A Finite State and Data-Oriented Method for Grapheme to Phoneme Conversion

Gosse Bouma
Alfa-informatica
Rijksuniversiteit Groningen
Postbus 716
9700 AS Groningen
The Netherlands
gosse@let.rug.nl

Abstract

A finite-state method, based on leftmost longest-match replacement, is presented for segmenting words into graphemes, and for converting graphemes into phonemes. A small set of hand-crafted conversion rules for Dutch achieves a phoneme accuracy of over 93%. The accuracy of the system is further improved by using transformation-based learning. The phoneme accuracy of the best system (using a large rule and a 'lazy' variant of Brill's algoritm), trained on only 40K words, reaches 99%.

1 Introduction

Automatic grapheme to phoneme conversion (i.e. the conversion of a string of characters into a string of phonemes) is essential for applications of text to speech synthesis dealing with unrestricted text, where the input may contain words which do not occur in the system dictionary. Furthermore, a transducer for grapheme to phoneme conversion can be used to generate candidate replacements in a (pronunciation-sensitive) spelling correction system. When given the pronunciation of a misspelled word, the inverse of the grapheme to phoneme transducer will generate all identically pronounced words. Below, we present a method for developing such grapheme to phoneme transducers based on a combination of hand-crafted conversion rules, implemented using finite state calculus, and automatically induced rules.

The hand-crafted system is defined as a two-step procedure: segmentation of the input into a sequence of graphemes (i.e. sequences of one or more characters typically corresponding to a single phoneme) and conversion of graphemes into (sequences of) phonemes. The composition of the transducer which performs segmentation and the transducer defined by the conversion rules, is a transducer which converts sequences of characters into sequences of phonemes.

Specifying the conversion rules is a difficult task. Although segmentation of the input can in principle be dispensed with, we found that writing conversion rules for segmented input substantially reduces the context-sensitivity and order-dependence of such rules. We manually developed a grapheme to phoneme transducer for Dutch data obtained from CELEX (Baayen et al., 1993) and achieved a word accuracy of 60.6% and a phoneme accuracy of 93.6%.

To improve the performance of our system, we used transformation-based learning (TBL) (Brill, 1995). Training data are obtained by aligning the output of the hand-crafted finite state transducer with the correct phoneme strings. These data can then be used as input for TBL, provided that suitable rule templates are available. We performed several experiments, in which the amount of training data, the algorithm (Brill's original formulation and 'lazy' variants (Samuel et al., 1998)), and the number of rule templates varied. The best experiment (40K words, using a 'lazy' strategy with a large set of rule templates) induces over 2000 transformation rules, leading to 92.6% word accuracy and 99.0% phoneme accuracy. This result, obtained using a relatively small set of training data, compares well with that of other systems.

2 Finite State Calculus

As argued in Kaplan and Kay (1994), Karttunen (1995), Karttunen et al. (1997), and elsewhere, many of the rules used in phonology and morphology can be analysed as special cases of regular expressions. By extending the language of regular expressions with operators which capture the interpretation of linguistic rule systems, high-level linguistic descriptions can be compiled into finite state automata directly. Furthermore, such automata can be combined with other finite state automata performing low-level tasks such as tokenization or lexical-lookup, or more advanced tasks such as shallow parsing. Composition of the individual components into a single transducer may lead to highly efficient processing.

The system described below was implemented using FSA Utilities,[1] a package for implementing and manipulating finite state automata, which provides possibilities for defining new regular expression oper-

[1] `www.let.rug.nl/~vannoord/fsa/`

`[]`	the empty string
`[`$R_1, \ldots, R_n$`]`	concatenation
$\{R_1, \ldots, R_n\}$	disjunction
`R^`	optionality
`ignore(A,B)`	ignore: A interspersed with elements of B
`A x B`	cross-product: the transducer which maps all strings in A to all strings in B.
`identity(A)`	identity: the transducer which maps each element in A onto itself.
`T o U`	composition of the transducers T and U.
`macro(Term,R)`	use Term as an abbreviation for R (where Term and R may contain variables).

