Lexical Storage and Phonological Change

GEERT BOOIJ

1 Introduction

Empirical investigations of the division of labour between storage and computation in language behaviour can be executed in a number of ways. An obvious one is that of psycholinguistic experimentation. Another way, which will be focused on in this paper, is language change: language change is a psycholinguistic laboratory of nature, a window on how speakers produce and interpret language. For example, if a language loses a phonological rule whereas the effects of that rule are preserved in a number of words, a possible explanation is that the outputs of that rule must have been stored at the stage when the phonological rule was still active, and thus survived after the loss of that rule. The question that I will address in this paper is how far phonological change provides evidence for what kind of phonological information about lexical items is stored in lexical memory. That is, whereas Kiparsky in his early work on phonological change since 1965 (compiled in Kiparsky 1982) focused on phonological change as evidence for the structure of the grammar, I will take a different, extra-grammatical perspective, and ask what we can learn from phonological change about lexical memory. Asking this question is also in line with Kiparsky's conclusion at the end of his book Explanation in Phonology that linguistic change does not provide as directly a window on the structure of the grammar as was hypothesized in Kiparsky (1968a): "Before we can exploit historical evidence for synchronic purposes we need a firm theory of the intervening factors …" (Kiparsky 1982: 234). Lexical storage is certainly one of these intervening factors that deserve more detailed investigation.

* I would like to thank Joan Bybee, Sharon Inkelas, Jaap van Marle, and Leo Wetzels for their comments on an earlier draft of this paper.
A necessary preliminary remark is that the issue of `storage versus computation' with respect to a specific regularity of a language, is not a matter of `either ... or'. The conclusion that a particular linguistic form must be stored in the lexicon, does not preclude the existence of a rule that accounts for most or all of the properties of that form. This position has not been a standard one in generative grammar, which has always been strongly influenced by the Bloomfieldian view of the lexicon as the `basic list of irregularities'. For instance, Kenstowicz (1994: 60) motivates the claim that predictable information is not stored lexically as follows:

(1) Generative grammar's answer to this question is based on the hypothesis that the human capacity for language is designed in such a way as to minimize the amount of information that must be stored in the speaker's mental lexicon."

However, a few pages later in the same book, Kenstowicz (1994: 69-70) points out that this point of view is no longer self-evident: "with the advent of neural science and more accurate estimates of the capacity of the human brain, this "economy of storage" argument is not compelling in and of itself". What I suspect is that the storage argument given in (1) has never been a serious consideration. Rather, the aim has always been to give an elegant and formally as simple as possible analysis of the distribution and alternation patterns of a language.

In the area of generative morphology, linguists have always been aware of the possible and even necessary simultaneity of rules and stored outputs. Many complex words, once formed, must be lexically stored because they have unpredictable formal and/or semantic properties. Nevertheless, the rules that created these complex words may still be productive. Jackendoff (1975) therefore advocated a view of morphological rules in which these rules function as redundancy rules with respect to existing complex lexical items, but this does not exclude that they can also be used creatively for the coining of new complex words. That is, predictable
information can be lexically stored, and redundancy rules tell us which part of that information is predictable, does not count as `independent information'. Similar views were defended by Aronoff (1976) and Booij (1977). In sum, we should avoid what has been called the `rule/list fallacy' by Langacker (1987: 29), i.e. the reasoning in which lists and rules are mutually exclusive.

The relation between lexical storage and morphological rules is even stronger, and of a more principled nature: a morphological rule does not exist without a set of listed words instantiating that rule. The native speaker first acquires complex words. It is on the basis of recurring patterns in sets of similar complex words with a systematic pairing of form and meaning that the speaker may conclude to the existence of a morphological rule, which then may result in extension of the set of words of that particular form.¹

This view of morphology presupposes that words (and idioms) are the units of lexical storage, not morphemes. The role of morphemes in the analysis of word structure is of a secondary nature, they play a role in establishing the relations between words, but are not the primary building blocks of complex words.²

As mentioned above, in generative phonology the prevailing tendency has been to reduce the lexical storage of allomorphy as much as possible. Allomorphy is accounted for in terms of one underlying form for each morpheme, and a set of rules for the computation of the actual surface allomorphs. In addition, predictable phonetic details are omitted from lexical representations: it is only phonemic distinctions, i.e. contrastive phonetic properties that are encoded in lexical representations. The adagium is: what can be computed, should not be stored.³ This relates to the well known abstractness controversy: the more abstract our phonology is, the more we can

¹ A more radical position not taken in this paper is that of connectionism, in which the distinction between rules and representations is denied (cf. Pinker 1999 for discussion of this issue).
² This is the standard view in the Dutch (and more generally, the European) morphological tradition (cf. Schultink 1962, Matthews 1974, Booij 1977, Van Marle 1985). It is also a cornerstone of Bybee’s morphological work (cf. Bybee 1988).
³ Early exceptions are Leben & Robinson (1977), Tiersma (1978), and Leben (1979) who argued that rules such as Trisyllabic Laxing in English are to be seen as parsing rules with primarily a redundancy rule function.
derive different surface forms from the same underlying representation. Linguists have tried to tackle this issue by looking at phonological change (Kiparsky 1968a), whereas psycholinguists have been trying to solve this problem by means of psycholinguistic evidence (cf. Lahiri & Marslen-Wilson 1992). In the beginning of the seventies, Bybee and Vennemann already argued in favour of a concrete view of phonological representations, partially on the basis of facts of language change (Hooper 1974, Vennemann 1974).

