In

Philip A. Luelsdorf (ed.)

Orthography and Phonology

Amsterdam/Philadelphia:

John Benjamins,

1987, 215-224

THE REFLECTION OF LINGUISTIC STRUCTURE IN DUTCH SPELLING

Geert E. Booij

1. *Introduction*

Most linguists correctly assume that speech is the primary medium of language, and that writing is only a secondary, derived medium for conveying language.[1] This is already clear from the fact that every existing natural language is spoken, but not always written.[2] One might therefore be inclined to conclude that spelling is nothing but a graphical (written or printed) representation of speech, the audible form of language. It is the aim of this paper to show that this is a much too simple conception of spelling. I will argue that an orthographical system represents different levels of language structure, taking my examples mainly from Dutch.[3]

2. *Spelling and phonology*

Dutch spelling[4] is often characterized as a more-or-less phonemic spelling, i.e. a spelling system in which letters or combinations thereof correspond to phonemes. By qualifying a spelling as phonemic, we already imply that written language not only abstracts away from the individual and situational variation in the pronunciation of words, but also from the effects of allophonic rules. That is, spelling does not represent speech, but a more abstract level of language structure.

However, it is not correct to consider Dutch spelling phonemic. For instance, the word *banaan* 'banana' can be pronounced as either [banán] or bənán]. In the second phonetic form, the /a/ has been reduced to a schwa due to a rule of unstressed vowel reduction (cf. Booij 1977, 1981). A phonemic spelling would provide the orthographic form *benaan* (*e* stands for schwa in Dutch), since the schwa is a phoneme of Dutch. Yet, the correct

spelling is *banaan*. This may suggest that Dutch spelling represents underlying forms instead of phonemic representations, but this would be too hasty a conclusion. It appears that both underlying forms, phonetic forms and intermediate levels can be represented in Dutch spelling, as will be illustrated below. Therefore, it is much more adequate to characterize the Dutch spelling system by first establishing a typology of phonological rules, and then determining of which kinds of rule the effects are visible in spelling.

Rule typology is a neglected issue in standard generative phonology, but has become more important in recent developments in generative phonology such as Natural Generative Phonology (cf. Hooper 1976) and Lexical Phonology (cf. Kiparsky 1982). We may distinguish the following types of rules (cf. Booij 1981: ch. 8):

A. *Rules of sentence phonology* These are the rules which apply across word boundaries, such as the rules of external sandhi. Well known examples are the rules of voice assimilation in many Indo-European languages. Usually, these rules are optional.

B. *Rules of word phonology* These are the rules that only apply within words. The following subclasses have to be distinguished:

(i) *obligatory rules*, like:

(1) Syllable-final devoicing
[− son] → [− voice] / ——)$_\sigma$ e.g. /hud/ 'hat' → [hut]

(2) Devoicing of obstruents before tautosyllabic voiceless obstruents
[− son] → [− voice] / —— [− voice])$_\sigma$ e.g. /ɣud+s/ 'good' → [ɣuts]

(3) Geminate simplification
$C_i\ C_i \rightarrow C_i$ e.g. /et+t/ 'eats' → [et]

Rules (2) and (3) together derive the phonetic form of (*hij*) *biedt* '(he) offers':

(4)

underlying form	/bid+t/
rule 2	t
rule 3	ø
phonetic form	[bit]

(ii) *morpholexical rules* (cf. Anderson 1974): rules which apply to a restricted (listed) class of words, or to words with a certain morphological property or specific morpheme. Well known examples are the German rules of Ablaut and Umlaut. Examples from Dutch are the rules which account for the allomorphy in the diminutive suffix, which has five allomorphs: *-tje*, *-je*, *-etje*, *-kje* and *-pje*:

(5)

traan 'tear'	traantje
kam 'comb'	kammetje
kant 'lace'	kantje
koning 'king'	koninkje
riem 'belt'	riempje

The specific form of the diminutive suffix is determined by the phonological form of the stem, but the rules only apply to diminutives. Morpholexical rules also tend to have (negative or positive) exceptions. For instance, several words have two diminutive forms (e.g. *weg* 'way' *-wegje/weggetje*, *kip* 'chicken' *-kipje/kippetje*), and in the pair *bloempje/bloemetje* (from *bloem* 'flower') we even find semantic differentiation: the meanings of the diminutive forms are 'little flower' and 'bunch of flowers', respectively.