Figure 1: A fragment of FSA regular expression syntax. A and B are regular expressions denoting recognizers, T and U transducers, and R can be either.

ators. The part of FSA's built-in regular expression syntax relevant to this paper, is listed in figure 1.

One particular useful extension of the basic syntax of regular expressions is the replace-operator. Karttunen (1995) argues that many phonological and morphological rules can be interpreted as rules which replace a certain portion of the input string. Although several implementations of the replace-operator are proposed, the most relevant case for our purposes is so-called 'leftmost longest-match' replacement. In case of overlapping rule targets in the input, this operator will replace the leftmost target, and in cases where a rule target contains a prefix which is also a potential target, the longer sequence will be replaced. Gerdemann and van Noord (1999) implement leftmost longest-match replacement in FSA as the operator

```
replace(Target, LeftContext,RightContext),
```

where Target is a transducer defining the actual replacement, and LeftContext and RightContext are regular expressions defining the left- and rightcontext of the rule, respectively.

An example where leftmost replacement is useful is hyphenation. Hyphenation of (non-compound) words in Dutch amounts to segmenting a word into syllables, separated by hyphens. In cases where (the written form of) a word can in principle be segmented in several ways (i.e. the sequence `alfabet` can be segmented as `al-fa-bet`, `al-fab-et`, `alf-a-bet`, or `alf-ab-et`), the segmentation which maximizes onsets is in general the correct one (i.e. `al-fa-bet`). This property of hyphenation is captured by leftmost replacement:

```
macro(hyphenate,
   replace([] x -, syllable, syllable)).
```

Leftmost replacement ensures that hyphens are introduced 'eagerly', i.e. as early as possible. Given a suitable definition of `syllable`, this ensures that wherever a consonant can be final in a coda or initial in the next onset, it is in fact added to the onset.

The segmentation task discussed below makes crucial use of longest match.

3 A finite state method for grapheme to phoneme conversion

Grapheme to phoneme conversion is implemented as the composition of four transducers:

```
macro(graph2phon,
   segmentation      % segment the input
 o mark_begin_end    % add '#'
 o conversion        % apply rules
 o clean_up       ). % remove markers
```

An example of conversion including the intermediate steps is given below for the word `aanknopingspunt` (*connection-point*).

```
input:  aanknopingspunt
s:      aa-n-k-n-o-p-i-ng-s-p-u-n-t-
m:      #-aa-n-k-n-o-p-i-ng-s-p-u-n-t-#
co:     #-a+N+k-n-o-p-I+N+s-p-}+n-t-#
cl:     aNknopINsp}nt
```

The first transducer (**segmentation**) takes as its input a sequence of characters and groups these into segments. The second transducer (**mark_begin_end**) adds a marker ('#') to the beginning and end of the sequence of segments. The third transducer (**conversion**) performs the actual conversion step. It converts each segment into a sequence of (zero or more) phonemes. The final step (**clean_up**) removes all markers. The output is a list of phonemes in the notation used by CELEX (which can be easily translated into the more common SAMPA-notation).

3.1 Segmentation

The goal of segmentation is to divide a word into a sequence of graphemes, providing a convenient input

level of representation for the actual grapheme to phoneme conversion rules.

While there are many letter-combinations which are realized as a single phoneme (`ch, ng, aa, bb, ..`), it is only rarely the case that a single letter is mapped onto more than one phoneme (`x`), or that a letter receives no pronunciation at all (such as word-final `n` in Dutch, which is elided if it is proceeded by a schwa). As the number of cases where multiple letters have to be mapped onto a single phoneme is relatively high, it is natural to model a letter to phoneme system as involving two subtasks: segmentation and conversion. Segmentation splits an input string into graphemes, where each grapheme typically, but not necessarily, corresponds to a single phoneme.

