MORPHON: Lexicon-based text-to-phoneme conversion and phonological rules

Anneke M. Nunn—Vincent J. van Heuven

Abstract

In this contribution MORPHON is outlined. This module provides the text-to-speech system with phonological rules. It will be argued that such rules are needed because the pronunciation of a sentence does not consist of the concatenation of the pronunciation of the constituting morphemes, but the pronunciation of morphemes is modified in certain contexts. These rules can only apply properly if exceptions can be listed in a lexicon, and if rules can refer to morphological and morpho-syntactic information. Therefore a lexicon-based approach to text-to-phoneme transcription conversion was chosen. Finally, the pronunciation accuracy of MORPHON is compared with that of two rule based text-to-phoneme transcription systems.

1. Introduction

An important step in text-to-speech conversion is the generation of the correct pronunciation representation on the basis of the input text. This requires a mechanism to convert spelled words of the input sentence into a phoneme representation. We will argue that these phoneme representations should be entered in a morpheme lexicon rather than be derived by rule.

Let us first note that the pronunciation of a sentence does not consist of the concatenation of the pronunciation of the individual words since the pronunciation of those words can be modified in certain contexts. As an example, consider the pronunciation of the sentences (1) and (2):

(1) a. De onderzoeker vond de verschillen significant.
     b. The researcher found the differences significant.
     c. də ʔɔn-dər-ˈzuˑ-kər ˈvɔnt də vər-ˈsχɪ-lən sɪχ-niˑ-fiˑ-ˈkɑnt

(2) a. De onderzoeker vond een significant verschil.
     b. The researcher found a significant difference.
     c. də ʔɔn-dər-ˈzuˑ-kər ˈvɔnt ən ˈsɪχ-niˑ-fiˑ-kɑnt fər-ˈsχɪl

One pronounces the first consonant of *verschillen* as a voiced [v] in (1) but in (2) it is changed to a voiceless [f] under the influence of the voiceless final obstruent of the preceding word. The word *significant* is spoken with final stress in (1) but in (2) the stress may be shifted to the first syllable due to a clash with the stressed syllable in *verschil*.

Generative phonology has provided us with rules that account for these processes. The formalization of such phonological rules on the word level has been the main object of this project, sentence prosody is accounted for in another module (cf. Dirksen—Quené, this volume). The name of the resulting rule set is MORPHON (it has MORphemes as its input and supplies them with their surface PHONeme representation). In this contribution we give a brief description of MORPHON, for a more detailed account, see Nunn (1991).

The organization of this contribution is as follows: In section 2 and 3 we will discuss the two stages of the derivation of the pronunciation, viz. spelling to underlying phoneme transcription and underlying to surface phoneme transcription conversion in more detail. In section 4 the role of the syntactic category of words for the proper application of the rules is discussed, and section 5 gives a comparison of MORPA-cum-MORPHON with two existing text-to-phoneme representation conversion modules for Dutch.

2. From spelling to underlying phoneme representation

Phoneme transcriptions can not be derived directly from orthographic input, for there is no one-to-one mapping between graphemes and phonemes in Dutch (Wester 1985; Berendsen—Langeweg—van Leeuwen 1986). Therefore phoneme transcriptions must be derived by one of the following two methods.

The first method consists in the application of a set of rules that map graphemes or grapheme combinations to phonemes. We will refer to this method as rule-based conversion. The second method consists of looking up the pronunciation of words in a lexicon (lexicon-based conversion). In MITalk, an English text-to-speech system (cf. Allen—Hunnicutt—Klatt 1987) the lexicon-based method was chosen, because a lexicon is needed anyway. The lexicon is necessary to analyze words for the purpose of pronunciation as grapheme-to-phoneme conversion rules do not work across morpheme boundaries, and to account for numerous exceptions to grapheme-to-phoneme conversion rules. Therefore it is elegant to use this lexicon for letter-to-sound conversion as well.

