochem
        b   Blaricum, Brunssum, Bussum, Hilversum

The sequences *-em* and *-um* occur exclusively in geographical names and are derived from *-haim* in most cases.[4] However, I assume that we are dealing with separate endings and not with one ending with two spelling variants. If *Hilversum* is compared to *Doetinchem* the relevant vowels are pronounced differently, as [ʏ] and [ɛ] respectively. This suggests that they are no longer to be considered one suffix. In most dissyllabic words, however, the contrast is less clear. *Lochem* and *Hattem* are pronounced with schwa, and a full vowel seems unacceptable. The same probably holds for *Brunssum* and *Bussum*, but it is hard to distinguish [ə] from [ʏ]. An indication that <u> can denote schwa is the alternation between *Gorinchem* and *Gorkum*, and between *Woudrichem* and *Woerkum*. The trisyllabic spelling is the older variant, which represents a historical pronunciation with a full vowel, see Van der Donk (1938). The dissyllabic variant represented a reduced pronunciation. Interestingly, it is not written with <e> but with <u>. It appears that *-em* and *-um* are no longer to be considered the same ending.

---

conclude that there is no reason to assume that *-ig* has a full vowel.

[3] *Loeris*/*loeres* and *stennis*/*stennes* have a variable spelling in [*Woordenlijst 1995*], contrary to Geerts & Heestermans (1995), who consider *loeres* a non-recognized form and do not list *stennes*. The sequence *-is* in *loeris* is never pronounced with [ɪ] and thus has the same pronunciation as words like *bakkes*, *wiedes*, etc. The ending *-is* never occurs after consonants other than /n/ in native words (with the exception of *kermis*). It seems that the <i> of *loeris* is etymological and not an indication of a full vowel. The word *stennes* or *stennis,* however, can be pronounced with an [ɪ].

[4] According to Gysseling (1960), the following words are derived from compounds with *-haim*: *Arnhem*, *Berchem*, *Blaricum*, *Bornhem*, *Haarlem*, *Hilversum*, *Loppersum*, *Edegem*, *Erembodegem*, *Hemiksem*, *Izegem*, *Maldegem*, *Merksem*, *Oudegem*, *Wevelgem*, *Woudrichem*, *Zelhem*, *Zwevegem*, *Hattem*, *Bakkum*, *Bochum*. However, *Kraainem*, *Brunssum* and *Merksem* are derived from compounds with *-hamma* and *Dokkum* from *Dokkingum*. Thus, words written with <em> are not consistently derived from another morpheme than words written with <um>.

The contrast between the pronunciation of the words under (2a) and (2b) suggests that these endings are to be treated the same way as *-ik* and *-nis*.[5] However, words such as *Lochem* are never pronounced with a full vowel, unlike words such as *monnik* and *kennis*.[6] The Arab-rule is optional, so the absolute absence of full vowels in words such as *Lochem* cannot be explained by this rule. These facts suggest that we are dealing with an underlying schwa. Unless further evidence is found, *-em* and *-um* pronounced as schwa or as a full vowel should not be considered the same suffixes. The pronunciation of the sequences *-em* and *-um* must be established for each word separately. Generally, dissyllabic words have an underlying schwa, and the other words have a full vowel (but cf. *Arnhem* and *Haarlem*). This pattern suggests that a reduction rule such as the Dutch Arab-rule has played a role at one stage, but that its effect has been lexicalized.

Phonological analysis of -is, -um and -us
Finally, consider *-is*, *-us* and *-um* in non-native words. Examples are given under (5):

(5)   *-is*:   dosis, salaris *-um*: valium, museum   *-us*:   catalogus, rebus

In *-is*, schwa alternates with [ɪ]. In the case of *-um* and *-us*, it is not so easy to determine the phonological representation, since [ə] and [ʏ] are hard to distinguish.[7] Trommelen assumes that *-um* has an underlying schwa on the basis of diminutive formation *valiumpje*, but judgements are difficult here, some speakers use *valiummetje*. Moreover, the choice of *valiumpje* does not necessarily imply that the relevant vowel is a schwa. Many speakers also select *-pje* after full vowels, e.g. in *Bethlehempje*, *requiempje* and *Joachimpje*. On the other hand, the stress pattern gives an important indication that we are dealing with full vowels in the endings of (5). Words ending in a syllable of which the vowel is a schwa never have antepenultimate stress (except when schwa is not preceded by a consonant as in

---

[5] Trommelen, who only discusses the ending *-um*, assumes that this sequence has an underlying schwa that is raised when it is preceded by another schwa. According to Trommelen, this is supported by the contrast between the diminutive forms of *Mokum* and *Hilversum*: *Mokumpje* but *Hilversummetje*, cf. *koninkje-wandelingetje* (*Blaricum* is not mentioned, but Trommelen would predict *Blaricumpje*). However, judgements are difficult with such words. Moreover, examples such as *pelgrimpje*, *pythontje* show that selection of a monosyllabic diminutive form does not always mean that there is an underlying schwa.

[6] The sequences *-um*, *-em* and *-ik* are not mentioned as affixes in De Haas &Trommelen (1993), but *-erik* is.

[7] De Coninck (1970) prescribes schwa or a full vowel for non-native words in *-is* (*dosis*, *salaris*), but full vowels only for non-native words in *-um* and *-us*.

terriër, *azië*), cf. *pant<u>o</u>ffel-*pant<u>o</u>ffel*, *ka<u>da</u>ver-*<u>ka</u>daver*. However, words composed with *-is*, *-us* or *-um* can have antepenultimate stress:

(6)   l<u>o</u>gicus, <u>cu</u>mulus, cat<u>a</u>logus, <u>oc</u>topus; <u>u</u>nicum, l<u>au</u>danum, sp<u>e</u>culum, cur<u>ri</u>culum; s<u>y</u>filis, <u>can</u>nabis, met<u>ro</u>polis

This stress pattern suggests that the final vowel is full, just like in words such as *m<u>o</u>nitor* and *p<u>o</u>sitron*.[8] Furthermore, words such as *catalogus* sometimes exhibit stress shift: *cat<u>a</u>logus*. These facts could be explained as a reinterpretation of a full vowel as schwa, which then causes stress shift. Words such as *normaliter* and *archipel* seem to support this analysis. It is either *nor<u>ma</u>liter* pronounced with [ɛ], or *norma<u>li</u>ter* with [ə]. I conclude that *-is*, *-um* and *-us* have full vowels. Other words in which [ɪ] and [ə] are both possible pronunciations such as *goj<u>i</u>m*, *sher<u>i</u>ff*, *top<u>i</u>c*, *goss<u>i</u>p*, *den<u>i</u>m* and *ac<u>i</u>d* will be considered to have underlying full vowels as well, whereas *lim<u>i</u>t*, *stenc<u>i</u>l* and *elix<u>i</u>r/elix<u>e</u>r* have an underlying schwa. [9]

Implications for orthography
Above it was argued that the affixes in (1) and (2) should be divided into two types on the basis of their phonological behaviour: the suffix *-ig* has an underlying schwa, and the suffixes *-ik* and *-nis* have an underlying full vowel. The sequences *-em* and *-um* are not considered affixes which are subject to the Morphological Principle, so words which contain these sequences occur in both categories:

(7)   **schwa**   -ig, Lochem, Dokkum
         full vowel   -ik, -nis, -is, Doetinchem, Hilversum

---

[8] Kager (1989:234) claims that words such as *catalogus* have an underlying schwa, but that the endings *-icum* and *-icus* are not part of the stress domain, which accounts for the antepenultimate stress. However, this would not explain the examples like *met<u>ro</u>polis* or *cat<u>a</u>logus*, nor words where a schwa is part of a suffix, e.g. *elektrode* and *communisme*. Such words always have penultimate stress (*elektr<u>o</u>de*, *commu<u>nis</u>me*), which shows that schwa always forms part of the stress domain, and that the stress attracting properties of schwa are not ignored in these cases. This makes Kager's analysis of *-icum* and *-icus* unlikely.

[9] De Coninck (1970) transcribes the relevant vowels in *stenc<u>i</u>l* and *limited* as schwas and those in *sheriff*, *tonic*, and *topic* as [ɪ] (the other words discussed here are absent in De Coninks dictionary). The alternating spelling of *elixir/elixer* does not reflect variation in the pronunciation: this word is transcribed with a schwa and the pronunciation with an [ɪ] is explicitly condemned. Consequently the spelling with an <i> is etymological, the variant with <e> more in accordance with the pronunciation.

This implies that the spelling <i> in -*ig* and <u> in words such as *Dokkum* is etymological rather than an indication of an underlying full vowel.[10] When the spelling is regular, doubling is expected after full vowels, but not after schwa. When we compare (1) and (7), we see that most spelling contrasts in (1) are caused by pronunciation differences and do not form inconsistencies, and that there is no need to assume that there is doubling after letters representing (underlying) schwa. However, one problem remains in (7), namely the fact that the final consonant of the suffix -*ik* is not doubled before vowels, although it has an underlying full vowel.

Why is there a difference between -*ik* and -*nis*? Te Winkel claims that *vonnis*, *monnik* and *botterik* are pronounced with a 'mute vowel' but *geschiedenis* with a full vowel, and that *vonnissen* is written with <ss> by analogy with *geschiedenissen*, see Te Winkel (1863:53). It is not clear whether there really was a pronunciation difference in Te Winkel's time. Te Winkel assumed that a mute vowel can also be pronounced as [ɪ], see Te Winkel (1860:6) which possibly used to be a reduced vowel in earlier versions of Dutch (hence Kollewijn's proposal for the spelling -*lik* for -*lijk*). Thus, it is only the stress pattern that distinguishes *botteriken* from *geschiedenissen*, and this is a very subtle difference. Since the spelling difference between *monniken* and *vonnissen* is quite old, see Siegenbeek (1805b), it is also possible that there was no difference in the pronunciation of *botteriken* and *geschiedenissen*, but that the spelling itself has had some influence. In any case, there is no difference in present-day pronunciation: both *geschiedenissen* and *viezeriken* are realized with a full vowel, so the spelling ought to have geminates in both cases. The spelling of -*ik* is really inconsistent with the current pronunciation.