Generative phonologists, however, differ in the degree to which they strive for redundancy-free representations, as shown by the debate on underspecification. For instance, in Dutch the velarity of a nasal is contrastive word-finally, but not before a velar obstruent. Should we therefore omit the specification for place in the lexical representation of velar nasals before velar stops? This is not obvious, and Anderson (1985: 136f.) has warned us that the idea that linguistic representations should be redundancy-free is by no means evident, and probably wrong. The principle of ‘lexicon optimization’ advocated by Optimality-Theoretical phonologists also implies that redundant, non-distinctive phonological information will be stored in lexical representations: the faithfulness condition implies that inputs should differ minimally from the corresponding outputs (cf. Archangeli & Langendoen 1997: 201ff).

Although there are differences in the degree of abstractness that generative phonologists allow for, most generative phonologists assume that at least the effect of automatic phonological rules should not be encoded in the lexical representations of morphemes and words, because they are always computable on the basis of purely phonological information. However, we should not take this position for granted given the storage capacity of the human memory.

An additional argument for redundancy-free lexical representations has been that by omitting redundant properties, and assigning these properties at the end of the phonological derivation by means of default rules, the inertness of such properties in phonological derivations follows. However, it has become clear that there is a substantial number of cases in which default or unmarked features such as the feature [coronal] for consonants do play a role in phonological processes, i.e. are not inert (cf. McCarthy & Taub 1992, Booij 1993, Hall 1994). Note, however, that it may be the case that underspecification may be necessary in cases of alternations (Inkelas 1995).
Even if one is willing to accept the storage of the effects of phonological rules, the information stored may still be abstract in the sense that it is phonemes that are stored, and not the actual details of the phonetic realization of these phonemes. This is what most generative phonologists assume, even those who advocate a concrete kind of phonology: the phonetic details, such as the acoustic parameters for a particular vowel are not specified as such in lexical representations, but accounted for by the set of language-specific rules of phonetic implementation. However, this position has recently been attacked, for instance in Flemming (1995) and Bybee (to appear b). This issue will come back in section 5.

In sum, there are three kinds of phoenetically relevant information of which the lexical representation is a point of discussion: (a) predictable phonetic properties that have a contrastive function in some contexts, for instance the place of articulation of nasals; (b) the effects of phonological processes that create alternations; (c) details of phonetic realization.

The point of departure for our discussion is that we know that the vastness of the lexical memory allows us to store much of what can be computed, even though it is completely regular, since it will speed up processing. In the area of morphology, there is solid evidence for the storage of frequent plural nouns in a number of languages (Baayen et al. 1997). The same applies to the syntactic level. As for syntax, there are thousands of noun phrases of the type `adjective + noun' such as yellow pages, red tape, green card, black hole, hard disk, little toe that are listed in our lexical memory because they function as names for things (cf. Jackendoff 1997, Chapter 7); yet this does by no means imply that we do not have a productive rule for the construction of such noun phrases.

Given these general considerations on the possibility of storage, and given that there is no conflict between storage and regularity/productivity, it is important to ask ourselves how we can find out about the actual realities of storage without a priori positions as to what should be
memorized in the form of lexical representations, and what in the form of rules. In this paper, I will use facts of phonological change as one possible type of evidence for storage.

Another general issue that should be addressed before going into the details of phonological change is that it is often taken for granted that the notions `lexical representation' in the sense of `the form of a word which is the starting point of morphological operations' and `stored information about a lexical item' coincide. This is not a logical necessity, however. Take a simple case like the following pair of nouns from Dutch:

(2) *hoed* `hat' [hut], plural *hoeden* [hudχn]

The standard assumption is to assume /hud/ as the lexical representation for *hoed*, and derive the phonetic form of the singular from this underlying form by means of the rule of Auslautverhärtung. However, it is also logically possible that we store [hut] as the singular form, and in addition the `basic form' /hud/, based on the alternation [hut]/[hudχn]. It is even possible that we do not store an underlying form like /hud/ at all, and only compute it from the stored phonetic form if we need to do so because we want to apply a morphological operation to that word. In other words, we then reverse the classical analysis: phonetic forms are stored, and underlying forms are computed. This option will be discussed in more detail below.

2 Phonemization of Allophonic Properties

A number of Germanic languages have seen the loss of the stop in word-final clusters of velar nasal + velar stop. For instance, Middle Dutch *coninc* [ko:nɔ k] `king' became *koning* [ko:nɔ ] in Modern Dutch. Thus, the property of the nasal consonant that it is velar before a velar stop lost its allophonic nature, and became phonemicized: words ending in [ ] now
contrast with words ending in [m] or [n]. Does this phonological change tell us something about the lexical storage of the place of articulation of the nasal before the following stop got lost, and thus about the storage of redundant properties?

In a classical rule-based analysis of phonological change, the answer is negative: the nasal can be underspecified at the underlying level (usually graphically represented as N). Two rules apply: the rule that spreads place of articulation from the stop to the nasal, and subsequently there is a rule of cluster simplification that deletes the stop (the addition of this rule is the change in the grammar of the adult). The surface form of coninc is therefore [ko:νə]. The child will then store this surface form as underlying form, thus giving phonemic status to the velar nasal.

This account of phonological change makes a sharp distinction between the two generations involved in a change, the adult generation, and the next generation that has to acquire the language system. It is doubtful, however, whether such a sharp distinction is justified. It is quite clear, that the adult speaker also continuously subjects the outputs of his language system to reanalysis, i.e. there is a continuous inspection of output forms (cf. Bynon 1977, Hopper & Traugott 1993). The possibility of inspection and reanalysis presupposes that these output forms are stored, have a certain degree of permanence in memory. Moreover, this also enables us to explain phonological change. In the present example, the loss of a final obstruent, the speaker can conclude that the final obstruent can be omitted without the loss of distinction between words, because the place of articulation of the nasal will suffice to distinguish a word like zing [νə] `to sing' from zin [zn] `sense'.