The same argument can be made for Dutch: grapheme-to-phoneme conversion rules do not work across morpheme boundaries in Dutch either. This has even more consequences in Dutch, as compounds are written as

one word. For instance, the rule that turns the grapheme sequence *sch* into the phonemes [sχ] should not work in a compound like *afdelingschef* 'departmental chief' because that would yield the incorrect pronunciation [ɑfdeːlɪŋsχɛf] instead of [ɑfdeːlɪŋʃɛf]. As in English, there are many exceptions to Dutch grapheme-to-phoneme conversion rules and some exceptions involve highly frequent words.

Moreover, just like letter-to-sound conversion rules, phonological rules as discussed above both are sensitive to morphological boundaries and subject to numerous exceptions as well. To illustrate the sensitivity of phonological rules to morphological boundaries, consider the following two words:

maandag	[maːn#dɑχ]	'monday'
maandabonnement	[maːnt#abɔnəmɛnt]	'monthly subscription'

Both words begin with the string *maand* but the /d/ is changed into [t] in the second word, where it is the final consonant of the first morpheme and not in the first word where it is the first consonant of the second morpheme. This can be explained as follows: the rule that affects the /d/, final devoicing, is sensitive to syllable structure and syllabification is sensitive to morphological structure of a word: compound boundaries are also syllable boundaries.

The main source of exceptions for languages such as Dutch is (word) stress assignment. As noted by, for instance, Langeweg (1988) and Kager (1989), Dutch word stress is only predictable to a certain extent (from syllable weight). According to Langeweg, in ninety percent of the monomorphemic words the stress can be accounted for by rules, the remaining ten percent show idiosyncratic behavior. For instance, words consisting of three open syllables regularly have penultimate stress but two deviating stress patterns occur, namely antepenultimate stress and final stress:[1]

[koː-ˈloː-niˑ]	'colony'
[ˈkoː-liˑ-briˑ]	'hummingbird'
[proː-zoː-ˈdiˑ]	'prosody'

Therefore a lexicon is necessary and we will use it for text-to-phoneme transcription conversion. In section 5 we will present an important advantage of this approach.

The lexicon cannot contain all words of a language, because new words are created every day. Therefore within the ASSP-project a lexicon was chosen that contains all morphemes, that is monomorphemic words and affixes, of Dutch (cf. Heemskerk—van Heuven, this volume). In order to look up their pronunciation, derived words must be decomposed into their constituent parts. This requires morphological analysis. For this reason

MORPA, a morphological parser, was developed (cf. Heemskerk—van Heuven, this volume). MORPA analyzes all words that are created by productive processes. Words formed by improductive processes are listed in the lexicon.

The lexical representation of each morpheme contains among other things the orthographic form, phoneme transcription (pron), and classification (cls) of the morphemes, and the category (cat) of the resulting word. The phoneme representation in the lexicon consists of the phonemic form of the morpheme supplemented with word stress and syllable structure. For instance, the representation of the word *monument* 'id.' is given in (3):

(3)   *monument*, [[pron,ˌmoː-nyˑ-ˈmɛnt], [cat,noun], [cls,stem]]

By choosing the lexicon-based approach, we can be sure that phonological rules generally operate on correctly analyzed words and exceptions can be treated efficiently. For instance, the irregular stress patterns of the words *ˈkolibrie* and *prosoˈdie* do not have to be derived because word stress is in the lexicon. Idiosyncratic stress behavior of nonnative derived words is not problematic either. If a syllable is exceptionally stressed it is usually in the suffix, so this stress pattern can either be noted in the lexical representation of the suffix, or accounted for by a special suffix label that triggers a specific stress rule. For example, the suffix *-ologie* in *methodologie* 'methodology' which has irregular final stress, has the representation /oː-loː-ˈɣiˑ/.

As in MITalk, grapheme-to-phoneme conversion rules are still used for words that MORPA fails to analyze, or words that should not be analyzed by MORPA such as proper names, acronyms, and digits (cf. van Leeuwen—te Lindert, this volume).