An overview of single or double consonants is given under (8):

(8)   **Regular**
    schwa, no gemination: -(er)ig (bezi**g**en), Loche**m**er, Dokku**m**er
    full vowel, gemination: Hilversu**mm**er, Arnhe**mm**er; -(e)nis (kenni**ss**en);
                           -is (dosi**ss**en), -um, -us (kroku**ss**en)
    Irregular
    full vowel, no gemination: - (er)ik (monni**k**en, viezeri**k**en)

We can conclude that the spelling in (1) is regular, except for the single <k> in words such as *monniken* and *viezeriken*, and that the contrast between *Loche**m**er*

---

[10] The generalization is that [ʏm] is written as <um> after [k] (*Bakkum*, *Dokkum*, *Blaricum*, *Gorkum*, *Mokum*, *Woerkum*) and after [s] unless this sound follows [k] (*Bussum*, *Brunssum*, *Hilversum*, *Loppersum* but *Merksem*, *Hemiksem*), and as <em> elsewhere. This generalization does not extend to non-geographical [əm], which is written as <em>: *stiekem*, *bloesem* etc.

versus *Doetinche**mm**er* and *Dokku**m**er* versus *Hilversu**mm**er* is unproblematic, since there is no reason to assume that the contrasting words contain the same suffix *-em* or *-um*.

Chapter 4 gave rules for the spelling of long and short vowels in non-native words. Deriving the spelling now seems straightforward. However, there are two factors which complicate the choice of single and double consonants: the indistinctness of the length contrast and the effect of non-native prefixes. Both factors will be discussed here.

Indistinctness of the length contrast
In native words it is always clear if a given vowel is long or short (although there are a few words in which the pronunciation varies: *witlo(o)f*, *kivi(e)t*). In non-native words the length contrast also exists, see for instance minimal pairs such as [hɑlo ~ hɑlo] (*halo* id. versus *hallo* 'hello'), [koma ~ kɔma] (*coma* id. versus *komma* 'comma'), [oto ~ ɔto] (*auto* 'car' versus *Otto* name). However, the contrast has less functional load, and it almost exclusively occurs in stressed syllables. Some rare examples in which spelling suggests a length contrast in an unstressed syllable are [d**o**serə] (*do**c**eren* 'to teach') versus [d**ɔ**serə] (*do**ss**eren* 'to slope'), [tr**a**serə] (*tra**c**eren* 'to trace') versus [tr**ɑ**serə] (*tra**ss**eren* 'to draw') and [**a**nalɪst] (*an**a**list* 'analyst') versus [**ɑ**nalɪst] (*a**nn**alist* 'historian').[11]

In words such as the following the contrast between long and short vowels which is suggested by the spelling is often not realized:

---

[11] In some words the spelling varies as well: *staniool* or *stanniol*, *allit(t)ereren*, *lit(t)eratuur*.

(9) a **ɑ**ɣrɛsi *agressie*   ~   **ɑ**ɣrəɣat *aggregaat*

　　　dɪlətɑnt *dilettant*   ~   ɪlyzi *illusie*

　　　kɔmite *comité*   ~   kɔmɪsi *commissie*

　　　fɔrmɑtØr *formateur*   ~   fɔrm**ɑ**ter *formatteer*

　　b **ɑ**nanɑs *ananas*   ~   k**ɑ**nabɪs *cannabis*

　　　pɑlɪʏm *pallium*   ~   m**ɑ**ratɔn *maraton*

　　c kɔrəlat *correlaat*   ~   satəlit *satelliet*

　　　ɑtlətik *atletiek*   ~   dɪltɑnt *dilettant*

　　d ɪmitatsi *imitatie*   ~   i**mi**ɣratsɪ *immigratie*

　　　sʏprəmatsi *suprematie*   ~   sʏprimer *supprimeer*

Despite the contrast in the spelling of the pairs of words in (9) which suggests a length contrast, all words are often realized the same way, i.e. with a short vowel in (9ab), with a schwa in (9c) and with a long vowel in (9d).

In the literature it is assumed that there is an underlying contrast in such words, which is neutralized by sound rules. The facts in (9ab) are described by an optional 'shortening' rule which is applied in pretonic unstressed syllables, see Cohen et al. (1978:49), Kager & Zonneveld (1989). Shortening of [a] is sometimes even applied in stressed position. This means that the underlying contrast is not self-evident in these positions either, as illustrated by the last two examples of (9a). The neutralization in (9c) is ascribed to vowel reduction, see Kager (1989:315). There is no rule in the literature which accounts for the 'lengthening' in (9d), but the alengthening rule that affects pretonic [i] and [y] before a dental cluster, see Trommelen & Zonneveld (1991), Zonneveld (1993), could be extended so that it accounts for the facts in (9d).

However, what is important here is not the exact formulation of rules, but to find out if the assumption of an underlying length contrast and neutralization rules is valid. If neutralization is the effect of productive rules, one would expect that it would not be difficult to determine the length of a given vowel. However, in practice this is not always so easy. For instance, the abolishment of the contrast between <n> and <nn> in certain contexts (see Appendix H) suggests that there was no sound contrast which enabled writers to choose the correct spelling. Where speakers do no longer recognize the underlying length contrasts, they can not derive the spelling from the pronunciation and the difference between single letters and geminates becomes etymological.[12]

---

[12] In the CELEX-database the effect of vowel shortening is sometimes encoded in underlying representations, but not consistently so, cf. /kɑpitel ~ kɑpital/ (*kapiteel-kapitaal*), /pɑpavər ~ pɑpaja/,

Let us therefore consider another option, namely that there is no underlying length contrast in unstressed syllables. In that case, the contrast between single letters and geminates has no basis in the pronunciation, so we predict that it causes difficulties in writing, since the choice between single letters and geminates is arbitrary. However, a generalization is missed, explaining why in most cases short vowels in unstressed syllables correspond to vowel letters followed by single consonant letters. Different factors cooperate to achieve this pattern. Short vowels, being in closed syllables, attract stress (see for instance Kager (1989:227), Trommelen & Zonneveld (1989:67). On the other hand, neutralization rules such as shortening are almost exclusively applied to unstressed syllables (the exceptions being words of the type *ananas*). Finally, unstressed short vowels are lengthened after the addition of non-native affixes: *consul-consulaat*, see Booij (1995:80–83). Together, these factors lead to a correlation between stressed syllables and short vowels, and unstressed syllables and long vowels. The choice between single and double letters could then be accounted for by the following rule of thumb: write geminates after stressed short vowels or after short vowels which were stressed in the base word. This generalization accounts for differences between *debat-deba**tt**eer* and *alcohol-alcoho**l**isch*. In fact, the rule for doubling is formulated this way in the South African spelling dictionary (see [*Afrikaanse woordelys*], p. 27).

To be able to choose between the two approaches, we must look for cases where they make different predictions. Words of which the spelling suggests that they are exceptions to this generalization are given under (10):

(10)   a   **doubling in unstressed syllable**
         te**rr**as   cal**l**igrafie   Bet**t**ina
         po**mm**ade   fla**tt**eren   ba**rr**icade
         secreta**ris**sen   cur**sus**sen   slalo**mm**en
      b   **doubling despite unstressed short vowel in base word**
         focus-focu**ss**eer   format-forma**tt**eer   fiat-fia**tt**eer
         bivak-biva**kk**eer   Boston-bosto**nn**eer   sabbat-sabba**tt**ist
      c   **no doubling in stressed syllable**
         ananas   **mar**aton
      d   **no doubling despite stressed short vowel in base word**
         pis**ton**-pisto**n**ist   pe**ron**-pero**n**ist   charla**tan**-charlata**n**erie
         Ja**pan**-Japa**n**oloog   ro**man**-roma**n**esk   maca**dam**-macada**m**iseer
         spi**on**-spio**n**eer etc.

_____

kapysein ~ kɑpitɛin/ (*capucijn-kapitein*).

The facts in (10) show that there is not a perfect correspondence between stress and doubling (the exceptional spelling of pairs such as *spion-spioneer* is the result of a spelling reform which is discussed in appendix H). It appears that consonants are also doubled after unstressed short vowels, provided that they are not affected by lengthening, so a Dutch counterpart to the South African rule would not always lead to the desired result (in fact the same holds for South African, cf. for instance *sekretarisse*). At first sight these facts argue in favour of an underlying length contrast and neutralization rules. However, it is also possible that judgements on the length contrasts in (10) are based on the spelling of these words. To decide in this matter, we need experiments which separate judgements on spelling and pronunciation. However, the design of such experiments falls outside the scope of the present study. For the present time, I assume that the facts in (10) argue in favour of the first option, i.e. that loan words have a length contrast in unstressed syllables.[13] I therefore decided to transcribe all these words with /a/ (conversely, *a* followed by geminates is always transcribed as /ɑ/ except for *catarraal*, *diarree*, *paraffine*, *parallax* and *parallel*).[14] I recognize that this approach has the disadvantage that in some cases spelling pronunciations will be listed, and that the real spelling problem is not addressed: pairs such as *ananas-cannabis* will automatically receive the correct spelling, since the sound representation is spelling-based. However, since the problem lies in the fact that sound distinctions to be represented by spelling are unclear, not in the spelling rules themselves, it is not of central interest to the present investigation.

The role of non-native prefixes

The second complication of the choice between single letters and geminates in non-native words is the presence of non-native prefixes. Some consonants in non-native words do not undergo gemination because of the morphological structure of the words, see (10a). However, we also find geminates at prefix boundaries if the last consonant of the prefix happens to be identical to that of the first consonant of the stem, see the examples in (11b-c):[15]

---

[13] Length contrasts such as the following could be the effect of letter-to-sound conversion rules, since a single vowel letter in a closed syllable is read as a short vowel: *laboratorium-lab*, *bioscoop-bios*, *coöperatief-coöp*, *professor-prof*, *locomotief-loc*, *populair-pop*, *Eduard-Ed*, *Thomas-Tom*, *Jacobus-Jacob*.

[14] De Coninck (1970) transcribes words such as *papaver*, *ananas*, *formateer*, *catarraal*, *diarree*, *paraffine*, *parallax* and *parallel* with [ɑ] only.