Segmentation is defined as:

```
macro(segmentation,
 replace(
  [identity(graphemes), [] x - ],[],[])
        ).
```

The macro `graphemes` defines the set of graphemes. It contains 77 elements, some of which are:

```
a, aa, au, ai, aai, e, ee, ei, eu, eau,
eeu, i, ie, iee, ieu, ij, o, oe, oei,..
```

Segmentation attaches the marker '-' to each grapheme. Segmentation, as it is defined here, is not context-sensitive, and thus the second and third arguments of replace are simply empty. As the set of graphemes contains many elements which are substrings of other graphemes (i.e. `e` is a substring of `ei`, `eau`, etc.), longest-match is essential: the segmentation of `beiaardier` (*carillon player*) should be `b-ei-aa-r-d-ie-r-` and not `b-e-i-a-a-r-d-i-e-r-`. This effect can be obtained by making the segment itself part of the target of the replace statement. Targets are identified using leftmost longest-match, and thus at each point in the input, only the longest valid segment is marked.

The set of graphemes contains a number of elements which might seem superfluous. The grapheme `aai`, for instance, translates as `aj`, a sequence which could also be derived on the basis of two graphemes `aa` and `i`. However, if we leave out the segment `aai`, segmentation (using leftmost longest match) of words such as `waaien` (*to blow*) would lead to the segmentation `w-aa-ie-n`, which is unnatural, as it would require an extra conversion rule for `ie`. Using the grapheme `aai` allows for two conversion rules which always map `aai` to `aj` and `ie` goes to `i`.

Segmentation as defined above provides the intuitively correct result in almost all cases, given a suitably defined set of graphemes. There are some cases which are less natural, but which do not necessarily lead to errors. The grapheme `eu`, for instance, almost always goes to '`|`', but translates as '`e,j,}`' in (loan-) words such as `museum` and `petroleum`. One might argue that a segmentation `e-u-` is therefore required, but a special conversion rule which covers these exceptional cases (i.e. `eu` followed by `m`) can easily be formulated. Similarly, `ng` almost always translates as `N`, but in some cases actually represents the two graphemes `n-g-`, as in `aaneengesloten` (*connected*), where it should be translated as `NG`. This case is harder to detect, and is a potential source of errors.

3.2 The Conversion Rules

The `g2p` operator is designed to facilitate the formulation of conversion rules for segmented input:

```
macro(g2p(Target,LtCont,RtCont),
  replace([Target, - x +],
             [ignore(LtCont,{+,-}), {-,+}],
                ignore(RtCont,{+,-})
          )
     ).
```

The g2p-operator implements a special purpose version of the replace-operator. The replacement of the marker '-' by '+' in the target ensures that g2p-conversion rules cannot apply in sequence to the same grapheme.[2] Second, each target of the g2p-operator must be a grapheme (and not some substring of it). This is a consequence of the fact that the final element of the left-context must be a marker and the target itself ends in '-'. Finally, the ignore statements in the left and right context imply that the rule contexts can abstract over the potential presence of markers.

An overview of the conversion rules we used for Dutch is given in Figure 2. As the rules are applied in sequence, exceptional rules can be ordered before the regular cases, thus allowing the regular cases to be specified with little or no context. The `special_vowel_rules` deal with exceptional translations of graphemes such as `eu` or cases where `i` or `ij` goes to '`@`'. The `short_vowel_rules` treat single vowels preceding two consonants, or a word final consonant. One problematic case is `e`, which can be translated either as `'E'` or `'@'`. Here, an approximation is attempted which specifies the context where `e` goes `'E'`, and subsumes the other case under the general rule for short vowels. The `special_consonant_rules` address devoicing and a few other exceptional cases. The `default_rules` supply a default mapping for a large number of

[2] Note that the input and output alphabet are not disjoint, and thus rules applying in sequence to the same part of the input are not excluded in principle.

graphemes. The target of this rule is a long disjunction of grapheme-phoneme mappings. As this ruleset applies after all more specific cases have been dealt with, no context restrictions need to be specified.

Depending somewhat on how one counts, the full set of conversion rules for Dutch contains approximately 80 conversion rules, more than 40 of which are default mappings requiring no context.[3] Compilation of the complete system results in a (minimal, deterministic) transducer with 747 states and 20,123 transitions.