3. From underlying to surface phoneme representation

In this section we will discuss which rules are needed to convert the phoneme representation found in the lexicon into the surface phoneme representation that is the input for synthesis modules. Subsequently we will present the formalization and organization of these rules.

3.1. Division of labor between lexical representation and rule set

At first sight it would be easiest to enter the surface representation of each morpheme in the lexicon. However, this is not possible since some

phonological processes work across morpheme boundaries, as was illustrated in (2). An example of such a process is progressive voicing assimilation:

/zɑk/ 'pocket'	⟹	[zɑk] in	[loːn#zɑk]	'pay packet'
			[ɪn##zɑ-kən]	'in pockets'
		[sɑk] in	[hɸːp#sɑk]	'hip pocket'
			[ɔp##sɑk]	'in one's pocket'

Another reason why a morpheme can receive more than one surface pronunciation, is that one or more processes that can be applied to it, are optional and depend on factors such as style, tempo etc. An example of a optional rule is n-deletion, that is illustrated below:

lopen 'to walk'	⟹	[loː-pə]	(n-deletion applied)	or:
	⟹	[loː-pən]	(n-deletion not applied)	

In generative phonology such variation is accounted for by choosing a sufficiently abstract form and deriving all surface variants from it by rule:

hɸːp#zɑk	loːn#zɑk		loː-pən	loː-pən	
s	–	assimilation		∅	n-deletion
hɸp#sɑk	loːn#zɑk		loː-pən	loː-pə	

We have adopted this approach for the derivation of surface forms in MORPHON.

However, in generative phonology processes that are both morpheme internal and obligatory are formalized by rules as well, because all regularities are expressed by rules rather than in the lexical representation. These rules will not be adopted by us, as it is more efficient to enter the effect of those processes in the lexical representation. For instance, word stress and syllable structure are entered in the lexicon. Subsequent modification of morphemes in context is accounted for by restressing and resyllabification rules. Therefore stress assignment and syllabification are lexical but restressing and resyllabification are part of our rule set.

In some cases resyllabification and restressing can be seen as reapplications of the rules for underived words to derived words. This holds for words derived with nonnative affixes, for instance the word ˌkrɪstalli ˈseer 'crystallize', that consists of the word *kristal* 'crystal' followed by the nonnative suffix *-iseer*, has exactly the same syllable structure and stress pattern as the comparable monomorphemic word ˌexperi ˈment 'id.'. Consequently, rules for underived words must be part of MORPHON, although real underived words receive their stress pattern and syllable

structure in the lexicon. Affixes must be classified as either native or nonnative. Nonnative affixes and roots are only syllabified and stressed after affixation has taken place. Some derivations are given below:

	underived stem	stem+nonnative affix	root+nonnative affix
input	krɪs-ˈtɑl	krɪs-ˈtɑl + iˑzeːr	ɑtmiˑniˑstr + aːtsiˑ
rules	-	ˌkrɪs-tɑ-liˑ-ˈzeːr	ˌɑt-miˑ-niˑ-ˈstraː-tsiˑ
	krɪs-ˈtɑl	ˌkrɪs-tɑ-liˑ-ˈzeˑr	ˌɑt-miˑ-niˑ-ˈstraː-tsiˑ

MORPHON therefore contains rules that work across morpheme boundaries as well as optional rules. The derivation of the surface pronunciation by MORPHON is illustrated in Figure 1:
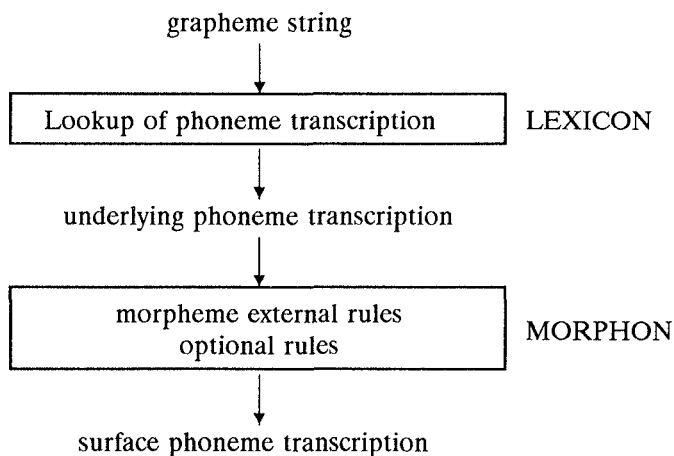