[15] Prefixes ending in a long vowel, such as *a-*, *de-*, *di-*, *(p)re-*, and nearly all dissyllabic prefixes never occur before double consonants: **prommenade* and **proffessor*. However, it is not always easy to

(11)   a   a**b**ortus, a**d**equaat, di**s**agio, i**n**actief, su**b**altern
       b   a**bb**atiaal, a**dd**uceren, de**ss**ous, di**ss**onant, i**nn**ocent, su**bb**oreaal
       c   a**bb**revieer, a**gg**regaat, a**tt**ractie, e**cc**lesia, e**ff**luent, di**ff**ractie, su**pp**lement,
           o**pp**ressie.

The words in (11b) behave like regular monomorphemic words with short vowels,
but the geminates in (11c) with geminates before liquids (where Consonant Doubling
is not applicable) can only be explained by their morphological structure. If the
words in (11a) had one underlying consonant and those in (11b) two, the spelling
would be straightforward. However, postulating two underlying identical consonants
is only motivated by the morphological structure of words such as *adduceren*, since
Dutch words do not have sequences of identical consonants within morphemes. Non-
native prefixes sometimes still form separate domains for phonological rules such as
syllabification, for instance in **in**actief and **des**infecteren. In these cases they are
easily recognized. Some prefixes even behave as parts of compounds: they have
stress and are combined with native words, see Booij (1981), for instance *subgroep*,
metataal. In these cases it is obvious that the prefixes are treated as separate
morphemes for spelling rules as well. The single consonant in such words is
comparable to that in compounds such as *wanorde*. However, in most words in
(11a), e.g. *anarchist* and *abortus*, the prefix is not a domain for phonological rules
like syllabification, cf. [ɑ-nɑr-xɪst]. Here it is not probable that speakers still
recognize non-native prefixes as separate morphemes. In the first place, the prefixes
are not productive: **anorganisatie*, **disironisch*. Secondly, the pronunciation of
some words suggests that prefixes have been reanalysed as part of an underived
word:

(12)   a   anɛkdotə (*anekdote*), sinonim (*synoniem*), syblim (*subliem*), redɑksi
           (*redactie*)
       b   prɛstatsi (*prestatie*), rɛspɛkt (*respect*)

The words in (12a) have a prefix that ends in a consonant, but the preceding vowel is
long, whereas the vowel final prefixes in (12b) are pronounced with short vowels.
The difference in pronunciation of such reanalysed words (*anekdote*) and words such
as *anarchist* has no effect on the resulting spelling (but it has repercussions for

---

distinguish prefixes with a short vowel from prefixes with a long vowel: cf. *asymmetrisch* versus
assimilatie.

hyphenation), although the derivation differs.[16] In words such as *anarchist* gemination is blocked by the prefix boundary, in words such as *anekdote* gemination is simply not applicable. In some cases the reanalysis has even resulted in the disappearance of geminates in the spelling, e.g. in *agressie* and most words with the prefix *trans-* (*trans+scendent* → *transcendent* etc. but *transsubstantiatie*).[17] We can conclude that Romance prefixes are not always considered separate morphemes. As underlying geminates are only motivated by the morphological structure in words such as *adduceren*, I assume that words such as *adduceren* now contain one underlying consonant only, just like *adequaat*.

When we list the different types of spelling for intervocalic consonants we see that even if the vowel is short, there can be no doubling, and even when the vowel is long there can be geminates. This means there are four different situations:

---

[16] Non-native prefixes also complicate hyphenation. [*Woordenlijst 1954*], p. LIII-LV gives the following instructions ([*Woordenlijst 1995*] is less explicit): hyphens can be inserted after common prefixes such as *ab-*, *ad-*, *in-* and *im-*, *sub-* and *sup-*, *trans-*, *anti-*, *con-*, *contra-*, *dis-*, *neo-*, *ortho-*, *peri-*, *kata-*, *meta-*, *mono-*, *para-*, but not after dissyllabic prefixes that have lost a vowel (*anekdote/ana+ekdote*, *para-allele/par+allel*), and according to the instructions for underived loan words elsewhere. Investigation of the examples suggests that 'morphological' hyphenation is blocked when this would incorrectly suggest a long vowel instead of a short one or vice versa: *pres-tatie/*pre-statie* ([prɛstatsi]), *res-pect/*re-spect* ([rɛspɛkt]), **syn-onim/syn-oniem* ([sinonim]), **red-actie/re-dactie* ([redɑksi]), but it is allowed where the pronunciation is unclear: *ab-erratie* ([abɛratsi]/[ɑbəratsi]) *bio-scoop* ([biɔskop]-[bioskop]).

[17] This is also observed in Neijt & Zuidema (1994b:52). Spelling adaptation has led to differences with other languages, cf. English *aggression, bisector* ~ Dutch *agressie, bissectrice*, French *abréviation* ~ Dutch *abbreviatie*, South African *stres, cappucino* ~ Dutch *stress, cappuccino* and German *Kaffee* ~ Dutch *café*.

(13)   A   B   C   D
         short vowel,   **short vowel,**   **long vowel,**   **long vowel,**
         geminate   **single letter**   **single letter**   **geminate**
         annalen   an+organisch   pro+motie   para+llel
         com+missie   a+symmetrisch   a+sociaal   dia+rree
         ab+brevieer   ab+rupt   pro+bleem

Even when underlying long-short distinctions are clear, the (underlying) pronunciation only allows us to choose between type A/B on the one hand, and type C/D on the other hand, but it does not distinguish between type A and B or between type C and D. The pronunciation thus does not suffice to choose between single and double consonant letters, and the morphological structure that enabled us to choose is no longer transparent. Therefore, one of the variants should be considered exceptional. I assume that the default spelling of consonants after short vowels is with geminates, just as in the native part of the lexicon. Words of type B and D (*anorganisch*, *parallel*) and geminates before liquids (*abbrevieer*) are exceptional.

   The difficulties with the choice between single and double letters in non-native words can be accounted for by the presence of near-automatical neutralization rules which only affect non-native words and by the effect of non-native prefixes.

Appendix G
Uniform spelling of morphemes as the effect of spelling conventions

This appendix discusses two sets of words that illustrate that the constant spelling of morphemes does not always imply that it is based on an abstract sound representation. In 4.5 it was concluded that words composed with non-native affixes are treated as morphemes. However, there are pairs of related non-native words with a uniform spelling that seem to indicate that this conclusion is not warranted. These examples, which were illustrated in chapter 4, are repeated here:

(1)   **Uniform spelling of related non-native words**
  a   [**s**olytsi]-[re**z**olytsi]   **s**olutie-re**s**olutie
  b   [provo**s**erə]-[provo**k**atsi]   provo**c**eer-provo**c**atie
  c   [ɑk**s**i]-[ɑk**t**if]   ac**t**ie-ac**t**ief

These words suggest that spelling encodes the pronunciation of non-native morphemes rather than the derivations as a whole, since the spelling seems to abstract from sound rules that affect the complex words, and represents the pronunciation of the constituting morphemes.

However, closer examination will show that this is not the case. This is already suggested by the fact that the rules responsible for the sound alternations in (1) are early, morpholexical rules whose effect is normally visible in Dutch orthography. Indeed spelling does not abstract from comparable sound rules, such as Learned Vowel Backing (*religieus-religiositeit, elitair-elitarisme*). It will therefore be argued that the uniform spelling in (1) has another explanation, namely the fact that the letters <s>, <c> and <t> can encode different sounds: /s/ or /z/, /s/ or /k/ and /t/ or /s/, respectively. In other words, the seemingly 'abstract' spelling is nothing more than the result of a spelling convention.

Alternation of [z], [s] and [k]
More examples of the type *solutie-resolutie* are given under (2):

(2) a pulsen   pulseer   impulsen   impulsief
       perverse   perversie   intens   intensief
    b serveren   reserveren   signeren   resigneren

First consider an analysis that posits an underlying /s/ and a voicing rule, see for
instance Kooij (1983). Wester (1987) even proposed that intervocalic voiced velar
and dental fricatives are derived from underlying voiceless fricatives by a voicing
rule, not only in non-native words but in all words. We saw in chapter 3 that this
analysis turned out to be untenable for native words such as *lief-lieve* and *vrees-
vrezen*, but it is possible that this analysis does hold for s/z-alternations in non-native
words. It would explain why [z] is often written as <s> in cases such as (1). The
following rule is now suggested:

(3)   /s/ → [z] after a long vowel or sonorant consonant

Support for a s-voicing rule is formed by acronyms such as *NASA* (*National
Aeronautics and Space Administration*). In these acronyms an <s> that is
pronounced as an [s] in the original words becomes voiced between vowels.
However, this analysis also has some drawbacks.
    A complication of this analysis that concerns phonology is the fact that there are
non-native words with voiceless fricatives after long vowels or sonorant consonants.
Some examples are given in (4a-b). There even are pairs of related words similar to
those in (2) in which there is no voicing, as illustrated in (4c):

(4) a [pɛrsel]   perceel   [ɑmbylɑnsə]   ambulance
       [provɪnsi]   provincie   [dɔserə]   doceren
    b [karusel]   carrousel   [reɣisØr]   regisseur
       [muserə]   mousseren   [pɛrson]   persoon
    c [balɑnzə]   balansen   [balɑnserə]   balanceren
       [kadɑnzə]   cadansen   [kadɑnserə]   cadanceren
       [səkwɛn(t)si]   sequentie   [kɔnsəkwɛnt]   consequent
       [sɪɣnifikɑnt]   significant   [ɪnsɪɣnifikɑnt] insignificant
       [sʏpozitsi]   suppositie   [presʏpozitsi] presuppositie

Wester (1987:79) observed that facts such as those in (4a) are written with <c>:
"Notice that 'c' behaves as 'a kind of s' that does not undergo voicing'". Wester
(1987:76) therefore claimed that "*c* must be considered a plosive as far as
orthography is concerned, and is converted to *s* or *k* after the application of the rule

that voices fricatives". This remark concerns the reading process, but it does not explain the phonological behaviour of the words in (4a).