3.3 Test results and discussion

The accuracy of the hand-crafted system was evaluated by testing it on all of the words wihtout diacritics in the CELEX lexical database which have a phonetic transcription. After several development cycles, we achieved a word accuracy of 60.6% and a phoneme accuracy (measured as the edit distance between the phoneme string produced by the system and the correct string, divided by the number of phonemes in the correct string) of 93.6%.

There have been relatively few attempts at developing grapheme to phoneme conversion systems using finite state technology alone. Williams (1994) reports on a system for Welsh, which uses no less than 700 rules implemented in a rather restricted environment. The rules are also implemented in a two-level system, PC-KIMMO, (Antworth, 1990), but this still requires over 400 rules. Möbius et al. (1997) report on full-fledged text-to-speech system for German, containing around 200 rules (which are compiled into a weighted finite state transducer) for the grapheme-to-phoneme conversion step. These numbers suggest that our implementation (which contains around 80 rules in total) benefits considerably from the flexibility and high-level of abstraction made available by finite state calculus.

One might suspect that a two-level approach to grapheme to phoneme conversion is more appropriate than the sequential approach used here. Somewhat surprisingly, however, Williams concludes that a sequential approach is preferable. The formulation of rules in the latter approach is more intuitive, and rule ordering provides a way of dealing with exceptional cases which is not easily available in a two-level system.

While further improvements would definitely have been possible at this point, it becomes increasingly difficult to do this on the basis of linguistic knowledge alone. That is, most of the rules which have to be added deal with highly idiosyncratic cases (often related to loan-words) which can only be discovered by browsing through the test results of previous runs. At this point, switching from a linguistics-oriented to a data-oriented methodology, seemed appropriate.

[3] It should be noted that we only considered words which do not contain diacritics. Including those is unproblematic in principle, but would lead to a slight increase of the number of rules.

4 Transformation-based grapheme to phoneme conversion

Brill (1995) demonstrates that accurate part-of-speech tagging can be learned by using a two-step process. First, a simple system is used which assigns the most probable tag to each word. The results of the system are aligned with the correct tags for some corpus of training data. Next, (context-sensitive) transformation rules are selected from a pool of rule patterns, which replace erroneous tags by correct tags. The rule with the largest benefit on the training data (i.e. the rule for which the number of corrections minus the number of newly introduced mistakes, is the largest) is learned and applied to the training data. This process continues until no more rules can be found which lead to improvement (above a certain threshold).

Transformation-based learning (TBL) can be applied to the present problem as well.[4] In this case, the base-line system is the finite state transducer described above, which can be used to produce a set of phonemic transcriptions for a word list. Next, these results are aligned with the correct transcriptions. In combination with suitable rule patterns, these data can be used as input for a TBL process.

4.1 Alignment

TBL requires aligned data for training and testing. While alignment is mostly trivial for part-of-speech tagging, this is not the case for the present task. Aligning data for grapheme-to-phoneme conversion amounts to aligning each part of the input (a sequence of characters) with a part of the output (a sequence of phonemes). As the length of both sequences is not guaranteed to be equal, it must be possible to align more than one character with a single phoneme (the usual case) or a single character with more than one phoneme (the exceptional case, i.e. 'x'). The alignment problem is often solved (Dutoit, 1997; Daelemans and van den Bosch, 1996) by allowing 'null' symbols in the phoneme string, and introducing 'compound' phonemes, such as 'ks' to account for exceptional cases where a single character must be aligned with two phonemes.