Figure 1. The derivation of the surface pronunciation by MORPHON

A subset of the phonological rules, i.e. the segmental rules (and only those that are phonologically motivated, not morpho-phonological rules) were investigated in the project FONWET (cf. Jongenburger—van Heuven, this volume). Rules that are optional or potentially redundant, are switched on or off in accordance with the findings of this project. Anticipating the findings of FONWET, all optional and potentially redundant rules were implemented. This makes MORPHON very flexible.

3.2. Formalization of the rules

MORPHON is formalized as a set of ordered rewrite rules of the form A $\Rightarrow$ B / X _ Y, which means that A is converted to B in the right context X and left context Y (cf. Chomsky—Halle 1968). An example of a MORPHON rule, viz. progressive voicing assimilation, the rule that devoices the /v/ in (2), is given in (4):

(4)   < –son, +cont, +voice> $\Rightarrow$ < –voice> / < –son, +segm, –voice> _

The notation slightly differs from that of phonologists, because MORPHON was implemented with the rule compiler TooLiP (cf. van Leeuwen 1989a). The use of a compiler has the advantage that computational and linguistic knowledge are separated (cf. Berendsen—Langeweg— van Leeuwen 1986). On the other hand, a serious disadvantage of TooLip is that it can only handle a linear string of phonemes. This means that syllable boundaries must be represented as segments and stress is denoted by labels. An example of a stress rule is given in (5), viz. a rule assigning tertiary stress to a vowel before a consonant:

(5)   vow $\Rightarrow$ <3stress> / _ cons

3.3. Organization of the rule set

We already saw that some words composed with nonnative affixes undergo stress assignment and other rules as if they were underived. Native suffixes are stress neutral (cf. van Beurden 1987; Trommelen—Zonneveld 1989). This can be illustrated by the following examples: *-ist* and *-iseer* are nonnative suffixes that trigger restressing; *-heid*, *-loos*, and *-achtig* are native suffixes, which do not influence the stress pattern of the word to which they attach:

kristal	kristalliseer	kristalachtig
[krɪsˈtɑl]	[krɪsɑliˈˈzeːr]	[krɪsˈtɑlɑχtəχ]
'crystal'	'crystallize'	'crystalline'

Recent theories of morphology and phonology explain this phenomenon as follows: stress is applied to underived words and words derived with nonnative affixes before native affixes are attached. As word formation takes place in the lexicon, the same must hold for phonological rules and the model mentioned above is referred to as "lexical phonology and morpholo-

gy". The word formation rules and the phonological rules that apply to its output are represented as levels in the lexicon. Rules that do not refer to internal structure apply outside the lexicon. The model is given in Figure 2.
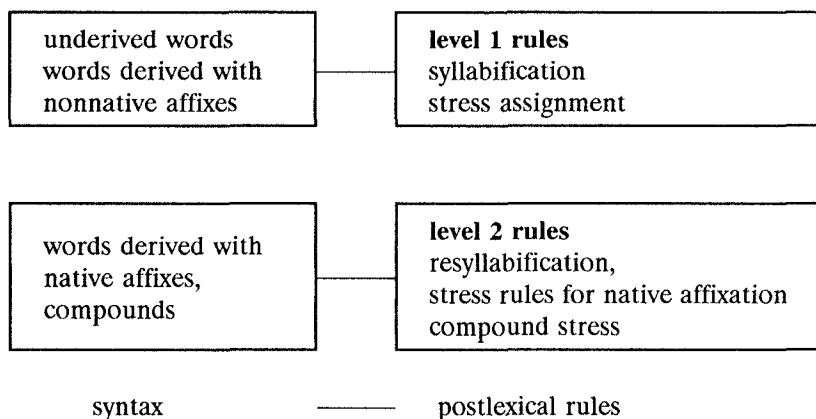
| underived words
words derived with
nonnative affixes | **level 1 rules**
syllabification
stress assignment |