(iii) *optional, style-dependent rules*. A relevant example is the rule of vowel reduction referred to above, which optionally reduces unstressed vowels to schwa.

(iv) *phonotactic rules*: rules which express restrictions with respect to the distribution of sound segments. For instance, in Dutch the nasal [n] never precedes a tautosyllabic /p/ or /k/: the nasal consonant is always [m] before /p/ and [ŋ] before /k/. Compare:

(6)

damp	[dɑmp]	*damk	[dɑmk]
*danp	[dɑnp]	*dank	[dɑnk]
*dangp	[dɑŋp]	dank	[dɑŋk]

Given the typology of rules outlined above, we are now able to give an adequate characterization of the Dutch orthographical system:

(i) Dutch spelling abstracts away from the effect of the rules of sentence phonology, i.e. the spelling represents words as spoken in isolation.
(ii) Dutch spelling does not represent the effect of optional, style-governed rules of word phonology, unless one wants to stress the informality of the kind of speech used (for instance in a written story).
(iii) The effect of morpholexical rules is always visible in spelling. Con-

sequently, the diminutive suffix with its five allomorphs also has five different written forms. For instance, we do not write *raamtje* 'little window' (/-tjə/ is generally assumed to be the underlying form of the diminutive suffix), but *raampje*, and *kannetje* 'little can', not *kantje*. This has a very desirable consequence: phonetic recoverability without recourse to syntactic context. For instance, if we were to spell *kantje* for *kannetje*, this form could also be interpreted as the diminutive form of *kant* 'lace'. Such ambiguities can only be resolved by taking context into account. Similarly, *bloem* 'flower' has two diminutive forms, with different meanings, as pointed out above. If spelling abstracted away from the effect of morpholexical rules, both words would be spelled as *bloemtje*, a graphical form which would be ambiguous both with respect to phonetic and semantic interpretation.

(iv) Dutch spelling does not abstract away from the effect of phonotactic rules either. We do not write *danp* instead of *damp*, although the phonetic realization [dɑmp] would be predictable. There is only one exception here: before a /k/ the velar nasal [ŋ] is spelled as *n*, e.g. *dank* instead of *dangk*.

(v) The last category of rules that needs discussion here is that of obligatory rules of word phonology. With respect to these rules the Dutch spelling system is rather inconsistent. For instance, we do abstract away from the effect of syllable-final devoicing of obstruents as far as /b,d,ɣ/ are concerned, but not with respect to the fricatives /v,z/:

(7)

rib 'rib'	ribben	(plural)	*but*:	brief 'letter'	brieven (plural)
hoed 'hat'	hoeden	(")		huis 'house'	huizen (")
rug 'back'	ruggen	(")			

Normally, Dutch spelling does not represent the effect of rule (2) (cf. *goeds* for /ɣud+s/ [ɣuts]), but in (*iets*) *vies* '(something) dirty' the phonetic form [vis] is represented instead of the underlying form /viz+s/.

Finally, the effect of Geminate Simplification (3) is usually represented in spelling, as in *eet*, but not in words with the female suffix *-ster* /fits+stər/ 'female cyclist': *fietsster*.

Summarizing, Dutch orthography never represents the effect of optional rules (rules of sentence phonology and stylistic word-internal rules). It always represents the effects of morpholexical rules. The effect of obligatory word-internal rules is represented partially. Consequently, Dutch spelling may represent both underlying forms, phonetic forms and intermediate levels of phonological derivation.

An illustrative example of an orthographical form which represents an intermediate step of a phonological derivation is the spelling of the past tense singular of verbs with stems ending in underlying /v/ or /z/, e.g. *roofde* 'to rob', past sg.':

(8)

underlying form	/rov+də/
syllabification	$(rov)_\sigma$ $(də)_\sigma$
rule 1	f
voice assimilation	v
phonetic form	[rovdə]

In this case, the orthographical form *roofde* represents the intermediate level of derivation /rofdə/.