Specification of place of articulation of the nasal in the word-final consonant cluster is also necessary for theory-internal reasons if we take a constraint-based approach to phonology instead of the traditional rule-based approach. The rule-based analysis makes crucial use of rule ordering of a particular type, non-bleeding order, and of the assumption that a phonological
change is to be seen as a rule that is always added at the end of the grammar. If we take a
Optimality-theoretical constraint-based approach, it is immediately clear that the velarity of the
nasal should be part of the lexical representation even before the loss of the final obstruent: an
underspecified representation will induce an extra violation of faithfulness compared to a fully
specified lexical representation, which goes against the principle of Lexicon Optimization. The
surfacing of the velar nasal is then straightforward: the change involved is that the constraint
*CC-Velar (a constraint belonging to the family of constraints on consonant clustering) that
forbids velar consonant clusters is ranked higher than Max-IO (a faithfulness constraint that
requires identity of input and output):

(3)

<table>
<thead>
<tr>
<th>ko:nɔ k</th>
<th>*CC-Vel</th>
<th>Max-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>ko:nɔ k</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>.Logicko:nɔ</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

If an underspecified lexical representation (forbidden by the principle of Lexicon Optimization,
at least if no alternation is involved, cf. Inkelas 1995) had been assumed for the velar nasal, we
would predict the underspecified nasal to surface as coronal [n] since this is the default value for
nasal consonants. (I assume that default values are expressed by markedness constraints such as
Coronal: the default place of articulation of consonants is coronal, and the place of articulation
must be specified, the constraint of Full Specification):
In conclusion, given a constraint-based analysis, the only way in which the velar nasal will surface after the diachronic process of velar cluster simplification is by specifying the place of articulation of this nasal consonant at the level of lexical representation. We should note, however, that this kind of evidence for lexical storage of predictable properties is theory-dependent, since, as we saw above, a rule-based analysis did not imply lexical storage of the place of articulation of the nasal.

### 3 Phonologization and Lexicalization: Vowel Lengthening in Dutch

Another relevant case of phonologization of allophonic properties is that of vowel lengthening in open syllables in Early Middle Dutch. This process affected both simplex nouns and complex nouns, and even word+clitic combinations.

\[(5) \quad \text{a. simplex nouns} \]

| name `name' \hspace{1cm} n[a:]me |
|----------|---------------------------------|
| smake `taste' \hspace{1cm} sm[a:]ke |
| stave `staff' \hspace{1cm} st[a:]ve |
b. singular

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>sch[ø]p `ship'</td>
<td>sch[e:]p-en</td>
</tr>
<tr>
<td>w[e]g `way'</td>
<td>w[e:]g-en</td>
</tr>
<tr>
<td>h[o:]l `hole'</td>
<td>h[o:]l-en</td>
</tr>
<tr>
<td>oorl[o:]g `war'</td>
<td>oorl[o:]gen</td>
</tr>
<tr>
<td>d[a:]g `day'</td>
<td>d[a:]g-en</td>
</tr>
</tbody>
</table>

c. word + clitic

saetic /s?t o skł/ [sa:tɔk] `sat I'

The singular-plural pairs in (5b) have been preserved in modern Dutch (cf. Booij 1995, Kager, this volume). This process of vowel lengthening is a manifestation of a much more general tendency in Germanic languages that stressed syllables must be heavy, i.e. cannot end in a short vowel, but contain minimally either a long vowel or a short vowel followed by a consonant (also known as Prokosch' Law).

In the first stage of this process in Early Middle Dutch (cf. Van Loon 1986: 89), the lengthening of the vowel was an allophonic process. In the course of time (probably in the 12th century), however, the lengthened vowel was phonemicized, i.e. became an underlying long vowel. How do we know this?

The first evidence concerns short vowels followed by geminate consonants. These were protected from lengthening because the first half of the geminate closes the syllable. In the 12th century, degemination took place. Yet, the short vowels that thus ended up in open syllables, did not lengthen anymore, and so we find many words with short vowels followed by only one consonant such as wikke [ɔkχ] `vetch'. That is, the following development took place:

---

5 See also Gussenhoven (this volume) and Kager (this volume) about vowel length in Dutch.
These changes are illustrated by the following minimal pair (Van Loon 1986: 89):

(7) Early Germanic   Early Middle Dutch   Modern Dutch
    wika [wɔka] ‘week’               weke [we:kχ]       week [we:k]
    wikkia [wɔkkia] ‘vetch’         wikke [wɔkkχ]   wik [wɔk]

The classical generative interpretation of this kind of phonemicization is as follows (Kiparsky 1968b): the initial change is the addition of a rule of degemination to the phonological system of the adult speakers. This rule of degemination made the alternation between long and short vowels in the output forms opaque: vowel lengthening is ordered before degemination, and this is a non-bleeding order. The next generation of language users therefore interpreted the length contrast as a phonemic contrast, and hence the length contrast became part of the lexical representation. In addition, the process of vowel lengthening in open syllables disappeared.

A second kind of evidence for the phonemicization of these initially allophonic contrasts for simplex nouns is provided by the later process of schwa apocope (13th century). This phonological change did not affect the length of the vowel in the simplex nouns ending in schwa: after apocope the vowels of words such as naam ‘name’ (<name) and smaak ‘taste’ (<smake) stayed long. Again, vowel lengthening became opaque, and the next generation had to
store the vowel length as part of the lexical representation. The phonemicization process even led to doublets such as *bar* `raw' - *baar* `ready (money) and *staf* `staff' - *staaf* `bar', a process that can be expected given the phonemicization of vowel length.