| words derived with
native affixes,
compounds | **level 2 rules**
resyllabification,
stress rules for native affixation
compound stress |

syntax   ————   postlexical rules

Figure 2. The lexical model

In the text-to-speech system that incorporates MORPHON, SPRAAK-MAKER (cf. van Leeuwen—te Lindert, this volume), morphology and phonology are separate modules, so the difference between the two suffix types cannot be explained by the relative ordering of certain morphological processes and phonological rules. Therefore the older approach of Chomsky—Halle (1968) to this problem has to be chosen: native suffixes and nonnative suffixes get different morpheme boundaries and level 1 rules are blocked by native morpheme boundaries. The lexical model is used in MORPA, however, to reduce the number of possible analyses for each word by taking ordering restrictions that follow from this model into account (cf. Heemskerk 1989; Heemskerk—van Heuven, this volume).

The organization in levels was adopted as the organizational principle of MORPHON. MORPHON consists of the following three modules:

MORPHON_1 contains rules for words derived with nonnative affixes. The most important rules on this level are syllabification and stress assignment. These are linear versions of the non-linear rules by Kager (1989). An advantage of our stress rules is that secondary stresses, which are relevant to stress-dependent segmental rules and stress adjustment rules, are generated as a by-product of main stress assignment. MORPHON_1 also contains some morpho-lexical rules, e.g. velar softening, the rule that

accounts for the alternation between [k] and [s] in words like *provoceer* [proːvoːˈseːr] 'provoke' and *provocatie* [proːvoːkaːtsiˑ] 'provocation'.

MORPHON_2 contains rules for words derived with native affixes and for compounds. These rules account, for instance, for stress attraction in words composed with *-lijk* and *-ig*, and rhythmical rules for words composed with *-baar, -loosheid* and *her-* (cf. Langeweg 1988; Trommelen—Zonneveld 1989). Compound stress and stress adjustment are also ordered in this module. We will return to these rules in section 5.

MORPHON_3 contains rules that are not sensitive to the internal structure of words, such as sound adjustment rules like (voicing) assimilation and n-deletion. These rules are discussed in detail in Jongenburger—van Heuven (this volume).

To illustrate MORPHON, the derivation of sentence (2) is given below:

input:	də	˜ɔn-dər#ˈzuˑk%ər		ˈvɔnd	ən	sɪχniˑfiˑC+ɑnt vər#ˈsχɪl
MORFON_1		˜ɔn-dər-ˈzuˑk%ər				–
velar softening						sɪχniˑfiˑkɑnt
syllabification						sɪχ-niˑ-fiˑ-kɑnt
stress						ˌsɪχ-niˑ-fiˑ-ˈkɑnt
MORFON_2						
syllabification		˜ɔn-dər-ˈzuˑ-kər				
stress adjustment						ˈsɪχ-niˑ-fiˑ-ˌkɑnt
MORFON_3						
glottal stop		ʔ˜ɔn-dər-ˈzuˑ-kər				
final devoic.				ˈvɔnt		
assimilation						fər-ˈsχɪl
output:	də	ʔ˜ɔn-dər-ˈzuˑ-kər		ˈvɔnt	ən	ˈsɪχ-niˑ-fiˑ-ˌkɑnt fər-ˈsχɪl

4. The role of syntactic category

So far we have only illustrated the necessity of segmentation and morpheme classification for the derivation of the correct pronunciation (recall that syllabification is sensitive to morpheme boundaries, and that native and nonnative affixes behave differently with respect to stress assignment).
    MORPA also provides morpho-syntactic information, viz. syntactic category. Some rules, notably stress rules, need this type of information. In

this section we will discuss rules that could only be formulated because the category of words is known, viz. compound stress rules and stress adjustment. We will not give the precise formulation of these rules in MORPHON but concentrate on their effect.