As our finite state system already segments the input into graphemes, we have adopted a strategy where *graphemes* instead of characters are aligned with phoneme strings (see Lawrence and Kaye (1986) for a similar approach). The correspondence

[4] Hoste et al. (2000b) compare TBL to C5.0 (Quinlan, 1993) on a similar task, i.e. the mapping of the pronunciation of one regional variant of Dutch into another.

```
macro(conversion,     special_vowel_rules o short_vowel_rules
                    o special_consonant_rules o default_rules  ).
macro(special_vowel_rules,
    g2p([e,u] x [e,j,}], [],  m)  %% museum
  o g2p(i     x @,       [],  g)  %% moedig(st)
  o g2p([i,j] x @,       l,   k)  %% mogelijkheid
  .... ).
macro(short_vowel_rules,
    g2p(e x 'E', [], {[t,t],[k,k],x,...})
    g2p({ a x 'A' , e x @, i x 'I' , o x 'O', u x '}' }, [], [cons, {cons , #}] )
     ).
macro(special_consonant_rules,
     g2p(b x p,         [], {s,t,#})
   o g2p([d,t^] x t,    [], {s,g,k,j,v,h,z,#})
   o g2p({ f x v, s x z},  [], {b,d})
   o g2p(g x 'G',       vowel, vowel)
   o g2p(n x 'N',       [], {k,q})
   o g2p(n x [],        [@],[#])
   ...).
macro(default_rules,
  g2p({ [a,a] x a, [a,a,i] x [a,j], [a,u] x 'M',  [e,a,u] x o, ....,
        [b,b] x b, [d,d] x d, ....,  [c,h] x 'x', [s,c,h] x [s,x], [n,g] x 'N', ...
      }, [], [] )
     ).
```

Figure 2: Conversion Rules

between graphemes and phonemes is usually one to one, but it is no problem to align a grapheme with two or more phonemes. Null symbols are only introduced in the output if a grapheme, such as word-final 'n', is not realized phonologically.

For TBL, the input actually has to be aligned both with the system output as well as with the correct phoneme string. The first task can be solved trivially: since our finite state system proceeds by first segmenting the input into graphemes (sequences of characters), and then transduces each grapheme into a sequence of phonemes, we can obtain aligned data by simply aligning each grapheme with its corresponding phoneme string. The input is segmented into graphemes by doing the segmentation step of the finite state transducer only. The corresponding phoneme strings can be identified by applying the conversion transducer to the segmented input, while keeping the boundary symbols '-' and '+'. As a consequence of the design of the conversion-rules, the resulting sequence of separated phonemes sequences stands in a one-to-one relationship to the graphemes. An example is shown in figure 3, where GR represents the grapheme segmented string, and SP the (system) phoneme strings produced by the finite state transducer. Note that the final SP cell contains only a boundary marker, indicating that the grapheme 'n' is translated into the null phoneme.

For the alignment between graphemes (and, indirectly, the system output) and the correct phoneme strings (as found in Celex), we used the 'hand-seeded' probabilistic alignment procedure described by Black et al. (1998). From the finite state conversion rules, a set of possible *grapheme* → *phoneme sequence* mappings can be derived. This *allowables*-set was extended with (exceptional) mappings present in the correct data, but not in the hand-crafted system. We computed all possible alignments between (segmented) words and correct phoneme strings licenced by the *allowables*-set. Next, probabilities for all allowed mappings were estimated on the basis of all possible alignments, and the data was parsed again, now picking the most probable alignment for each word. To minimize the number of words that could not be aligned, a maximum of one unseen mapping (which was assigned a low probability) was allowed per word. With this modification, only one out of 1000 words on average could not be aligned.[5] These words were discarded.The aligned phoneme

Word	aalbessen (*currants*)						
GR	aa -	l -	b -	e -	ss -	e -	n -
SP	a +	l -	b -	@ +	s +	@ +	+
CP	a	l	b	E	s	@	

Figure 3: Alignment

[5]Typical cases are loan words (*umpires*) and letter words (i.e. abbreviations) (*abc*).

method	training data (words)	phoneme accuracy	word accuracy	induced rules	CPU time (in minutes)
Base-line		93.6	60.6		
Brill	20K	98.0	86.1	447	162
Brill	40K	98.4	88.9	812	858
lazy(5)	20K	97.6	83.5	337	43
lazy(5)	40K	98.2	87.0	701	190
lazy(5)	60K	98.4	88.3	922	397
lazy(10)	20K	97.7	84.3	368	83
lazy(10)	40K	98.2	87.5	738	335
lazy(10)	60K	98.4	88.9	974	711
lazy(5)+	20K	98.6	89.8	1225	186
lazy(5)+	40K	99.0	92.6	2221	603

Figure 4: Experimental Results using training data produced by graph2phon

string for the example in figure 3 is shown in the bottom line. Note that the final cell is empty, representing the null phoneme.