This account in terms of rule opacity does not explain, however, why the alternation between long and short vowels was kept in certain cases such as those in the singular-plural pairs in (5b) above. When vowel lengthening disappeared, the length alternation should also have disappeared. The fact that it was kept for a number of words shows that in these cases the plural forms must have been stored in their phonetic form, with a long vowel in the stressed syllable, at a time when the alternation was still governed by a phonological rule. In addition, it also shows that plural forms of nouns can be lexically stored, a conclusion that is confirmed by recent psycholinguistic evidence, as I already mentioned above (cf. Baayen et al. 1997).

The traditional generative account of this change (Kiparsky 1968b) would read as follows. Each new generation of speakers has to discover the rules of its native language on the basis of the phonetic outputs encountered. First, they will simply store the phonetic output forms of words. Once they have discovered the phonological regularities involved, they will replace each set of related phonetic output forms with one common underlying phonological form (so-called restructuring). If, for some reason, a particular phonological regularity (such as vowel lengthening in open syllables, the process discussed here) is not grasped by the new generation, there will be no restructuring: the stored phonetic forms will remain memorized as is, and if new words or word forms are coined, they will not be affected by that rule. Hence, the rule disappears, and only those words for which the different surface form had already been stored in lexical memory will keep that alternation.

However, this reasoning clearly presupposes that output forms are stored in the initial stages of acquisition. The question then is, why native speakers would change, and even erase that information about output forms in a later stage of acquisition, once they have discovered
the phonological regularity involved. Given the vastness of human memory, there is no reason not to keep that information in lexical memory, and to use the phonological rule that can be discovered on the basis of related output forms for the computation of the phonetic shape of new words and word forms only. Thus, we get a natural division of labour between storage and computation: output forms, once heard, may be stored, and rules, once discovered, are used to interpret phonetically similar forms as instantiations of the same morpheme, and for the computation of the phonetic forms of new words.

For those words where there was no alternation involved, such as the words *weke* versus *wikke* given in (7) the assumption of storage in memory of the output forms by adult speakers is also necessary. Only if we assume that predictable length differences are stored, can we explain why the rule of degemination was added to the adult grammar: the addition of degemination has been made possible by the length difference, because the latter could now serve to keep words phonetically distinct. The language user is able to survey and inspect the phonetic output forms of words in his memory, and can thus conclude that the distinction between single and geminate consonants is redundant given a concomitant distinction in vowel length. The emergence of degemination can therefore be seen as the result of storage in speakers’ memory of output forms. Similarly, the possibility of schwa apocope may also be seen as the result of storage of the phonetic output forms: since the vowel length distinction was kept in memory, the word-final schwa could disappear. This then made it necessary to consider vowel length as information that is part of the lexical representation.

Given this analysis, we avoid constructing a complete gap between two generations of language speakers, and are not forced to assume that language change is only caused by the imperfect learning of the next generation that has to acquire the language on the basis of outputs. Language change is also an effect of the adult language user who is able to analyze the phonetic forms of words which are accessible in his memory.
The consequence of the loss of transparency of a phonological process (rule opacity) is that it is no longer possible to compute or store an underlying form for a word that differs from its phonetic form (unless an alternation is involved). A distance between phonetic form and underlying form is only possible in the case of surface-true transparent processes. In the case at hand this means that the plural forms of these nouns, with their long vowels, must be stored, also for new generations, since their phonetic forms cannot be computed. This is also the conclusion of Kager (this volume).

Since in Dutch the singular noun is identical to the stem that is the basis for morphological operations, it is the short vowel that appears in such cases:

(8)  \[\begin{array}{lll}
\text{singular noun} & \text{plural noun} & \text{derived word} \\
gebr[ε]k \ 'handicap' & \text{gebr[ε]ken} & \text{gebr[ε]kkig \ 'handicapped'} \\
h[\text{][}] \ 'hole' & h[o:]len & h[\text{][}]lletje \ 'diminutive' \\
sch[øp] \ 'ship' & \text{sch[e:]}pen & \text{sch[ø]pper \ 'skipper'} \\
g[\text{][}d] \ 'god' & g[ø:den] & g[\text{][}]ddelijk \ 'divine' \\
sp[ε]l \ 'game' & \text{sp[e:]}len & \text{sp[ε]lletje \ 'diminutive'} \\
w[ε]g \ 'road' & w[e:gen] & w[ε]ggetje \ 'diminutive'
\end{array}\]

In a number of cases the alternation that existed between singular and plural nouns disappeared after the loss of the rule of vowel lengthening (so-called analogical or paradigmatic leveling). This applies to, for instance, the following Dutch nouns which have a short vowel in their last syllable in both singular and plural forms:

(9)  \[\begin{array}{ll}
\text{singular noun} & \text{plural noun} \\
bissch[\text{][}]p \ 'bishop' & \text{bissch[\text{][}]ppen}
\end{array}\]
The traditional interpretation of these facts is that since the plural forms of these words were not stored, the regular forms with short vowels will show up after the loss of the rule of vowel lengthening. However, since the position is taken here that plural forms can be stored in their output form, the following interpretation is called for: the original plural forms of these words with long vowels did have a lexical representation of their own, but their frequencies, and thus their resting level of activation was not high enough to block the formation of a regular plural form without vowel lengthening, i.e. on the basis of the underlying form of the singular noun. This is also the point of view defended in Wetzels (1981: 95-97; 1984: 595) as to analogical leveling: analogical leveling is nothing else but the application of productive rules that are not blocked by the existence of stored forms.

4 Auslautverhärtung and Lexical Storage

So far, our findings with respect to the lexicalization of vowel lengthening are in line with the conclusion of Wetzels mentioned above. However, he argued that paradigmatic leveling only takes place if rules are no longer automatic phonological rules, and hence require storage of allomorphs. Only if rules are no longer phonologically conditioned will the different allomorphs of a morpheme be lexically stored, which then may give rise to leveling by using the `wrong' allomorph. Since, according to Wetzels, the outputs of automatic phonological rules are not stored, they will not lead to paradigmatic leveling. For instance, the alternations created by the automatic rule of devoicing of obstruents in coda position have, with a few exceptions to be
discussed below, not been leveled in Dutch. Therefore, we will now have a more detailed look at coda devoicing.