In line with Wester's analysis, we might postulate that [s] is derived from the underlying plosive /k/ in these cases, which would account for most exceptions in (4a). This brings us to the examples of (1b). To account for surface fricatives in words such as *provoceren* and *doceren*, we need a rule that converts /k/ to [s] before front vowels. The (extrinsic) ordering of this rule with respect to fricative voicing would then explain the difference in voicedness between *doceren* and *doseren*:

(5) a   /s/ → [z] after a long vowel or sonorant consonant   (/doserən/ → [dozerə])
    b   /k/ → [s] before a front vowel   (/dokerən/ → [doserə])
        ordering: a aplpies before b

However, rule (5b) also has exceptions and it presupposes unlikely productive word formation rules and morpholexical rules, and it cannot account for the facts in (4b).

The examples in (6) show that not all relevant words undergo rule (5b):

(6)   **pronunciation**   **spelling**
      [mɑrkɑnt] ~ [mɑrkerə]/*[mɑrserə]   markant-mar**k**eren
      [fraŋko] ~ [fraŋkerə]/*[franserə]   franco-fran**k**eren
      [vɪskozitɛit] ~ [vɪskØs]/*[vɪsØs]   viscositeit-vis**k**eus
      [hipotek] ~ [hipotekɛːr]/*[hipotesɛːr]   hypothee**k**-hypothe**c**air

Romance words such as *provocatie-provoceer* can not be seen as synchronically derived from a common base in the Dutch language, since we cannot predict the pronunciation of 'derived' forms, see Booij & Van Santen (1995:45). Even if the formal relation is clear, the semantic relation may be obscure, cf. for instance *radicaal-eradiceren.* Rule (5b) thus is at the most a historical rule just like other non-native rules whose effect is visible in orthography such as Learned Vowel Backing. Speakers of Dutch still use such rules as a 'via-rule', but not as a productive rule, see Booij (1995:79). Consequently, we cannot posit an underlying /k/ for [s], so the fact that voicing does not apply in words such as *provoceren* remains problematic.

The second drawback of Wester's analysis of pairs such as *solutie-resolutie* is formed by incorrect predictions of the spelling of some words. If Wester's analysis were correct, we would predict that words which Wester claims to have an underlying /s/ are always written with <s>. This prediction comes out for the words such as *meson*. However, there are similar words that are written with a <z>:

(7)   meson ~ horizon   censuur ~ azuur
         diffuse ~ precieze   forensen ~ accijnzen

In some words in (7), <z> is the result of a spelling change in the Dutch language. However, in Wester's analysis /s/ and /z/ are separate sounds that are written differently, and there is no reason why these two spellings should converge. To account for facts such as *azuur* we would need an additional writing convention that changes an /s/ to a <z>. Alternatively, we could assume that the underlying representation has changed. We would also need a complementary writing convention that changes a /z/ to an <s> in words such *forensen*. There does not seem to be much regularity: both /s/ and /z/ may be written as <s> and <z>, and not all voiceless sibilants that form exceptions to voicing are written as <c>, as shown in (4).

Summarizing, the facts in (4) and (7) show that there is no one-to-one relation between /s/ and <s> or /z/ and <z>. Wester introduced abstract /k/ and /s/ to account for sound-spelling relations, but this abstract segment causes more problems than it can solve. It turns out that in Wester's analysis sound-to-spelling correspondences are quite complicated after all. For these reasons, I will not adopt Wester's analysis.

Let us therefore consider an alternative approach. In this analysis words such as *doseren* and *forensen* have an underlying /z/. The derivation of the surface pronunciation (e.g. [forɛnzdə]) and [forɛnzis]) is now straightforward: nothing happens. We do not need an abstract /k/ to distinguish *doceren* from *doseren*; /s/ and /z/ suffice to derive the surface pronunciation variants. This analysis removes the phonological problems of Wester's analysis. However, now we must account for the fact that we write an <s> instead of a <z> in *doseren*, *forensen*, *diffuse* etc. The explanation is easily found. We are dealing with an etymological spelling here, a relict of Latin spelling that did not use a <z> (except in Greek loans such as *zefier*). However, in some cases we write a <z>. This is not problematic, since an etymological spelling may be regularized, and the regular spelling of /z/ in native words is <z>, e.g. *zaal*, *deze* (in chapter 4 we saw that we can predict spelling adaptation in some contexts).

If words such as *doseren* have an underlying /z/, and the spelling <s> is the effect of an etymological writing convention, the correspondence between sounds and spelling is more transparent than it is in Wester's analysis: the etymological spelling <s> is being adapted by replacement with regular <z>, starting at the end of morphemes. A rule such as 'write /z/ as <s>' is no longer necessary. <c> is considered an etymological spelling. The fact that <c> may represent both /k/ and /s/

is due to a historical rule that changed /k/ to [s] before front vowels. Spelling abstracts from this rule even though the effect is lexicalized. The rules for the spelling of /s/ and /z/ after long vowels or sonorant consonants in non-native words are given in (8) and (9), respectively:

(8) Write /z/   as <z> at the end of morphemes   (precie**z**e)
                     as <s> elsewhere   (preci**s**ie)

(9) Write /s/   as <c> before front vowels   (do**c**ent)
                     as <s> elsewhere   (per**s**oon)

The new view on sound-spelling relations also offers an alternative, non-phonological explanation for facts such as *NASA*. Wester assumed that the surface [z] is caused by the voicing rule. However, words like *NASA* originate as follows: the initial graphemes of the (most important) words become phonemes, and these phonemes form a new word, see Reker & Streekstra (1988). Therefore, in order to pronounce an acronym, we have to apply letter-to-sound conversion rules. The rule that is relevant here is the complement of rule (9b):

(10)   Read <s> as /z/ between vowel letters (and as /s/ elsewhere)

It is rule (11) that accounts for the pronunciation with a [z] in words like *NASA*. This rule may also be responsible for the 'voicing' in words of which the pronunciation is not known. For instance, loans such as *Alfonso* and *pisang* may be pronounced with a [z] instead of the original [s] for this reason. It is also the spelling which induces the pronunciation [ses] when the name *Cornelis* ([**k**ɔrnelɪs]) is abbreviated as *Cees* rather than *Kees*: <c> before front vowels is read as [s].

It appears that the 'spelling convention' analysis allows for a simpler phonological analysis of the facts. There is no need for abstract representations with /s/ or /k/ which never surface, or for different underlying representations for words that have a similar pronunciation such as *precisie-precieze* and *censuur-azuur*. Moreover, this analysis offers more insight into the spelling system.

Alternation of [(t)s] and [t]
Another set of words that have a spelling that is more constant than their pronunciation is given in (11):

(11)   [ɑ**ksi**]-[ɑk**tif**]   ac**t**ie-ac**t**ief
            [dekoratsi]-[dekoratif]   decora**t**ie-decora**t**ief
            [vakɑn**(t)si**]-[vakɑn**t**]   vacan**t**ie-vakan**t**

In *actie* and *decoratie*, <t> corresponds to [(t)s], but in the related words *actief* and
decoratief <t> corresponds to [t]. Again we need to determine whether there is a
common underlying form for the related words reflected in orthography.

   At first sight, it seems that these alternations can be accounted for by the
postulation of an underlying /t/. In that case, the derivation of the spelling of words
such as *actie* is straightforward (/t/ → <t>), but the pronunciation [(t)s] must be
accounted for by a morpholexical rule that applies before certain (semi-)suffixes
only: *-ie* (*actie*), *-iaal* (*differentiaal*, *potentialis*), *-ieel* (*confidentieel*), *-iaat*
(*novitiaat*, *initiatie*, *initiator*, *nuntiatuur*), *-io-* (*ratio*, *station*, *functioneel*), and
similar suffixes (e.g. *silentium*, *nuntius*, *dementia*, *Tertiair*, *penitentiaris*, *patiënt*,
spatiëren, *tendentieus*, *gentiaan*). However, there are exceptions in these contexts:

(12)   a   *Cynthia*, *lithium*, *parathion*, *promethium*
        b   *peripetie*, *parodontium*, *syncytium*

In (12a) [t] is written as <th>, so we could postulate that <th> represents a sound that
does not undergo spirantization. However, we already discussed the problems of
abstract sounds that only differ from other sounds by not undergoing a certain rule.
Moreover, this does not always work out well: <th> represents [ts] in *forsythia*.
Many speakers pronounce *lithium* and *promethium* this way, and when speakers are
asked to combine *apathisch* with *-ië they* pronounce the new non-existent word as
[ɑpatsiə]. On the other hand, (12b) illustrates that there also are words in which <t>
(without <h>) does not correspond to [(t)s]. Therefore, there is no reason to assume
that <th> and <t> encode different sounds and the words in (12) remain problematic,
unless we are not dealing with a productive rule.[1] For this reason, Booij (1995:79)
concludes that alternations such as those in (11) are to be accounted for by a 'via

---

[1] Another clue that we are not dealing with a productive rule is that the context of the rule becomes much
simpler when it is considered a historical rule. In Brink (1970), a SPE-like analysis of Dutch phonology,
words like *functie* and *functioneer* are derived from an underlying representation /fʏŋktɪon/ by rules such as
N-deletion, Vowel Reduction and t/s-alternation, see Brink (1970:197–199). In modern versions of
phonology such abstract analyses are no longer postulated, but we may consider this derivation a historical
sound change. In Booij's formalization the rule is applied before [i] and before [i] followed by a vowel initial
suffix, but diachronically both *functie* and *functioneer* are words in which [i] is followed by a vowel-initial
suffix. Some of the exceptions to the rule are explained as well: *sympathie* is not derived from *sympathio(n)*,
so the rule did not apply.

rule' instead of a productive rule.

A second drawback of the abstract analysis concerns orthography. There are pairs of words that exhibit *t-(t)s* variation, while one of them is not written with <t>, but with <x>, <s> or <ts>, which corresponds to the surface pronunciation:

(13)   a   anorexia-anorectisch   coaxiaal-coactisch
             fluctueer-fluxie   dyslexie-dyslectisch, etc.
         b   epilepsie-epileptisch   ellips-elliptisch, etc.
         c   intelligent-intelligentsia   staat-staatsie

When we adopt the analysis proposed above, we should assume an underlying /t/, but the spelling does not correspond to this putative underlying sound. Therefore, we need rules that convert /kt/ into <x> (*fluxie*), and /t/ to <ts> (*staatsie*) or <s> (*epilepsie*). Thus, the analysis with a underlying /t/ is not ideal.