4.2 The experiments

For the experiments with TBL we used the μ-TBL-package (Lager, 1999). This Prolog implementation of TBL is considerably more efficient (up to ten times faster) than Brill's original (C) implementation. The speed-up results mainly from using Prolog's first-argument indexing to access large quantities of data efficiently.

We constructed a set of 22 rule templates which replace a predicted phoneme with a (corrected) phoneme on the basis of the underlying segment, and a context consisting either of phoneme strings, with a maximum length of two on either side, or a context consisting of graphemes, with a maximal length of 1 on either side. Using only 20K words (which corresponds to almost 180K segments), and Brill's algorithm, we achieved a phoneme accuracy of 98.0% (see figure 4) on a test set of 20K words of unseen data.[6] Going to 40K words resulted in 98.4% phoneme accuracy. Note, however, that in spite of the relative efficiency of the implementation, CPU time also goes up sharply.

The heavy computation costs of TBL are due to the fact that for each error in the training data, all possible instantiations of the rule templates which correct this error are generated, and for each of these instantiated rules the score on the whole training set has to be computed. Samuel et al. (1998) therefore propose an efficient, 'lazy', alternative, based on Monte Carlo sampling of the rules. For each error in the training set, only a sample of the rules is considered which might correct it. As rules which correct a high number of errors have a higher chance of being sampled at some point, higher scoring rules are more likely to be generated than lower scoring rules, but no exhaustive search is required. We experimented with sampling sizes 5 and 10. As CPU requirements are more modest, we managed to perform experiments on 60K words in this case, which lead to results which are comparable with Brill's algoritm applied to 40K words.

Apart from being able to work with larger data sets, the 'lazy' strategy also has the advantage that it can cope with larger sets of rule templates. Brill's algorithm slows down quickly when the set of rule templates is extended, but for an algorithm based on rule sampling, this effect is much less severe. Thus, we also constructed a set of 500 rule templates, containing transformation rules which allowed up to three graphemes or phoneme sequences as left or right context, and also allowed for disjunctive contexts (i.e. the context must contain an `'a'` at the first or second position to the right). We used this rule set in combination with a 'lazy' strategy with sampling size 5 (lazy(5)+ in figure 4). This led to a further improvement of phoneme accuracy to 99.0%, and word accuracy of 92.6%, using only 40K words of training material.

Finally, we investigated what the contribution was of using a relatively accurate training set. To this end, we constructed an alternative training set, in which every segment was associated with its most probable phoneme (where frequencies were obtained from the aligned CELEX data). As shown in figure 5, the initial accuracy for such as system is much lower than that of the hand-crafted system. The experimental results, for the 'lazy' algorithm with sampling size 5, show that the phoneme accuracy for training on 20K words is 0.3% less than for the corresponding experiment in figure 4. For 40K words, the difference is still 0.2%, which, in both cases, corresponds to a difference in error rate of around 10%. As might be expected, the number of induced rules

[6] The statistics for less time consuming experiments were obtained by 10-fold cross-validation and for the more expensive experiments by 5-fold cross-validation.

method	training data (words)	phoneme accuracy	word accuracy	induced rules	CPU time (in minutes)
Base-line		72.9	10.8		
lazy(5)	20K	97.3	81.6	691	133
lazy(5)	40K	98.0	86.0	1075	705

Figure 5: Experimental results using data based on frequency.

is much higher now, and thus CPU-requirements also increase substantially.