It is a uncontroversial assumption within mainstream generative phonology that the effects of automatic neutralisation rules are not encoded in lexical representations. For instance, Kenstowicz & Kisseberth (1979: 49) argue on the basis of the Russian rule of word-final devoicing of obstruents that it would be wrong to list both the allomorph with a voiced obstruent and the one with a voiceless obstruent in the lexicon, and to assume a selection rule for the allomorphs stated as follows:

(10) "If a morpheme has alternants that differ with respect to the voicing of a final obstruent, select the alternate with a final voiceless obstruent when the morpheme appears at the end of a word; otherwise, select the alternant with a voiced obstruent."

The authors then add that the "basic criticism is that this sort of analysis fails to adequately characterize the rule-governed nature of the voicing alternation in Russian" (p. 49). That is, such an analysis implies that two rules have to be assumed for Russian, a phonological rule of devoicing, and a morphological rule of allomorph selection.

Language change again provides a window on this issue: if we do not lexically store the outputs of a neutralisation rule (with the effect that each output is a possible underlying form), we predict that paradigmatic leveling will not take place. I will therefore have a look at a process similar to Russian devoicing, coda devoicing of obstruents in Dutch. Final devoicing in Dutch is a productive generalization: voiced obstruents cannot occur in codas. If there are two allomorphs for a lexical item with an alternation between a voiced and a voiceless obstruent, it is generally assumed that the underlying form that is stored, is the form that ends in a voiced
obstruent. If this voiced obstruent ends up in coda position, it is devoiced by the rule of final devoicing:

(11) rib [röp] / ribben [röbχn] `rib, sg/pl'

    hoed [hut] / hoeden [hudχn] `hat, sg/pl'

    slaaf [slaːf] / slaven [slaːvχn] `slave, sg/pl'

    kiez [kis] / kiezen [kizχn] `molar, sg/pl'

    vlieg [vliːx] / vliegen [vliːpχn] `fly, sg/pl'

Interestingly, in some cases the alternation is lost. For instance, although we have the alternation hand/handen `hand, sg/pl' [h̩mt]/[h̩ndχn], there is an idiomatic phrase bijdehand [bœidhɔnt] `lit. at the hand, bright' that functions as an adjective. One of its inflected forms in prenominal position is bijdehante [bœidhɔntε] i.e. there is no alternation between voiced and voiceless obstruent anymore.

In Van Loey (1964: 54) the rise of forms such as bijdehante is interpreted as a case of paradigmatic leveling (analogy), which would suggest that the allomorph [h̩mt] of hand is stored. However, the problem for this account is that it does not explain why paradigmatic leveling almost never takes place with respect to effects of final devoicing. Another interpretation is therefore called for: the language learner only computes an underlying form that differs from the phonetic one if (i) the two words involved are related, and (ii) the two surface forms are relatable by means of a transparent rule. In the case of bijdehand, with its inflected form bijdehante, this adjective has no transparent formal relation to the word hand `hand’ with its plural form handen, and thus it will be stored with a final /t/. In other words, we can only get underlying forms that differ from the surface forms if there is a transparent morphological relation between the words in which the different allomorphs occur (cf. Wetzels
Thus, the form *bijdehante* is not a counterexample to Wetzels’ (1981) claim that automatic phonological rules do not lead to leveling.

Another interesting case is the word *stad* [st\textipa{d}] `town’ with the irregular plural form *sten* [ste:d\textipa{n}]. The plural noun not only exhibits vowel lengthening, but also vowel quality change, from /a:/ to /e:/.

Consequently, the underlying form of *stad* is apparently computed as /st\textipa{d}/, as shown by new coinings such as the verb *statten* `to do shopping in town’ derived through conversion form the noun *stad* `city’, and the inhabitant name *Lelystatter* derived from the toponym *Lelystad*.

A similar phenomenon can be seen in Afrikaans for a number of nouns that have plural forms with lengthened vowels in Dutch. Apparently, the distance between *smid* `blacksmith’ and *smeden* (plural), for instance, is too large for speakers of Afrikaans, and thus they do not conclude to a common stem /sm\textipa{d}/. Thus, the lexical form of *smid* will be /sm\textipa{t}/ rather than /sm\textipa{d}/, and consequently the regular plural *smiten* has arisen.

(12) 

<table>
<thead>
<tr>
<th>Dutch</th>
<th>Afrikaans</th>
</tr>
</thead>
<tbody>
<tr>
<td>smid [sm\textipa{t}] /smeden [sme:d\textipa{n}]</td>
<td>smit [sm\textipa{t}] /smiten [sm\textipa{t}\textipa{n}] `smith, sg/pl’</td>
</tr>
<tr>
<td>god [\textipa{t}] / goden [\textipa{t}\textipa{n}]</td>
<td>god [\textipa{t}\textipa{t}] / gotten [\textipa{t}\textipa{n}] `god, sg/pl’</td>
</tr>
<tr>
<td>lid [l\textipa{t}] / leden [le:d\textipa{n}]</td>
<td>lit [l\textipa{t}] / litten [l\textipa{t}\textipa{n}] `member, sg/pl’</td>
</tr>
<tr>
<td>rad [r\textipa{t}] / raderen [ra:d\textipa{r}\textipa{n}]</td>
<td>rat [r\textipa{t}] / ratten [r\textipa{t}\textipa{n}] `wheel, sg/pl’</td>
</tr>
</tbody>
</table>

The point here is that precisely the nouns with the unproductive vowel length alternation exhibit paradigmatic leveling with respect to the still transparent rule of coda devoicing.