In earlier linguistic interfaces for Dutch text-to-speech conversion such as GRAFON and FONPARS, primary stress is indiscriminately assigned to the leftmost member of any compound. However, the stress pattern of compounds depends on the syntactic category of the top node (cf. Langeweg 1988). If the compound is a noun or verb, stress is assigned to the left-hand member and if it is a adjective, adverb or preposition to the right-hand member. This is illustrated by the following compounds:

(6)   *noun or verb*   *adjective, adverbial, etc.*
        'onrecht   'injustice'   on'echt   'unreal'
        'kolenschop   'coal shovel'   boven'op   'on top of'

Syntactic category is not only crucial for the determination of compound stress in regular cases but also for some exceptions. Compounds that consist of a lexicalized phrase or contain a phrase exhibit a deviating stress pattern. In phrases stress is on the right-hand member in the cases that concern us here, even if the category of the resulting word is noun or verb (the elements between brackets are lexicalized phrases):

        [rode 'kruis]verpleegster   'red cross nurse'
        [buiten 'boord]motor   'outboard motor'
        [stad 'huis]   'city hall'

This is accounted for by the Phrasal Stress Rule that stresses the right-hand element irrespective of word category. Some phrases can be distinguished formally, for instance, words that contain an inflected adjective like *rode 'kruisverpleegster* are almost always phrases.

        [gemene 'best]land   'commonwealth country'
        [ronde 'tafel]conferentie   'round table conference'

Not all phrases can be identified this way. Therefore some highly frequent phrases (e.g. *stadhuis*) are listed in the lexicon.

Syntactic category is also important for rhythmic adjustment rules that reverse the prominence pattern laid down by word and compound stress rules (cf. Visch 1990). For instance, the primary and secondary stresses are interchanged in adjectives that are followed by a word that begins with a primary stressed syllable. This rule that is called stress shift, is illustrated in (7):

(7)   ˌrespec ˈtabel   ˈrespec ˌtabel ˈman   'respectable man'
      kon ˈtant   ˈkontant ˈgeld   'cash'
      ˌregio ˈnaal   ˈregio ˌnaal ˈdagblad   'regional newspaper'

Stress shift can not be formulated in rule-based convertors, because the rules cannot refer to the syntactic category of a word. However, since this rule is optional, this is not disastrous.

However, there is a similar rule, called stress retraction, that reverses prominence within adjectival compounds. This rule is obligatory when the whole compound is in focus (but not when *donkerblauw pak* 'deep blue suit' contrasts with *donkerbruin pak* 'dark brown suit'). The effect of this retraction is illustrated in (8):

(8)   ˌdonker ˈblauw   ˈdonker ˌblauw ˈpak   'deep blue (suit)'

We conclude that for the correct assignment of stress to compounds and lexicalized phrases, as well as for stress adjustment and stress retraction word category is crucial. This is yet another justification of our approach, as this type of morpho-syntactic information is made available by lexicon-based letter-to-sound conversion only.

5. Comparison with earlier grapheme-phoneme conversion systems

In section 2 and 4 we claimed that lack of morphological information such as segmentation, morpheme classification and syntactic category, may well lead to errors in grapheme-to-phoneme conversion and phonological rules. Therefore it is to be expected that MORPA-cum-MORPHON as a lexicon-based convertor has a better pronunciation accuracy than rule based-systems. The pronunciation accuracy of MORPA-cum-MORPHON was compared with that of two grapheme-to-phoneme conversion modules for Dutch that are currently available, namely GRAFON (Berendsen—Don—Langeweg 1986; van Leeuwen 1989a) and FONPARS (Kerkhoff—Wester—Boves 1984).

The test material consisted of isolated words in order to keep the test simple (although some MORPHON-rules have a larger scope, there are few rules that *only* work across word boundaries). This implies that the effect of PROS, the SPRAAKMAKER-module that takes care of, among other things, disambiguation of MORPA-analyses and phrasing (cf. Quené—Kager, this volume; Dirksen—Quené, this volume), and TEXTSCAN, the module that handles punctuation, abbreviations, etc. (cf. van Holsteijn, this volume), could not be included in the evaluation.