Now consider the alternative analysis. If we assume that the surface pronunciation is the same as the underlying representation, i.e. /s/ after obstruents and /t s/ elsewhere, we no longer need a productive sound rule. As suggested by Booij, there may be a via-rule, which would explain the pronunciation of *apathië* as [ɑpɑtsiə]. The derivation of the surface pronunciation thus only involves the deletion of /t/ after sonorant consonants (and after vowels as well for some speakers). Spelling abstracts from the historical rule, so we need the following writing conventions to derive the spelling of the words in question:

(14)   **The spelling of /s/**
         a   /ks/ → ct   before *-ie*, *-ioneer*, etc.   (*fractie*)
         b   /ks/ → x   elsewhere   (*existent*)
         c   /ts/ → t   before *-ie*, *-ioneer*, etc.   (*situatie, preventie*)
         d   /ts/ → ts   elsewhere   (*fatsoen, kwetsuur*)
         e   /s/ → t / p _   before *-ie*, *-ioneer*, et c.   (*conceptie*)
         f   /s/ → c / son _   [−back]   (*sociaal*)
         g   /s/ → s   elsewhere   (*persoon, pressie*)

Again the 'spelling convention' analysis does not presuppose abstract representations with /t/ that never surfaces, or different underlying representations for words that have a similar pronunciation such as *perfectie-flexie*. Moreover, it offers more insight into the spelling system. The words in (14c) can be considered adapted to the Dutch pattern.

We can conclude that the constant spelling in pairs such as *solutie-resolutie*,

provoceer-provocatie and *actie-actief* is best accounted for by a writing convention. Consequently, the uniform spelling of these words does not contradict the conclusion of 4.5 that non-native morphemes do not form domains for phoneme-to-grapheme conversion rules.

Appendix H
Overview of spelling changes since 1804

This appendix gives an overview of the changes that Dutch spelling underwent since its first formalization in 1804. For an overview of decision-making with respect to spelling reforms see Geerts et al. (1977), Geerts (1994) and Neijt & Nunn (1997). The overview presented here consists of the rules of Siegenbeek and Te Winkel, the spelling reform of 1936–47, and the spelling reform of 1995, respectively. This overview focuses on the spelling of words and does not pay much attention to topics such as the placement of spaces or capitals.

Siegenbeek (1805a) and Te Winkel (1863)
By 1800, Dutch spelling was becoming uniform, but there still was variation in the following cases: the spelling of long vowels (e.g. /a/ as was written as *aa*, *a* or *ae*; /i/ as *ie*, *i*, *ij* or *y*), the spelling of diphthongs (*ei*, *ij* or *y*; *aau*, *au* or *ou*) and semi-diphtongs (*aai*, *aay*, *aaij* or *aj*) and the spelling of obstruent clusters: *magt* or *macht*, *kachel* or *kagchel*. The regulations of Siegenbeek, published in 1804, decided on most of these issues. Siegenbeek chose to write *zaai-zaaijen*, *magt* and *kagchel* and vowel geminates for long vowels. The other types of spelling variation remained. In addition to the spelling regulations, Siegenbeek published a spelling dictionary in 1805 to regulate the spelling of individual words, particularly with respect to the issues mentioned above.

The choices made by Siegenbeek were not acceptable to everyone in the Netherlands, and a slightly different set of regulations was adopted in Belgium. Therefore, initiators of a new dictionary of Dutch tried to introduce writing conventions that were acceptable in Belgium as well as in the Netherlands. The compilation of these conventions was entrusted to Te Winkel, who had written an explanation of Siegenbeek's spelling rules. Te Winkel's rules were published in 1863 as 'Grondbeginselen der Nederlandsche spelling' (principles of Dutch orthography). Based on the principles that Te Winkel designed, Te Winkel and De Vries published [*Woordenlijst 1866*]. This spelling has therefore since then been known as the 'De Vries and Te Winkel spelling'. It was officially accepted in Belgium in 1864 and in the Netherlands in 1883.

The spelling rules as formulated by Te Winkel are almost the same as those of Siegenbeek. A few controversial conventions were adapted, and the spelling of some words was changed, e.g. *muziek* instead of *muzijk*. The following general changes took place (in Te Winkel (1863) we still find *blaauw*, *vleyen*

and *gooy-gooyen*, but this was changed in 1884):

(1) **Sounds**   **Siegenbeek's spelling**   **Te Winkel's spelling**
  a [ɛi]   m**y**, vl**ey**en   m**ij**, vl**ei**en
     [ɑu]   bl**aauw**   bl**auw**
     [j]   goo**i**-goo**ij**en   goo**i**-goo**i**en
     [x]   ma**g**t, ka**gch**el   ma**ch**t, ka**ch**el
  b [e]   -**e**ren, -**e**le   -**ee**ren, -**ee**le

Like his predecessor Siegenbeek, Te Winkel did not want to reform etymological spellings, since the original spelling is in accordance with the etymological principle and gives a correct representation of the pronunciation (sometimes through the pronunciation rules of the foreign language). A Dutch spelling would be awkward, for instance in the case of *masjine* (*machine*) or *patroelje* (*patrouille*). Furthermore, since the etymological spelling is already given, no writing rules have to be found. Spelling reform would presuppose the formulation of new spelling rules which would either be consistent but very drastic (and thus violate the Principle of Common Practice), or not drastic and thus inconsistent. Finally, Te Winkel (1863:67–69) argued that even with an adapted spelling a word will be recognizable as non-native, so reform would not really lead to a simplification.

The spelling reform of 1946–1954

In 1946 and 1947, a spelling reform took place in Belgium and the Netherlands, respectively. This reform consisted of new spelling rules and a new spelling dictionary. In order to be able to appreciate the spelling changes as experienced by language users, I will compare the words in [*Woordenlijst 1954*] with the last unchanged, i.e. [*Woordenlijst 1914*]. The reform affected native as well as loan words.

First consider the changes of native words. These consisted of some simplifications in line with the proposals by Kollewijn (1916). The distinction between *ee* and *e* or *oo* and *o*, *sch* for [s] (except in *-isch*) were abolished and the case marker *-n* became optional:

(2) **Te Winkel**   [***Woordenlijst* 1954**]
  stenen 'to sigh'   *steenen* 'stones'   *stenen/stenen*
  tonen 'toes'   *toonen* 'sounds'   *tonen/tonen*
  mensch   *mens*
  *de**n** vermolmde**n** boom*   *de vermolmde boom*

Some other changes are the following: *mia(a)uw* → *miauw*, *gewikst* → *gewiekst*, *giegelen* → *gie**ch**elen*, *Hunebed/hu**nn**ebed* → *Hu**n**ebed*, *kievit* → *kievi(e)t* and

*ie*zegrim → *i(e)*zegrim.

Loan words were also changed in some cases. These changes, laid down in [*Woordenlijst 1954*] affected hybrid words only; foreign words kept their deviating spelling. Two types of changes were carried out: the abolishment of certain spellings and the introduction of spelling variants.

Examples of letter combinations which were abolished are <ph> for [f] and <rh> for [r]. For instance *physica*, *rhapsodie* became *fysica* and *rapsodie*. Other letter combinations were abolished in certain contexts only. For instance, <th> was changed to <t> word-finally, before a consonant and after <ch> and <f>. A new rule was given for the spelling of [n] in certain contexts: "After -*io* (-*jo*, -*yo*, -*eo*), *n* is written as a single letter, but the ordinary gemination rule is applied when the *n* is followed by a 'schwa'." [Na -*io* (-*jo*, -*yo*, -*eo*) wordt *n* enkel geschreven. Volgt echter op de *n* een 'sjwa' dan geldt de gewone regel van verdubbeling.]. Before 1954, some of the words mentioned in the rule were written with <nn>, e.g. *stationneeren*. When we compare the old spelling of the words with their French origins, we see that the difference between <n> and <nn> is probably etymological. All words that were written with <nn> in 1914 had <nn> in the French word as well, although the reverse is not true, cf. *conventioneel-conventionnel*:

(3)	**1914**	**French**	**1914**	**French**
a	collatio**nn**eeren	collatio**nn**er	commissio**nn**air	commissio**nn**er
	impressio**nn**isme	impressio**nn**isme	statio**nn**eeren	statio**nn**er
	pensio**nn**eeren	pensio**nn**er	petitio**nn**eeren	pétitio**nn**er
	petitio**nn**ement	pétitio**nn**ement	reactio**nn**air	reactio**nn**air
	revolutio**nn**air	revolutio**nn**air	sanctio**nn**eren	sanctio**nn**er
	spio**nn**eeren	espio**nn**er	statio**nn**air	statio**nn**aire
b	constitutio**n**eel	constitutio**n**el	correctio**n**eel	correctio**n**el
	meridio**n**aal	méridio**n**al	natio**n**aal	natio**n**al
	pleo**n**asme	pléo**n**asme		
c	confessio**n**eel	confessio**nn**el	conventio**n**eel	conventio**nn**el
	emotio**n**eel	émotio**nn**able	functio**n**aris	fonctio**nn**aire
	mario**n**et	mario**nn**et	missio**n**aris	missio**nn**aire
	pensio**n**aat	pensio**nn**at	pio**n**ier	pio**nn**ier
	provisio**n**eel	provisio**nn**elle	additio**n**eel	additio**nn**el
	ratio**n**eel	ratio**nn**el	traditio**n**eel	traditio**nn**el

The spelling of the examples in (3a) and (3b) may be considered etymological. The examples in (3c) show that at least some words no longer had etymological geminates in Dutch by 1914. This suggests that, possibly as the effect of shortening, the sound representation of words written with <nn> did not differ so much from that of words written with <n> that this difference prevented the

simplification of <nn>. The new spelling rule in [*Woordenlijst 1954*] can then be seen as a further simplification: the etymological geminates that are not sufficiently supported by the pronunciation are removed. The result is that the spelling is closer to the pronunciation. However, words like *spionage*, *pensionaat* and *stationeer* were subjected to the same rule, although they are related to words with a short vowel: *spion*, *pension* and *station*. These words were probably pronounced with a short vowel as is still the case with *stationnement* and *petitionnement*. The fact that the pronunciation with a short [o] is possible now in *spionage*, etc., is probably an effect of the spelling. The provision "but the ordinary gemination rule is applied when the *n* is followed by a 'schwa" generally distinguishes native suffixes (doubling) from non-native ones (no doubling), but also prescribes doubling in *petitionnement* and *stationnement*. Words with a double or single <n> are also pronounced differently, see De Coninck (1970):

(4)   *petitioneer*   [pətisjɔner]   *petitionnement*   [pətisjɔnəmɛnt]
       stationeer   [stɑsjoner]   *stationnement*   [stɑsjɔnəmɛnt]

It is not clear whether the short vowel in *stationeer* may be considered a spelling pronunciation, but at the moment the spelling is compatible with the (possibly adapted) pronunciation.