In the case of the words with the vowel length alternation discussed above, there is also internal evidence for the storage of the effect of coda devoicing in the lexical entry, because...
vowel length plays a role in the distribution of fricative consonants. The generalization is that /v/ and /z/ only occur after long vowels, the \( v/z \)-constraint.\(^6\) Interestingly, we find the following alternations

\[(13)\] *nouns*

- **graf** [\(\pi\Psi f\)] \(\rightarrow\) **graven** [\(\pi ra:v\chi n\)] \(\sim\) `grave, sg/pl'
- **hof** [\(hjf\)] \(\rightarrow\) **hoven** [\(ho:v\chi n\)] \(\sim\) `court, sg/pl'
- **glas** [\(\pi l\Psi s\)] \(\rightarrow\) **glazen** [\(\pi la:z\chi n\)] \(\sim\) `glass, sg/pl'
- **staf** [\(st\Psi f\)] \(\rightarrow\) **staven** [\(sta:v\chi n\)] \(\sim\) `staff, sg/pl'
- **verlof** [\(verl\]f] \(\rightarrow\) **verloven** [\(verlo:v\chi n\)] \(\sim\) `permission, sg/pl'

*verbs, past tense, sg/pl*

- **las** [\(l\Psi s\)] \(\rightarrow\) **lazen** [\(la:z\chi n\)] \(\sim\) `read' 
- **genas** [\(\pi \chi n\Psi s\)] \(\rightarrow\) **genazen** [\(\pi cha:z\chi n\)] \(\sim\) `cured'
- **gaf** [\(\pi \Psi f\)] \(\rightarrow\) **gaven** [\(\pi a:v\chi n\)] \(\sim\) `give'
- **vergaf** [\(ver\pi \Psi f\)] \(\rightarrow\) **vergaven** [\(ver\pi a:v\chi n\)] \(\sim\) `forgave'

*noun-verb pairs*

- **draf** [\(dr\Psi f\)] `trot' \(\rightarrow\) **draven** [\(dra:v\chi n\)] `to trot'
- **lof** [\(l\]f] `praise' \(\rightarrow\) **loven** [\(lo:v\chi n\)] `to praise'

\(^6\)There are four exceptions: the loan words *mazzel* ‘good luck’, *puzzel* ‘puzzle’, *razzia* ‘idem’, and the name for railway kiosks *wizzl*, a form intentionally coined as a marked form with a high attention value.
In all these cases, short vowels are followed by a voiceless fricative, whereas long vowels are followed by a voiced fricative. In other words, all allomorphs obey the constraint on the distribution of /v/ and /z/. In these cases, the voicelessness of the final obstruent is also part of the underlying form, although this information is predictable, and although these obstruents alternate with voiced ones. We know this on the basis of the forms of complex words derived from such nouns: these words are derived either from a form with short vowel + voiceless obstruent, or from a form with long vowel + voiced obstruent; the combination short vowel + voiced obstruent does not occur:

(14)  

**denominal word formation**

---

gl[a:]zenier `stained glass artist'  *glazzenier

h[f]elijk `polite'    h[o:]veling `courtier'

*hovveling

best[Y]fjing `staff'    st[a:]ven `to prove'  *bestavving

l[f]elijk `praiseworthy'    l[o:]ven `to praise  *lovvelijk

On the other hand, if the noun ends in a stop or a velar fricative, this restriction does not apply, and morphology can apply to underling forms with a voiced stop:

(15)  

**singular**  **plural**  **derived word**

god  goden  g[ ][ ]ddelijk `divine'

bad  baden  b[Y]dden `to bathe'

weg  wegen  w[ ]getje `small road'
The $v/z$-constraint thus restricts the distance between phonetic form and underlying form.

Van Loey (1964: 54-55) also mentions a number of cases in which leveling in favour of the allomorph with voiceless obstruent took place, where no additional alternations are involved that make the two allomorphs different in other respects as well. Examples from 17th century Dutch are:

(16)  *Middle Dutch* gewaet `dress' / gewaden `pl.' > *17th century Dutch* gewaeten

*Middle Dutch* cieraet `ornament' / cieraden `pl.' > *17th century Dutch* cieraeten

In these cases the allomorphs with final [d] have been restored in modern Dutch.

There are also cases in present-day Dutch of this kind of leveling. An example is the adjective *boud* `bold'. Originally, the inflected form of this adjective is *boude*, but most speakers of Dutch say *boute*, which shows that the alternation d/t has been leveled out. What these examples show is that leveling in favour of the allomorph with voiceless obstruent is possible. This kind of leveling can be due to imperfect learning: the language user did not yet compute the correct underlying form of the noun on the basis of the alternation between singular and plural form, although the alternation is transparent. Therefore, the underlying form is identical to the phonetic form of the singular noun.

We may thus hypothesize that we store phonetic forms and that underlying forms are computed on the bases of stored phonetic forms (a similar position is taken in Leben & Robinson 1977 and Leben 1979). Underlying forms will only differ from phonetic forms in so far that difference follows from transparent phonological rules. If the distance between two forms is too big either semantically (the case of *bijdehand*) or because the formal differences are too big (*stad-steden* and the Afrikaans cases), the underlying forms will not differ from the surface forms. A structural constraint such as the Dutch $v/z$-constraint will also block the
computation of an underlying form that is different from the surface form of a word, even though there is an automatic alternation involved.

5 Lexical Diffusion

In the early stages of generative phonology, phonological change was seen as the addition of a rule at the end of the grammar. That is, the lexical representations do not change, but the corresponding surface form is affected by the application of the added rule. It is only the new generation that will have different lexical representations with the effect of the added rule lexically encoded, except when it is a rule that creates alternations. In the latter case, lexical representations will not necessarily change (Kiparsky 1968b).