A test file was created that consisted of words from newspaper text, compounds and words with a very low frequency. The most frequent words were deleted, because MORPA-cum-MORPHON, when functioning in the text-to-speech system SPRAAKMAKER, will only convert words that are not found in HIFREQ, a dictionary of the 10,000 most frequent word forms (cf. Lammens, this volume). Of the remaining words of the three files the first 700 words were selected. Acronyms and words with punctuation marks such as hyphens and apostrophes, proper names and words with spelling errors were deleted, as MORPA-cum-MORPHON were not designed to handle these words. A file of 1,985 words remained. This test file was submitted to each of the three systems and the transcriptions of these systems were classified as correct or incorrect. A word is considered correct if it has a proper phoneme transcription, which means that all appropriate non-optional phonological rules have been applied. Furthermore, the words must have the correct syllable structure and stress pattern. The results of the comparison are given in Table 1:

Table 1.   Percentage of words that receive a correct phoneme transcription

LEXICON BASED MORPA-cum-MORPHON	RULE BASED	
	GRAFON	FONPARS
78%	66%	60%

On the basis of the data in Table 1 the following observations can be made: MORPA-cum-MORPHON has a considerably better pronunciation accuracy than either GRAFON or FONPARS. This is a remarkably good result, given the fact that we are dealing with low frequency words. The difference in accuracy between the two rule-based systems is quite small, but GRAFON is to be preferred.

On the basis of the test results, we can also make some observations on MORPA-cum-MORPHON in its own right. Seventy-eight percent of the words are assigned a correct phoneme transcription. Sixteen percent of the errors could be traced back to the phonological module, the remaining errors are due to faulty morphological analysis. Of the errors made by MORPA 82 percent led to an incorrect pronunciation representation. As expected, segmentation errors almost always led to an incorrect phoneme transcription. Interestingly, category assignment errors also cause incorrect pronunciation, though less often. This bears out the importance of syntactic category information.

In order to see how MORPHON would perform when applied to perfectly analyzed words, the morphological analyses were corrected by hand, and again converted by MORPHON. MORPHON then provides correct phoneme transcriptions. For ninety-five percent of the words the remaining errors made by MORPHON have to do with stress assignment.

MORPA and MORPHON were still being developed at the time that this evaluation was performed (July/August 1990), for instance, MORPA did not yet contain the algorithm that selects the most plausible analysis from the list of alternatives on the basis of frequency information. The results would probably be even better if the evaluation were repeated now, since MORPA now provides the correct analysis for ninety-two percent of the words (cf. Heemskerk—van Heuven, this volume).

6. Conclusions

In this contribution we argued that the pronunciation representation should be derived from spelled input via the lexicon because this lexicon is needed anyway for exceptions and to provide grapheme-to-phoneme conversion rules with syntactic category. The phoneme representations in lexicon do not render phonological rules superfluous, because some processes are optional, potentially redundant, or sensitive to information outside the morpheme. The phonological rules benefit from the syntactic category now made available. Segmentation and morpheme type information are crucial for almost all rules, whereas word category information is relevant for compound stress and stress adjustment rules. Some other improvements of the phonological rules are independent of morphological information, for instance the translation of some properties of non-linear stress theories in linear rules. Finally, the problem of exceptions is solved by listing exceptional words in the lexicon. The results of a comparison of MORPA-cum-MORPHON with other text-to-phoneme transcription conversion systems proved that MORPA-cum-MORPHON has a considerably better pronunciation accuracy, but is considerably slower.

Notes

1. When relevant, syllable boundaries are indicated in the transcriptions by hyphens; primary stress is indicated by a superior vertical stroke, secondary stress by a inferior vertical stroke, and tertiary stress by a tilde. Affix boundaries are indicated by '%', word boundaries by '#'.