    The second type of reform of the spelling of loan words meant that two spelling variants were allowed in many cases, the *voorkeurspelling* (preferred spelling) and the *toegelaten spelling* (permitted spelling). In most cases, the preferred spelling was the etymological spelling, and the other form the version that is adapted to Dutch: *quasi* (also *kwasi*), but there are also many examples of the opposite, e.g. *cilinder* (also *cylinder*), *cipres* (also *cypres*), *kwadraat* (also *quadraat*). In many cases, the words of which the adapted version was the preferred variant were already used in the adapted version (these were the words that Te Winkel accepted in a simplified spelling. Some examples are *kasuaris*, *katafalk*, *kapitein*, *karakter*, *kanarie*, *kameel*, *kapucijn* etc.).

    The effect of this reform was a very unsystematic partial adaptation of hybrid words to the native regularity, notably with respect to the letter <c>. An overview of the changes is given under (5). The overview was compiled by screening [*Woordenlijst 1914*] for words of which the spelling was changed, so it is probably not exhaustive. It only lists spelling changes as the result of partial reform of the spelling of a sound; complete abolishment of spellings such as *ph* are omitted. Spellings separated by 'and' have a different meaning:

(5) Examples of words changed in 1954

*ac**u**stiek/a**koe**stiek → a**koe**stiek*   *adh**ae**sie → adh**e**sie*

*a**e**sthetisch → **e**sthetisch*   *a**e**ther → **e**ther*

*a**e**sculaap → **e**sculaap*   *a**e**quator → **e**quator*

*algebra**i**sch → algebra**ï**sch*   *anci**e**nniteit → anci**ë**nniteit*

*anon**y**miteit → anon**i**miteit*   *arch**ae**oloog → arch**e**oloog*

*art**i**st → art**ie**st*   *as**y**l → as**ie**l*

*athen**ae**um → athen**e**um*   *banke**te**eren → banke**tt**eren*

*bi**lli**oen and bi**lj**oen → bi**lj**oen*   *bivou**a**k → biva**k***

*bran**i** → bran**ie***   *br**u**sk → br**uu**sk*

*belved**e**re → belved**è**re*   *ca**b**aal and **k**abaal → **k**abaal*

*ca**c**ographie → **k**a**k**ografie*   *cadaver → **k**adaver*

*c**ae**suur → c**e**suur*   *catholicisme → **k**atholicisme*

*cath**o**/**k**atholiek → **k**atholiek*   *ch**inaas-/s**inaasappel → sinaasappel*

*c**l**ericaal → **k**lericaal*   *c**l**iniek → **k**liniek*

*c**omedie and **k**omedie → **k**omedie*   *c**omisch → **k**omisch*

*com**p**tant/con**t**ant → con**t**ant*   *contr**ô**le → contr**o**le*

*c**oqu**et- → **k**o**k**et-*   *c**osmetiek → **k**osmetiek*

*c**osmo- → **k**osmo-*   *c**rit- → **k**rit-*

*c**ritiek and **k**ritiek → **k**ritiek*   *c**ylinder → c**i**linder*

***suikerij**/cichorei → cichorei*   *d**ae**mon(isch) → d**e**mon(isch)*

*decast**e**re → decast**è**re*   *diar**rh**ee → diar**r**ee*

*ego**i**st → ego**ï**st*   *electr- → elektr-*

*entr**é**e → entr**e**e*   *enth**u**- → enth**ou**-*

*é**tagère → **e**tagère*   *Farize**ë**r → Farize**e**ër*

*p**hoenix**/f**eniks → f**eniks*   *flot**ille → flot**ielje***

*fo**u**rage, fourag**ee**ren → f**oe**rage, f**oe**rageeren*   *fo**u**rgon → f**oe**rgon*

*fo**u**rier → f**oe**rier*   *fri**kk**adel → fri**k**adel*

*ge**c**ostumeerd → ge**k**ostumeerd*   *gram**oph**one → gra**mm**o**f**oon*

*g**u**inje → g**i**enje*   *hach**é**e and **hachis** → hach**e**e*

*harmoni**c**a → harmoni**k**a*   *Hebre**ë**r → Hebre**e**ër*

*hom**oe**opaat → hom**e**opaat*   *improdu**c**tief → improdu**k**tief*

*huurce**ê**l → huurce**e**l*   *inh**ae**rent → inh**e**rent*

*inse**c**t → inse**k**t*   *i**oo**n → i**o**n*

*jubil**é** → jubil**e**e*   *kalm**u**s → kalm**oe**s*

***kr**af/**kar**af → **kar**af*   ***k**analje/**c**anai**ll**e → **c**anai**ll**e*

*klavecim**baal** → klavecim**bel***   *ko**k**erellen → ko**kk**erellen*

*kop**ïi**st → kop**ii**st*   *lingu**i**st → lingu**ï**st*

*lin**i**aal → lin**e**aal*   *lo**c**aal and lo**k**aal → lo**k**aal*

*lo**c**aliteit → lo**k**aliteit*   ***loupe** → **loep***

*Ma**cch**iavellisme → ma**ch**iavellisne*   *ma**hom**edaans → m**ohamm**edaans*

*ma**e**cenas → m**e**cenas*   *mari**o**-/mar**j**olein → mar**j**olein*

marechaussee/maréchaussée → *marechaussee*

millioen → *miljoen*

minute → *minuut*

muziekkorps/corps → *muziekkorps*

October → *oktober*

pedestal → *piëdestal/pedestal*

piramide/pyramide → *piramide*

praedicaat → *predikaat*

praeparaat → *preparaat*

praeses → *preses*

purim → *purim/poerim*

quaesti- → *kwesti-*

quantiteit- → *kwantiteit-*

quint- → *kwint-*

ragout → *ragoût*

rapalje → *rapaille*

reliquie → *relikwie*

resumtie → *resumptie*

risqueeren → *riskeren*

châle/sjaal → *sjaal*

Sanskrit → *Sanskriet*

sceptisch → *sceptisch/skeptisch*

souverein → *soeverein*

sekuur → *secuur*

sexueel → *seksueel*

soloecisme → *sol(o)ecisme*

syring/sering → *sering*

talmud → *talmud/talmoed*

tirannij/ie → *tirannie*

tram → *tram/trem*

tracteeren → *trakteren*

triumph → *triomf*

vanielje/vanille → *vanille*

wolfram → *wolfra(a)m*

milliard → *miljard*

millionnair → *miljonair*

mirre/myrrhe → *mirre*

muscus → *muskus*

orientalist → *oriëntalist*

phaenomenaal → *fenomenaal*

poenitent → *penitent*

praefix → *prefix*

praesens → *presens*

priëel → *prieel*

quadraat- → *kwadraat-*

qualificatie- → *kwali-*

quarto → *kwarto*

quitantie → *kwitantie*

ran(t)soen- → *rantsoen-*

recru- → *rekru-*

request(eeren) → *rek(w)est(eren)*

reveille → *reveil*

saccharine → *sacharine*

Sadduceër → *Sadduceeër*

savoyekool → *savooiekool*

schacheren → *sjacheren*

secreet/sekreet → *secreet*

sequester- → *sekwester (sequ-)*

soleneel → *solemneel*

soirée → *soiree*

tabouret → *taboeret, tabouret*

tyran/tiran → *tiran*

tornooi → *to(e)rnooi*

tractaat (traktaatje) → *traktaat*

trillioen → *triljoen*

vacantie → *vakantie*

villatje → *villaatje*

zefi(e)r → *zefier*

The spelling reform of 1995

In 1995, a spelling reform was carried out which consisted of the following changes:

(6) **Spelling reform of 1995**

1954	**1995**
• spelling variants	• one spelling[1]
• spelling of linking morpheme	• spelling of linking morpheme
• (*-e* or *-en*) based on number	• based on plural ending
• diaeresis in some compounds (*naäpen*)	• hyphen in compounds (*na-apen*)
• â, ô, û (*bâton, entrecôte, ragoût*)	• a, o, u (*baton, entrecote, ragout*)[2]

The remaining spelling variant was the preferred spelling (*voorkeurspelling*), except in the following cases:

(7) **Changes in the preferred spelling**

antikrist → antichrist	catheter → katheter
croquet → kroket	dioxyde → dioxide
elektrokuteren → elektrocuteren	elektrokutie → elektrocutie
emfaze → emfase	fotocopie → fotokopie
fotocopiëren → fotokopiëren	harmonika → harmonica
insekt → insect	komplot → complot
komplotteren → complotteren	korpus → corpus
kwaker → quaker	lambrizeren → lambriseren
lambrizering → lambrisering	macrocosmos → macrokosmos
mediaevist → mediëvist	microcosmos → microkosmos
oxydatie → oxidatie	oxyde → oxide
oxyderen → oxideren	prae → pre
praeses → preses	prakkizeren → prakkiseren
praktizeren → praktiseren	predikatief → predicatief

---

[1] Not all variation was abolished: equivalent spelling variants (marked by 'see also' [zie ook] or brackets) are often still present, e.g. *plafon(d), aanbeeld* or *aambeeld, Purim* or *Poerim* and *schibbolet* or *sjibbolet*. The allowed variant *toast* became equivalent to *toost*. However, in some cases spelling variants of equal status were deleted, e.g. *synchroon/synkroon, cichorei/cikorei* became *synchroon, cichorei*. On the other hand new variants were introduced: next to *apropos, jioe-jitsoe, poelet, choqueren, yuca, koosjer* and *kousjer, morille* and *madam*, the dictionary now also contains *à propos, jiu-jitsu, poulet, shockeren, yucca, kosjer, morielje* and *madame*. New words were sometimes introduced in two variants: *bibs* or *bips, had(z)ji, Jahwe(h), müsli* or *muesli, rauwdouwer* or *rouwdouwer, sanseviëria* or *sanseveria, spinnaker* or *spinaker, staatsieportret* or *statieportret, stampei* or *stampij, stennis* or *stennes* and *tekkel* or *teckel*

[2] However, î and ê were not abolished, cf. *maîtresse, crêpe.*

produkt → product   produktie → productie
produktief → productief   produktiviteit → productiviteit
propaedeuse → propedeuse   propaedeutisch → propedeutisch
publikatie → publicatie   quantum → kwantum
vredestractaat → vredestraktaat   vulcanisatie → vulkanisatie
vulcaniseren → vulkaniseren

The spelling reforms that were applied to non-native words in 1954 were not applied to new loan words, so that some new inconsistencies arise: *questionnaire* next to *miljonair*, *dédain* next to *detente* and *urethra* next to *antraciet*.