A problem for this view is that phonological changes in progress often affect lexical representations but are also simultaneously the source of phonological alternations. For instance, Dutch is subject to a process of /d/-weakening in which intervocally, /d/ is replaced with the glide /j/ before a following, suffix-initial schwa (Booij 1995: 90). This causes alternations of the following type:

(17)  

\[
\begin{align*}
good \text{'good'}, \text{ inflected form } & \text{goed-e, phonetic form } [\pi\text{ud}x] \text{ or } [\pi\text{uj}x] \\
\text{rood} \text{'red'}, \text{ inflected form } & \text{rod-e, phonetic form } [\text{ro:d}x] \text{ or } [\text{ro:j}x]
\end{align*}
\]

Application of this allomorphy creating rule of /d/-weakening is lexically governed. For instance, it does not apply to the plural form of \textit{hoed-en} `hats' [hud\textit{x}n]: the phonetic form [hu\textit{j}x\textit{n}] is impossible. Weakening also applies optionally to the derived adjective \textit{goed-ig} `good-natured' [\textit{pud}\textit{d}x\textit{x}] or [\textit{puj}x\textit{x}], but obligatorily to the derived noun \textit{goeierd} [\textit{puj}z\textit{rt}] `good-natured person': the form [\textit{pud}z\textit{rt}] is impossible. But other adjectives with stem-final /d/ do not
always allow weakening; the adjective *wreed* `cruel', for instance, does not have the inflected form [wre:jχ] for *wrede*, the only form possible is [wre:dχ]. Similar observations apply to processes such as */d/-deletion in Dutch (Booij 1995: 90). The lexically governed nature of these alternations implies that they must be stored lexically, even in the case of morphologically completely regular inflected forms of adjectives such as *goed-e*.

The facts of *d*-weakening support Kiparsky's (1988, 1995) claim that it is lexical, i.e. neutralizing rules that exhibit lexical diffusion, since the distinction between */d/ and */j/* is phonemic in Dutch. What is essential from the perspective of this paper is that they support the view that information provided by rule and information that is lexically stored are not exclusive with respect to each other.

These allomorphy facts imply that the recognition system of the language user must have a certain robustness because (s)he has no problems in relating *goed* to *goeie* and *goeierd* notwithstanding the phonological differences in the lexical representation of the shared part of these words, the lexical morpheme *goed*. That is, apparently the phonetic forms of a morpheme do not have to be computed by rule from a common underlying form in order to be recognizable as allomorphs of that morpheme (contra Lahiri & Marslen-Wilson 1992).

In the case of */d/-deletion an intervocalic */d/ is deleted, and replaced with a predictable hiatus filling glide. For instance, the inflected adjective *oud-e* `old' can be pronounced as [judχ] or [ju χ]. Both forms must be stored because it cannot be predicted which forms allows for *d*-deletion. The phonetic difference even correlates with a semantic difference in the case of the *A+N* phrases *oude hoer* [judχ hu:r] `old whore' (the literal meaning) versus *ouwe hoer* [ju β hu:r] `talkative person'.

Labov (1981, 1994) proposed to distinguish two types of phonological change: change that is phonetically gradual, and affects all relevant words, and change that is phonetically abrupt, replaces a phoneme with another one, and is lexically gradual, i.e. exhibits lexical
diffusion. Kiparsky (1988) argued that the distinction between phonetically gradual and phonetically abrupt changes coincides with the distinction between postlexical and lexical phonological rules. The rules of /d/-weakening and /d/-deletion can indeed be considered lexical rules since they are neutralizing. As expected, they have exceptions, and thus exhibit lexical diffusion. Lexical diffusion always creates surface opacity for rules since the speaker will find forms that have not undergone the rule. Therefore, as stated above, opacity will lead to lexical storage in the sense that for each phonetic form of such words a distinct lexical entry has to be created. This in its turn explains why semantic distinctions may correlate with phonological differences, as in the pair *oude hoer* / *ouwe hoer* discussed above.

What about the effects of postlexical rules? The usual assumption (cf. Kiparsky 1988: 399) is that the effects of such rules are not encoded in lexical representations. In recent work, Bybee (to appear a) has questioned these assumptions. She argues that a change may be both phonetically and lexically gradual. This implies that the effects of such gradual changes must be lexically stored, in the lexical representations of individual words.

Vowel reduction in Dutch (Booij 1982, 1995; Van Bergem 1995) is a potentially interesting phenomenon in this respect because it has both lexical and postlexical properties. The basic generalization is that at the phonetic level all unstressed vowels are affected by reduction in the sense that their actual realisation can be quite far away from the target values of the acoustic parameters of these vowels (Van Bergem 1995). In addition, there is a lexical process of vowel reduction of vowels in unstressed (non-word-final) syllables that has already affected many words whose full vowels have been replaced with schwa, for instance:

(18) televisie  `television`
    álgebra   `algebra`
    serenáde `serenade`
recláme `publicity'
betón `concrete'
repetítie `rehearsal, test'
conferéntie `conference'

Moreover, there is also a large number of words that exhibit vowel reduction in more casual speech only, i.e. there is still alternation:

(19) banáan `banana' [baːnaː] / [bχnaːn]
polítie `police' [poːliː] / [pχliː]
minúut `minute' [minyt] / [mχnyt]