It will be clear from this description of Dutch orthography (i) that it represents different levels of abstraction and different levels of phonological structure, and (ii) that it is best described by making use of an independently motivated typology of phonological rules.

3. *Spelling and hierarchical structure*

So far, we described Dutch spelling as a system that represents linear sequences of sound segments at varying levels of abstraction. However, it also provides syntactic information. Capital letters mark the beginning of sentences and also mark proper names. Punctuation is used to mark the boundaries between constituents and the end of clauses. Spacing is used to mark word boundaries. In many cases these word boundaries indicated by spacing are both syntactic and phonological word boundaries, since usually syntactic word boundaries coincide with underlying phonological word boundaries. On the phonetic level, however, phonological word boundaries are often blurred by syllabification across word boundaries, as is illustrated by the phonetic syllabification pattern of *op een oor* 'on an ear': $(ɔ)_\sigma$ $(pə)_\sigma$ $(no{:}or)_\sigma$. When the number of syntactic word boundaries is not identical to that of phonological word boundaries, as is the case with compounds which contain more than one phonological word (see below), spacing gives priority to syntactic word boundaries.

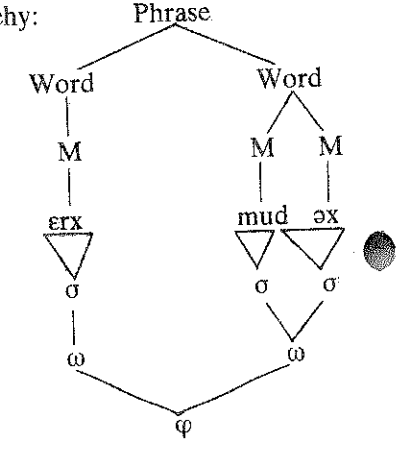
Thus, these orthographical conventions provide a wealth of information about the syntactic structure of a sentence. This is very important, because the phonetic form of a sentence does not always directly reflect its syntactic structure. Therefore, a graphic representation of the phonetic form of a sentence would be very inconvenient from the perceptual point of view.

One should realize, however, that sequences of sound segments are hierarchically structured in two different ways, since there are two hierarchies to be distinguished: a morphological-syntactic hierarchy on the one hand, and a phonological one on the other:

(9) i the morphological-syntactic hierarchy: morpheme (M), word (W), phrase (Ph);
ii the phonological hierarchy: syllable (σ), phonological word (ω), phonological phrase (φ).

The idea that sequences of sound segments are structured in two different ways has been stressed repeatedly by Pike (e.g. Pike 1979), and has been also pointed out by, for instance, Halle and Clements (1982) and Booij (1983). Let me first illustrate the difference between these hierarchies by means of an example from Dutch, the structure of the phrase *erg moedig* 'very courageous'. The word *moedig* is derived from *moed* 'courage' by means of the suffixation of *-ig*:

(10) morphological-syntactic hierarchy:



phonological hierarchy:

The difference between the two hierarchies is clear when we look at *moedig*: the internal morphological structure *moed-ig* is non-isomorphic with the phonological structure *moe-dig*.

Lack of isomorphism between the two hierarchies is also found in compounds and certain types of derived word. For instance, the compound $[[oog]_N\ [arts]_N]_N$ 'lit. eye doctor' is one syntactic word, but consists of two phonological words, *oog* and *arts*. If *oogarts* were one phonological word, we would expect the syllabification pattern $(o)_\sigma\ (\gamma arts)_\sigma$, in accordance with the Maximal Onset Principle. The existence of a syllable boundary after *oog* is also proven by the phonetic form of this compound: $(ox)_\sigma\ (arts)_\sigma$ with syllable-final devoicing of the underlying /γ/. It is known for many languages that the internal morphological boundary of compounds coincides with a syllable boundary. Since the phonological word is the domain of syllabification, we have to assume that compounds consist of more than one phonological word.