5 Concluding remarks

We have presented a method for grapheme to phoneme conversion, which combines a hand-crafted finite state transducer with rules induced by a transformation-based learning. An advantage of this method is that it is able to achieve a high level of accuracy using relatively small training sets. Busser (1998), for instance, uses a memory-based learning strategy to achieve 90.1% word accuracy on the same task, but used 90% of the CELEX data (over 300K words) as training set and a (character/phoneme) window size of 9. Hoste et al. (2000a) achieve a word accuracy of 95.7% and a phoneme accuracy of 99.5% on the same task, using a combination of machine learning techniques, as well as additional data obtained from a second dictionary.

Given the result of Roche and Schabes (1997), an obvious next step is to compile the induced rules into an actual transducer, and to compose this with the hand-crafted transducer. It should be noted, however, that the number of induced rules is quite large in some of the experiments, so that the compilation procedure may require some attention.

References

Evan L. Antworth. 1990. *PC-KIMMO : a two-level processor for morphological analysis.* Summer Institute of Linguistics, Dallas, Tex.

R. H. Baayen, R. Piepenbrock, and H. van Rijn. 1993. *The CELEX Lexical Database (CD-ROM).* Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Alan Black, Kevin Lenzo, and Vincent Pagel. 1998. Issues in building general letter to sound rules. In *Proceedings of the 3rd ESCA/COCSADA Workshop on Speech Synthesis*, pages 77–81, Jenolan Caves, Australia.

Eric Brill. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21:543–566.

Bertjan Busser. 1998. TreeTalk-D: a machine learning approach to Dutch word pronunciation. In *Proceedings TSD Conference*, pages 3–8, Masaryk University, Czech Republic.

W. Daelemans and A. van den Bosch. 1996. Language-independent data-oriented grapheme-to-phoneme conversion. In *Progress in Speech Synthesis*, pages 77–90, New York. Springer Verlag.

Thierry Dutoit. 1997. *An Introduction to Text-to-Speech Synthesis.* Kluwer, Dordrecht.

Dale Gerdemann and Gertjan van Noord. 1999. Transducers from rewrite rules with backreferences. In *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 126–133, Bergen.

Veronique Hoste, Walter Daelemans, Erik Tjong Kim Sang, and Steven Gillis. 2000a. Meta-learning of phonemic annotation of corpora. ms., University of Antwerp.

Veronique Hoste, Steven Gillis, and Walter Daelemans. 2000b. A rule induction approach to modelling regional pronunciation variation. ms., University of Antwerp.

Ronald Kaplan and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics*, 20(3).

L. Karttunen, J.P. Chanod, G. Grefenstette, and A. Schiller. 1997. Regular expressions for language engineering. *Natural Lanuage Engineering*, pages 1–24.

Lauri Karttunen. 1995. The replace operator. In *33th Annual Meeting of the Association for Computational Linguistics*, pages 16–23, Boston, Massachusetts.

Torbjörn Lager. 1999. The μ-TBL System: Logic programming tools for transformation-based learning. In *Proceedings of the Third International Workshop on Computational Natural Language Learning (CoNLL'99)*, pages 33–42, Bergen.

S. C. G. Lawrence and G. Kaye. 1986. Alignment of phonemes with their corresponding orthography. *Computer Speech and Language*, 1(2):153–165.

Bernd Möbius, Richard Sproat, Jan van Santen, and Joseph Olive. 1997. The Bell Labs German text-to-speech system: An overview. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 2443–2446, Rhodes.

J. R. Quinlan. 1993. *C4.5: programs for machine learning.* Morgan Kaufmann Publishers, San Mateo.

Emmanuel Roche and Yves Schabes. 1997. Deterministic part-of-speech tagging with finite-state transducers. In Emmanuel Roche and Yves Schabes, editors, *Finite state language processing*, pages 205–239. MIT Press, Cambridge, Mass.

Ken Samuel, Sandra Carberry, and K. Vijay-Shanker. 1998. Dialogue act tagging with transformation-based learning. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, Montreal.

Briony Williams. 1994. Welsh letter-to-sound rules: rewrite rules and two-level rules compared. *Computer Speech and Language*, 8:261–277.