These facts of vowel reduction nicely fit into Kiparsky's two-stage theory of phonological change: "… the phonetic variation inherent in speech, which is blind in the neogrammarian sense, is selectively integrated into the linguistic system and passed on to successive generations of speakers through language acquisition" (Kiparsky 1995: 642). Vowel reduction, originally a purely phonetic process motivated by ease of articulation, could become a lexical rule, subject to lexical diffusion, because the schwa is a phoneme. As a lexical rule, it can be seen as a process in which the place of articulation features of vowels in unstressed syllables are removed from the lexical representations of words; subsequently, a default rule will fill in the relevant features of the schwa (cf. Kiparsky 1995: 642-47 for this interpretation of lexical diffusion). Simultaneously, vowel reduction is a postlexical rule applying to vowels in syllables without lexical stress, and a process of phonetic implementation for vowels (even those with lexical stress) that are not stressed in a particular utterance.
The alternation between full vowel and schwa is also visible in related pairs of words that differ in whether the relevant vowel is stressed or not; if the vowel is unstressed it can or must reduce:

(20)  

(a) persóon `person'  pèrsonéel `staff'  
percént `procent'  pèrcentáge `id.'  
pastóor `priest'  pàstoráal `pastoral'

(b) proféet `prophet'  pròfetéer `to prophesy'  
juwéel `juwel'  jùwelier `jeweler'  
gène `embarrassment'  genánt `embarrassing'

The words in (20a) exhibit optional reduction, whereas in those in (20b) the vowel in italic can only be realized as schwa. That is, the words in (20a) are subject to the postlexical rule of vowel reduction whereas the words in (20b) exhibit lexical diffusion effects. This implies that a morpheme like juweel will have two different lexical representations, /jy eːl/ when it is an independent word, and /jy ɣl/ when occurring in juwelier. This does not cause any computational problem since words are the units of storage, and thus the two allomorphs will automatically appear in the right context. The only problem that the language user has to solve here, is the recognition problem: how can both forms be recognized as forms of the same morpheme (a prerequisite for the semantic analysis)? But this is part and parcel of commanding a language since allomorphy that is not reducible to one underlying form is a widespread phenomenon, as we saw above (cf. also Booij 1997a,b; 1998).

Van Bergem (1995) rightly qualifies this process of vowel reduction as a sound change in progress, which came into being due to interpretation by the native speaker of acoustic vowel reduction as a process of replacement of full vowel with schwa, and exhibits lexical diffusion,
since in some words the unstressed vowel can only be realized as a schwa, as we saw above. Furthermore, this kind of reduction is boosted by high frequency: in high frequency words unstressed vowels are reduced more easily and more frequently than in low frequency words. This is to be expected since vowel reduction reduces lexical contrasts, and thus impedes word recognition. High frequency, on the other hand, boosts recognition, and can thus compensate for the negative effects of vowel reduction.

The question then arises if the postlexical rule of optional vowel reduction in unstressed syllables is phonetically gradual, or should be interpreted as the replacement in lexical representations of the full vowel with schwa by means of a phonological rule, just as in the case of the lexical rule of vowel reduction. If the first position is taken, this may have the further implication that the degree of reduction of the vowel is encoded in lexical representation. This position is argued for in Bybee (to appear a,b) for some other cases of reduction, and implies that details of the phonetic realization of phonemes are lexically specified. Can we say something about this on the basis of Dutch vowel reduction?

The data in Van Bergem (1995: 121) show that there is a positive correlation between frequency and acoustic parameters: in high frequency words the formant values for the unstressed vowels are farther away from the target values than in low frequency words. Moreover, this correlates with how native speakers perceive these vowels: "the average number of schwa responses increases when the spectral distance between the test vowel and its target increases" (Van Bergem 1995: 125). These results are explained if the growing distance between the acoustic parameters of unstressed vowels in words of high frequency and those of the non-reduced correlates is lexically encoded: each time a vowel is reduced, its formant values move away from those of the unreduced vowel. This is the kind of explanation advocated in Bybee (to appear, a). The theoretical implication of this step is that lexical representations are
not redundancy-free as far as phonetic details are concerned, a position also defended by Ohala & Ohala (1995).

This position as to the lexical specification of vowel reduction effects is supported by the observation that data concerning optional vowel reduction can be obtained by means of introspection: native speakers of Dutch systematically know that certain words are more susceptible to vowel reduction than other. For instance, they know that the /i/ in minuut is susceptible to reduction, whereas the /i/ in pilóot `pilot' is not. This suggests that such information is stored in memory.

The conclusion to be drawn is that phonological rules may be productive and automatic, and yet at least some of their outputs are lexically stored because the process exhibits lexical diffusion. In addition, it seems that phonetic details concerning the pronunciation of vowels can be stored, given the facts concerning the gradual erosion of unstressed vowels. Note, however, that this kind of stored information does not necessarily lead to a proliferation of distinct underlying forms (in the sense of 'bases for morphological operations') for a morpheme. This is a crucial difference between lexical and postlexical rules: it is only the effects of lexical rules that may lead to more than one underlying form for a particular morpheme.

6 Conclusions

In this paper, we saw that the standard view in generative phonology of the balance between storage and computation has to be reconsidered: there is a wealth of evidence for the position that predictable information is stored in the lexicon. First, recent theoretical developments in phonology imply that predictable information about morphemes must nevertheless be stored in the lexicon. Secondly, data concerning phonological change show that computable information concerning the phonetic realization of morphemes nevertheles has to be stored lexically. I also
proposed that we should take a radical step with respect to the relation between underlying form and phonetic form: it is not the phonetic form that is computed by the speaker, but rather the underlying form. Like storage in general, storage of phonetic forms of words will speed up processing; it is only when we coin a new word that computation of the underlying form of the base word is necessary.

These conclusions do not refute the position that the human language faculty has a dual structure: a lexicon with stored representations and rules. The native speaker does need rules for the perception and production of novel forms. What, however, these conclusions do refute is the position that computation and storage of information with respect to the same process or regularity are mutually exclusive.

7 References


Gussenhoven, C. (this volume) Vowel Duration, Syllable Quantity and Stress in Dutch.


Kager, R. (this volume). Lexical Irregularity and the Typology of Contrast..