The official Dutch spelling marks syntactic boundaries only, and therefore there will be no space between the parts of a compound. On the other hand, many English compounds are written with an internal space, and thus English orthography gives priority to the phonological hierarchy. Interestingly, there is also a tendency among spellers of Dutch, in particular the less skilled ones, to separate the parts of compounds by spaces, in accordance with their phonological intuitions.

Certain classes of derived word also consist of more than one phonological word. Generally, one might distinguish two kinds of affixes, cohering affixes and non-cohering ones (cf. Booij 1983). Cohering affixes fuse phonologically with their stems, while non-cohering affixes form phonological words of their own. Certain Dutch suffixes, like *-achtig* '-like', *-heid* '-ness' and *-baar* '-able' are independent phonological words.[5] Compare, for instance, the suffixes *-ig* and *-achtig*, which have the same meaning, with respect to syllabification (cf. Booij 1977: 1981):

(11) $[[rod]_A\ ig]_A$ 'reddish' $(ro)_\sigma\ (dəx)_\sigma$
$[[rood]_A\ achtig]_A$ 'reddish' $(rot)_\sigma\ (ax)_\sigma\ (təx)_\sigma$

Phonologically, words like *roodachtig* have the same structure as compounds, which is corroborated by the fact that they also have the same stress pattern: primary stress on the first phonological word, secondary stress on the second.

Certain prefixes show the same kind of behaviour with respect to syllabification, e.g. the prefixes *ont-* and *ver-*:[6]

(12) $[ont\ [aard]_N]_V$ 'to degenerate' $(ɔnt)_\sigma\ (a{:}rt)_\sigma$
$[ver\ [as]_N]_V$ 'to cremate' $(vɛr)_\sigma\ (as)_\sigma$

These observations with respect to the non-isomorphism between morphological-syntactic and phonological structure are not only relevant with respect to spacing, but also with respect to hyphenation rules. The hyphenation rules of Dutch crucially depend on the phonological hierarchy, i.e.

the hyphen comes after a syllable boundary, not after a morpheme boundary (for instance, we hyphenate *moedig* as *moe-dig*, not as *moed-ig*). Compare this to the English spelling which puts a hyphen after a morphological boundary, if possible (e.g. *morphem-ic*, not *morphe-mic*). A derived word like *roodachtig* will be hyphenated as *rood-achtig*. At first sight one might be inclined to conclude that here morphological structure is the determining factor. However, we have seen above that in this word a morphological boundary coincides with a syllable boundary, and hence we can conclude that, after all, the hyphenation rules of Dutch are consistently phonologically oriented. The same applies to prefixed words, which are hyphenated as follows: *ont-aard*, *ver-as* etc..

If possible, in Dutch orthography a higher phonological category takes precedence over a lower one. That is, Dutch spelling prefers a hyphen after a phonological word boundary to a hyphen after a syllable boundary. Thus, *rood-achtig* is preferred to *roodach-tig*, and *ont-aarden* to *ontaar-den*.

In addition to sound sequences which form one syntactic word, but consist of more than one phonological word, we also find the inverse situation, sound sequences which are independent syntactic words, but are not independent phonological words. Clitics are examples of such sound sequences. For instance, in Dutch *-ie* is the weak form of the pronoun *hij* 'he'. This clitic always fuses phonologically with the preceding word, as is clear from its behaviour with respect to syllabification, e.g. *Kan-ie komen*? 'Can he come': $(k\alpha)_\sigma$ $(ni)_\sigma$ $(ko)_\sigma$ $(mən)_\sigma$. That is, *kan-ie* is one phonological word. Therefore, Dutch spelling requires the use of a hyphen before *ie*. Note also, that since *-ie* is phonologically dependent on a preceding word, it can never occur at the beginning of a sentence: **Ie komt* 'He comes' is impossible.

4. *Conclusions*

In conclusion, we have seen that Dutch orthography represents a number of levels of linguistic structure. First of all, it represents both phonological and syntactic information. The differences between the phonological and morpho-syntactic structurings of sound sequences appeared to be crucial for a proper account of spacing and hyphenation rules, and made it also possible to interpret the difference between Dutch and English spelling with respect to both spacing and hyphenation conventions, and to interpret the deviations from the spacing norm in the spelling

of compounds by unskilled spellers of Dutch.