It can be concluded that the spelling reforms did not change the basic character of Dutch orthography.

Samenvatting

Het Nederlands heeft een alfabetische spelling waarbij elke klank wordt weergeven door een letter of lettercombinatie. Er is echter in het Nederlands geen perfecte één-op-één-relatie tussen klanken en letters (verder fonemen en grafemen genoemd). Dit is het gevolg van factoren als een gebrek aan geschikte letters om een unieke code voor alle klanken te verzorgen, het in de loop van de tijd optreden van klankveranderingen die niet gepaard gaan met aanpassingen in de spelling, en de instroom van leenwoorden met een afwijkende spelling. Hierdoor zijn er verschillende soorten spellingsregels nodig die vastleggen hoe een klank in een gegeven context wordt geschreven.

Kennis van de Nederlandse spellingsregels is van theoretisch en praktisch belang. Een beschrijving van het spellingssysteem stelt ons in staat om vast te stellen hoe de spelling klankrepresentaties vastlegt en welke schrijfwijzen regelmatig zijn en welke uitzonderlijk. Daarnaast kan kennis van het spellingssysteem nuttig zijn voor het schrijfonderwijs, voor de consistentie van de spelling in woordenboeken, en voor toepassingen als spellingcontrole.

Ondanks theoretisch en praktisch nut van een beschrijving van de spelling bestond er nog geen volledig overzicht van spellingsregels en de uitzonderingen daarop. Natuurlijk zijn de belangrijkste spellingsprincipes al vastgelegd door Te Winkel (1863), die de eerste officiële spellingregeling voor Nederland en België introduceerde. Deze principes luiden als volgt:

(1) **Regel der Beschaafde Uitspraak**
Geef door letterteekens al de bestanddeelen op, die in een woord gehoord worden, wanneer het door beschaafde lieden zuiver uitgesproken wordt.
Regel der Gelijkvormigheid
Geef, zooveel de uitspraak toelaat, aan een zelfde woord en aan ieder deel, waaruit het bestaat, steeds denzelfde vorm.
Regel der Afleiding
Bij de keus der gelijkluidende letterteekens beslist de afleiding of de oudere vorm uit den tijd, toen de nu gelijk geworden klanken nog duidelijk onderscheiden konden worden.
Regel der Analogie
[…] de woorden wier spelling noch door de uitspraak, noch door de gelijkvormigheid, noch door de afleiding wordt bepaald, worden op dezelfde wijze geschreven als andere, wier spelling met zekerheid bekend is en die oogenschijnlijk op overeenkomstige wijze gevormd zijn.

De Regel der Beschaafde Uitspraak en de Regel der Afleiding staan ook wel bekend als het Fonologische en Etymologische Principe. De voorwaarde 'zooveel de uitspraak toelaat' bij de regel der Gelijkvormigheid wordt wel het Terugleesbaarheidsprincipe genoemd. De Regel der Analogie en de Regel der Gelijkvormigheid worden hier samen aangeduid als het Morfologische Principe.

Deze principes uit 1863 vormen nog steeds de basis van de Nederlandse spelling, want de spelling is sindsdien niet essentieel gewijzigd. De veranderingen die hebben plaatsgevonden, kunnen worden gekarakteriseerd als het verminderen van de invloed van het Etymologische Principe. Te Winkels spellingprincipes wijzigden de op dat moment bestaande schrijfpraktijk niet ingrijpend. Dit wijst erop dat deze principes niet alleen de juiste spelling *voorschrijven*, maar ook de bestaande spellingsgewoonten *beschrijven*.

Nauwkeurige bestudering van de spelling van Nederlandse woorden laat zien dat er ook generalisaties te maken zijn over de Nederlandse spelling die Te Winkel niet noemde, en die hij mogelijk ook niet heeft gezien. Voorbeelden zijn het verschil tussen <i> en <ie> in *neurie-neuriën* en de afwisseling van enkele en dubbele letters in *raam-hamer*, *brem-emmer*. Geen van Te Winkels principes verklaart deze afwisselingen, sterker nog, ze zijn in strijd met het Fonologische en Morfologische Principe. Dit soort context-afhankelijke spellingvariatie komt teveel voor en vertoont teveel regelmaat om van een inconsistentie te kunnen spreken. Het lijkt erop dat er een ander principe is dat deze schending van principes afdwingt, maar dat principe noemt Te Winkel niet. Dit onderzoek is bedoeld om dergelijke impliciete generalisaties op te sporen.

Werkwijze
Om inzicht te verkrijgen in het Nederlandse spellingssysteem ben ik als volgt te werk gegaan. Uit de CELEX-taaldatabank werd de spelling en uitspraak van ongeveer 45.000 trefwoorden (ongelede woorden en woorden met voor- en achtervoegsels) gehaald. Vervolgens stelde ik een set spellingsregels samen die in een computerprogramma verwerkt werden. Met behulp van dit programma werd de uitspraak van de woorden omgezet in spelling, en de aldus verkregen spelling werd vergeleken met de spelling uit de databank. Op deze manier kon het effect van de spellingsregels geëvalueerd worden en konden uitzonderingen worden opgespoord. Wanneer de berekende spelling afweek van de eigenlijke spelling, heb ik geprobeerd dit te verhelpen door het verbeteren van de spellingsregels. De woorden die ook na verbetering van de spellingsregels niet van een correcte spelling konden worden voorzien, worden als uitzonderingen aangemerkt (zie Appendix C-E). Het uitgangspunt van de spellingsregels werd gevormd door de spellingsvoorschriften in de meest recente editie van het Groene Boekje uit 1995. Ook werden regels uit rapporten van spellingcommissies en uit verschillende beschrijvingen van delen van de spellingssystematiek gebruikt.

Spellingssysteem

Te Winkels principes suggereren dat spelling als volgt wordt afgeleid uit de uitspraak:
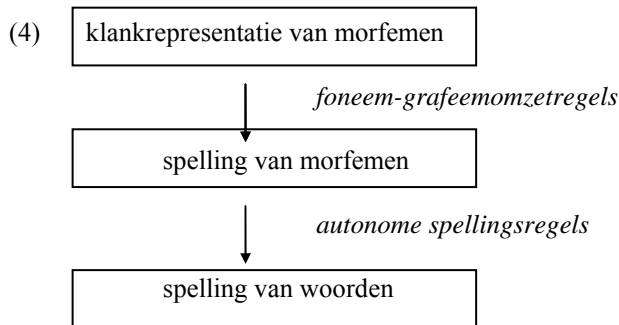
(2)
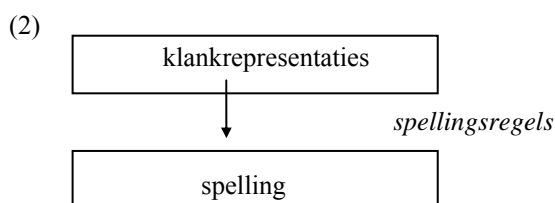
In veel gevallen levert model (2) een goed uitgangspunt om het spellingssysteem te beschrijven, maar er zijn ook verschijnselen die niet of niet optimaal verantwoord kunnen worden binnen model (2) en die twee bijstellingen van het model nodig maakten.

De eerste wijziging van model (2) die hier wordt voorgesteld betreft het toevoegen van een tweede soort spellingsregels. Niet alle spellingpatronen kunnen goed beschreven worden wanneer ze louter als een weergave van klanken gezien worden. Sommige spellingsverschijnselen zijn weliswaar voorspelbaar, maar op basis van letters in plaats van klanken. Zo zijn afwisselingen van enkele en dubbele letters in *raam-ramen* en *laf-laffe* en de plaats van afbreekposities beter voorspelbaar op basis van lettergrepen (en letters) dan op basis van lettergrepen in de uitspraak (en klanken), zoals geïllustreerd in (3):

(3)	**uitspraak**	**voorspelde spelling**	**juiste spelling met lettergrepen**
a	ba-jərt	*bajerd (vgl. bajes)	baai-erd
b	he-lɑl	*helal (vgl. heler)	heel-al
c	lɑ-fɑrt	*laffaard (vgl. laffe)	laf-aard
d	pɔ-xə	*pochchen (vgl. rogge)	po-chen
	jɔ-ŋə	*jongngen (vgl. bonnen)	jon-gen

De voorbeelden in (3) laten zien dat spelling beter beschreven kan worden wanneer die niet in één keer van de uitspraak wordt afgeleid, maar via een tussenstap, een onderliggende spellingsrepresentatie op grond waarvan lettergrepen kunnen worden gemaakt. Om de verschijnselen in (3) beter te kunnen beschrijven, heb ik (2) uitgebreid met een tweede soort spellingsregels, autonome spellingsregels. Dit levert het volgende spellingmodel op (morfemen zijn de woorddelen waarvan in het Morfologische Principe sprake is):