Secondly, it appeared that rule typology rather than notions like 'underlying form' and 'phonetic form' is the key to an adequate description of the orthographical conventions of Dutch.

Generally, we have shown that orthography is much more than a representation of 'how it sounds'. By means of abstraction from the effects of certain phonological rules in orthographical representation we achieve graphemic invariance for many morphemes of Dutch, which is undoubtedly convenient from a perceptual point of view, in particular because the skilled reader normally does not read with some phonological form as intermediary between graphemic form and meaning: skilled readers read like Chinese. However, this does not mean that graphemic invariance is an absolute condition for a workable spelling system. Above, we saw that Dutch morphemes ending in /v/ or /z/ have two graphical repesentations. Another source of graphical variation is the lack of a sufficient number of graphemes for the vowels of Dutch: long vowels are represented by the letters of their short counterparts in the following way: in closed syllables one vowel grapheme stands for a short vowel, two identical ones for its long counterpart. In open syllables, one vowel grapheme stands for a long vowel. Consequently, a morpheme like /bɑk/ 'bake' has two graphemic representations, *bak* and *bakk* (as in *bakken* 'to bake', infinitive'). Similarly, /et/ 'eat' has the spelling forms *eet* and *et* (as in *eten* 'to eat, inf.').

Therefore, it is still an open issue which level of phonological abstraction would be the most optimal one for an orthographical system. That is, I have not claimed here that the present Dutch spelling system is the best one that one can think of. The moral to be drawn from the analysis of Dutch orthography given above is that those who think that a more phonetically oriented spelling would by definition be more adequate because such a spelling is the only feasible one from a scientific point of view, are simply wrong.

NOTES

1. See Tauli (1977) for criticism of the position that writing and speech are equivalent, independent media of language.

2. A possible exception is Modern Standard Arabic, which is nobody's mother tongue and is not a spoken language at all (Coulmas 1983: 473).

3. See Smith (1980) for similar conclusions.

4. A general introductory book on Dutch spelling is Booij e.a. (1979).

5. Historically, these suffixes derive from words. Related suffixes are found in several Germanic languages, for instance German and Danish.

6. I do not want to claim that all Dutch prefixes are independent phonological words. Certain prefixes may be considered syllable-appendices to phonological words instead. The crucial point is that prefix boundaries always coincide with syllable boundaries.

BIBLIOGRAPHY

Anderson, S.R. 1974. *The organization of phonology*. New York: Academic Press.

Booij, G.E. 1977. *Dutch morphology. A study of word formation in generative grammar*. Dordrecht: Foris.

Booij, G.E. 1981. *Generative Fonologie van het Nederlands*. Utrecht: Het Spectrum.

Booij, G.E. 1983. "Principles and parameters in prosodic phonology," *Linguistics* 21, 249-80.

Booij, G.E. e.a. 1979. *Spelling*. Groningen: Wolters-Noordhoff.

Coulmas, F. 1983. "Linguistic problems of literacy-introduction," *Journal of pragmatics* 7, 467-77.

Halle, M. and G.N. Clements. 1982. *Problem book in phonology*. Cambridge Mass.: MIT Press.

Hooper, J.B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Kiparsky, P. 1982. "From cyclic phonology to lexical phonology." In: H. van der Hulst and N. Smith (Eds.) *The structure of phonological representations, Part 1*, 131-76. Dordrecht: Foris.

Pike, K.L. 1979. "Universals and phonetic hierarchy." *Proceedings of the 9th Int. Congress of Phonetic Sciences Vol. II*, 48-52.

Smith, P.T. 1980. "Linguistic information in spelling." In: U. Frith (Ed.) *Cognitive processes in spelling*. London: Academic Press, 33-50.

Tauli, V. 1977. "Speech and spelling." In: J. Fishman (Ed.) *Advances in the creation and revision of spelling systems*. The Hague-Paris: Mouton, 17-36.