(4) | klankrepresentatie van morfemen |

foneem-grafeemomzetregels

| spelling van morfemen |

autonome spellingsregels

| spelling van woorden |

In dit model worden lettergrepen gevormd op basis van letterreeksen. De spellingsverschillen tussen woorden als *baaierd* en *bajes* kunnen eenvoudig verantwoord worden doordat de regels die letters groeperen tot lettergrepen <j> anders behandelen dan <k>. Het bijzondere gedrag van *heelal* en *lafaard* in vergelijking met *heler* en *laffe* ligt aan het feit dat in de spelling lettergreepgrenzen soms samenvallen met het begin van morfemen terwijl lettergrepen in de uitspraak meestal worden toegekend als in een ongeleed woord. De verschillen tussen *rogge*, *jongen* en *pochen* worden veroorzaakt doordat de regels die lettergrepen vormen gevoelig zijn voor de groepering van letters in grafemen (letters en vaste lettercombinaties). Zo zijn de regels gevoelig voor het verschil tussen grafemen die bestaan uit één of twee letters (*g* en *ch*) en voor het verschil tussen twee letters die al dan niet één grafeem vormen (*ch* en *ng*).   Het feit dat spelling niet louter een weergave vormt van de uitspraak, wordt niet door een van Te Winkels principes verantwoord. Daarom heb ik een extra principe voorgesteld, het Grafotactisch Principe. Dit Principe bepaalt dat letterreeksen aan bepaalde welgevormheidseisen moeten voldoen. Zo mogen er bijvoorbeeld geen identieke letters in dezelfde lettergreep staan (*lopen/*loopen*). Aan deze eis wordt echter niet voldaan wanneer de terugleesbaarheid in het gedrang komt (we schrijven *goochem*, niet **gochem*), hetgeen laat zien dat het bovengenoemde Terugleesbaarheidsprincipe ook voor autonome spellingsregels geldt.

De introductie van autonome spellingsregels maakt het mogelijk om te verantwoorden dat niet alle spellingsregels het Morfologische Principe gehoorzamen. Dit principe schrijft voor dat morfemen altijd hetzelfde geschreven worden. Dit is het geval in (6a), maar niet in (6b):

(6)   **De spelling van verwante woorden**
  a   trouw   ge*trouw*d   *getroud   (vgl. koud)
       aai   *aai*en   *ajen   (vgl. bajes)
  b   raam   **raam*en   ramen   (vgl. hamer)
       ham   **ham*en   hammen   (vgl. jammer)

Regels zoals die in (6a) die tot een gelijkvormige spelling van verwante woorden leiden zijn foneem-grafeemomzetregels, terwijl de regels onder (6b) die wel afwisselingen veroorzaken autonome spellingsregels zijn. Het contrast kan dus verantwoord worden door aan te nemen dat het Morfologische Principe alleen foneem-grafeemomzetting betreft, en niet autonome spellingsregels. Met andere woorden, de eerste stap van het afleiden van de spelling werkt op morfemen, de tweede op woorden.

De autonome spellingsregels bieden ook een nieuw perspectief voor de feiten in (7). Op het eerste gezicht lijken hier de klankregels verscherping (het vervangen van een stemhebbende medeklinker door een stemloze tegenhanger) en degeminatie (het vervangen van twee identieke medeklinkers door een enkele) inconsistent weergegeven te worden in de spelling:

(7)   **Gelijkvormige spelling**   **Ongelijkvormige spelling**

       Uitspraak   **Spelling**   **Uitspraak**   **Spelling**
  a   [strɑndə-strɑnt] stranden-stran**d**   [levə-lef]   leven-lee**f**
       [hɛbə-hɛp]   hebben-he**b**   [vrezə-vres]   vrezen-vree**s**
       [draɣə-drax]   dragen-draa**g**
  b   [hat-hat]   haat-haa**tte**   [hat-ɣəhat]   haat-gehaa**t**
       [lɑnt-lɑndə]   land-land**de**   [lɑnt-ɣelant]   land-gelan**d**

De voorbeelden in (7a) suggereren dat verscherping van /d/, /b/ en /ɣ/ niet in de spelling wordt weergeven, maar van /v/ en /z/ wel. Deze inconsistentie verdwijnt wanneer we aannemen dat de afwisseling het gevolg is van een autonome spellingsregel verscherping, die losstaat van de klankregel en die alleen van toepassing is op <v> en <z>. De voorbeelden in (7b) suggereren dat degeminatie wel op woordeinde maar niet binnen woorden wordt weergegeven. Ook deze feiten zijn niet langer inconsistent wanneer we een autonome spellingsregel aannemen die wel werkt op identieke letters in dezelfde lettergreep (*ge-haatt*) maar niet over lettergreepgrenzen heen (*haat-te*). We kunnen nu aannemen dat spelling in (7) altijd de klankrepresentatie van morfemen weergeeft, en dat de variatie net als bij *raam-ramen* pas later geïntroduceerd wordt. Door gebruik te maken van autonome spellingsregels kunnen we dus zowel de gelijkvormige als

de ongelijkvormige spelling van morfemen verantwoorden.

Een tweede bijstelling van het eenvoudige model (2) is nodig voor de beschrijving van leenwoorden. Een strikte toepassing van het Etymologische Principe op leenwoorden betekent dat leenwoorden de spelling uit de taal van afkomst hebben en dus in principe uitzonderlijk zijn. Toch zijn er veel generalisaties te maken over de spelling van leenwoorden, maar andere dan voor inheemse woorden. Om deze generalisaties te kunnen benutten bij de spellingsbeschrijving is het wel nodig dat we leenwoorden kunnen herkennen (zonder daarbij de spelling zelf te gebruiken). Zoals al geobserveerd door Te Winkel kan aan deze voorwaarde inderdaad worden voldaan door te bepalen hoe inheemse woorden zich wat betreft uitspraak en verbuiging gedragen (ze hebben bijvoorbeeld per morfeem niet meer dan een volle klinker), en alle woorden die die eigenschappen niet hebben (b.v. *aria*) leenwoorden te noemen.

Generalisaties over de spelling van leenwoorden zijn soms gebaseerd op de meest voorkomende schrijfwijzen uit de taal van herkomst. Zo wordt de klank /ɑu/ in inheemse woorden meestal als <ou> geschreven maar in leenwoorden vaker als <au>. Bovendien worden leenwoorden soms op een systematische manier gedeeltelijk aangepast aan de Nederlandse regelmaat zoals wordt geïllustreerd in (8)

(8) a **Inheemse woorden**
   /i/ → ie   *niet, gieter*
   /e/ → ee   *deze, zee*
   /z/ → z   *zaak, ezel*
   b **Leenwoorden**
   /i/ → ie in de laatste lettergreep   *papier, olie*
   /i/ → i elders   *liter, piste*
   /e/ → ee in de laatste lettergreep   *trochee, systeem*
   /e/ → e elders   *menuet, mechanisch*
   /z/ → z in de laatste lettergreep   *precieze*
   /z/ → s elders   *isolatie*

(In *precieze* staat de /z/ in de laatste lettergreep van het morfeem *precies*. De <s> in de ongelede vorm is het gevolg van de bovengenoemde spellingsregel verscherping.) Veel spellingswijzigingen uit 1954 blijken volgens het patroon in (8) te verlopen: *asyl* en *soirée* werden *asiel* en *soiree*, en *cylinder* en *aesculaap* werden *cilinder* en *esculaap* (niet **cielinder* en **eesculaap*). Regels als die in (8) worden niet gegeven in het Groene Boekje of elders in de literatuur (behalve in het geval van de /i/, waarvoor in de editie van het Groene Boekje uit 1954 wel aparte regels waren).

Om patronen als die in (8) te beschrijven werden de foneem-grafeemomzetregels gesplitst in regels voor inheemse woorden en leenwoorden.

Autonome spellingsregels houden geen rekening met het verschil tussen inheemse woorden en leenwoorden. De twee soorten spellingsregels hebben verschillende eigenschappen wat betreft toepassingsdomein (morfeem of woord), context (klanken of letters), en de gevoeligheid voor de herkomst van woorden.

De hier voorgestelde spellingsregels voorspellen voor 85% van de 45.000 woorden de spelling correct. Met andere woorden, computerprogramma's of schrijvers die deze regels gebruiken, moeten de spelling van 15% van de woorden opslaan. Wanneer we de woordenschat uitsplitsen in inheemse woorden en leenwoorden blijkt dat 95% van de inheemse woorden regelmatig is. De meeste woorden waarvan de spelling niet voorspelbaar is bevatten de klanken /ɛi/ of /ɑu/ (b.v. *wei-wij*, *rouw-rauw*). Van de overige uitzonderingen heeft een deel een historische spelling (b.v. *thuis*, *erwt*); een ander deel bestaat uit leenwoorden die alleen aan hun schrijfwijze als zodanig herkenbaar zijn (b.v. *hetze*, *cirkel*). Van de leenwoorden heeft 73% een voorspelbare spelling. Ter vergelijking, wanneer we de spelling van leenwoorden berekenen met behulp van de foneem-grafeemomzetregels voor inheemse morfemen wordt maar 25% correct gespeld (dat wordt 30% procent wanneer we wel de juiste regel voor de klank /i/ gebruiken).

Acknowledgements

I would like to thank all those people who supported me in writing this book:

- Anneke Neijt for her inspiring enthusiasm
- Henneke de Bruijn, Simon van Dreumel, Bob Nunn, Petra Poelmans and Edith Schouten and Peter Tiemeijer for their invaluable help
- my colleagues in Nijmegen for the pleasant coffee breaks

Curriculum Vitae

Anneke Nunn was born in Billericay, Great Britain on October 3, 1965. In 1984 she graduated from the Eindhovens Protestants Lyceum (gymnasium α). She then entered the University of Utrecht where she received her master's degree (cum laude) in Dutch Language and Literature in 1989. From 1989 to 1991 she was a researcher for the national research program 'Analysis and Synthesis of Speech' at the Phonetics Department of Leiden University. In 1992 and 1993 she was a research assistant for the Spellingcommissie (spelling (reform) committee). In 1994 she became a junior researcher at the Department of Dutch Language and Literature of the University of Nijmegen. After finishing the investigation that resulted in this thesis, she started working for Van Dale Lexicografie.