Comparing manually-developed and data-driven rules for P2P learning

Linsen Loots[†], *Marelie Davel*[‡], *Etienne Barnard*[‡], *Thomas Niesler*[†]

[†]Department of Electrical & Electronic Engineering, University of Stellenbosch, South Africa
[‡]HLT Research Group, Meraka Institute, CSIR, South Africa

Abstract

Phoneme-to-phoneme (P2P) learning provides a mechanism for predicting the pronunciation of a word based on its pronunciation in a different accent, dialect or language. We evaluate the effectiveness of manually-developed as well as automatically-derived P2P rules for British to South African English pronunciation conversion. Using the freely-available Oxford Advanced Learners Dictionary of Contemporary English (OALD) as source, the two approaches to P2P conversion are compared with a manually-developed South African English pronunciation dictionary. We show that, when the British English pronunciation is known, a small manually-derived rule set is able to approximate the South African pronunciation surprisingly well. Furthermore we demonstrate that the best performance is achieved by data-driven P2P learning, which proves to be a better mechanism for pronunciation prediction than both manually-derived P2P rules as well as data-driven grapheme-to-phoneme (G2P) conversion.

1. Introduction

The pronunciation dictionary is a key component in speech technology systems, such as systems for automatic speech recognition (ASR) or text-to-speech conversion (TTS). Since the pronunciation dictionary lies at the interface between text and speech (or letter and sound), it is a crucial determinant of the performance of such systems. Significant effort has therefore been expended in the creation of pronunciation dictionaries for the world's dominant languages [1].

In many instances, such as for Spanish, German and most Southern Bantu languages, the mapping between letter and sound is highly regular. In these cases machine-learning algorithms can readily create extensive high-quality pronunciation dictionaries from training sets as small as a few thousand words [2]. However, when the relationship between orthography and pronunciation is less regular, a substantial effort is required to develop reliable dictionaries. English is an extreme case in this regard. For example, words such as *enough, though, through, plough, cough, hiccough* demonstrate that it is often very difficult to predict the pronunciation of an English word from its spelling. Consequently, many person-years of work have gone into the creation of large pronunciation dictionaries in the major English accents.

The differences between English accents can be sufficiently large to prevent the respective pronunciation dictionaries from being shared [3]. In developing a dictionary for a specific English accent, one is therefore faced with a choice: either begin dictionary development for the required accent anew, or in some way convert the pronunciations in an existing dictionary to the desired accent. Given the substantial effort required for the development of a large, high-quality pronunciation dictionary, the latter option is preferable if it can achieve sufficient accuracy.

In this paper, we focus on two approaches to the transformation of pronunciations from an existing and freely-available British English dictionary to standard South African English. In the first, pronunciation conversion rules are created manually, based on phonological knowledge of the source and target accents. In the second approach, a certain number of South African pronunciations are created manually, and subsequently phoneme-to-phoneme (P2P) conversion rules are derived automatically. Both rule sets can be used to transform further pronunciations between the two accents.

2. Standard South African English

Standard South African English (SSAE) refers the pronunciation perceived to be used by the majority of first language speakers [4]. SSAE therefore refers to the received pronunciation, or the perceived 'proper' pronunciation for words. This accent is influenced by four main South African English (SAE) variants, namely, White SAE, Black SAE, Indian SAE and Cape Flats English. These include strongly accented English variants that are not included in SSAE, and are not addressed in this paper.

Various British to South African English adaptations are known, although the context in which these adaptations occur are typically not codified. Examples of adaptations include devoicing of $/z/$ and $/dh/$ (using ARPABET notation), vowel reduction, and the KIT split. The KIT split is one of the most striking features of SSAE. It refers to the behaviour of the $/ih/$ vowel in British received pronunciation, which is realised as two distinct allophones in SSAE: one as $/ih/$ and the other a variant that merges with $/ax/$. For example "kit" and "thin" are realised as the same phoneme in British English, but as two distinct phonemes ($/ih/$ and $/ax/$) in SSAE.

3. Dictionaries

3.1. British English

A number of popular English pronunciation dictionaries are publicly available. If we restrict our attention to standard British English, the most widely-used dictionaries include:

- *BEEP*, the *British English Example Pronunciation Dictionary*, containing over 250 000 words [5]. This dictionary was developed for the purposes of large-vocabulary speech recognition, by combining a number of public resources and creating additional pronunciations for words occurring in the Wall Street Journal corpus.
- *OALD*, the *Oxford Advanced Learners Dictionary of Contemporary English*, a dictionary aimed at general users, with approximately 63 000 words [6].

Although these dictionaries differ somewhat in the conventions employed (e.g. the use of syllabic consonants and the phonemic symbols chosen), they are generally similar in quality. We have chosen to work with OALD in order to facilitate comparisons with other research.

3.2. South African English

SSAE pronunciations were obtained from SAEDICT, a pronunciation dictionary under development at Stellenbosch University [7]. All pronunciations in SAEDICT were transcribed by the same linguistic specialist, to ensure its consistency. Transcriptions were chosen to reflect commonly accepted SSAE pronunciations. SAEDICT is considerably smaller than most other publicly available English pronunciations, and currently contains 36 956 entries.

3.3. Phoneset

ARPABET was chosen as the common phoneset in which to analyse the dictionaries. Both BEEP and OALD use the same phoneset, which can be mapped directly to ARPABET. SAEDICT was transcribed in a phoneset, based on IPA, developed to describe the languages of Southern Africa [8]. This was converted to ARPABET by means of a mapping based on the closest IPA symbol.

3.4. Wordlist

In order to compare pronunciations present in both the OALD-derived dictionary and SAEDICT, the set of words common both dictionaries was determined. In addition, all words with pronunciation variants in either dictionary were excluded from our analysis. The final set of common words contained 16 996 entries. As is standard in the evaluation of pronunciation dictionaries, all entries were weighted equally, even if some were closely related to others.

4. Accent mapping by manually-developed rules

As suggested in Section 1, the manual development of an entire dictionary is typically a time- and effort-intensive process. In addition, depending on the skills of the dictionary developers and the amount of verification implemented (which again increases the cost of the dictionary), the manual development process may itself include errors.

This raises the question: how closely can a manually developed SSAE dictionary be approximated through a set of fairly simple adaptation rules applied to an existing British English dictionary? An initial analysis specifically relating to the KIT vowel was performed in [9]. Here, a small set of rules was developed and evaluated based on a set of 400 manually annotated words. Building on these results and informed by the performance of a newly developed SSAE TTS system [10], a small set of manual rules was developed for the conversion of OALD to SSAE pronunciations.

It was found that most of the differences between OALD and SSAE could be captured with four categories of rules, relating to (1) the KIT split, (2) syllabic consonants, (3) /z/ devoicing, and (4) /iy/ reduction. In the remainder of this section, these rules are described in more detail.

4.1. The KIT split

Dealing with the KIT split requires the most complex set of rules. In fact, we found that morphological analysis (MA) is required in order to predict the correct KIT pronunciations accurately in all cases. In the absence of a full MA implementation, direct graphemic analysis is utilised. While this approach compensates to an extent for the lack of MA, results obtained are less accurate than would have been possible if MA was utilised.

An interesting phenomenon identified in [9] relates to the concept of vowel harmony, where the pronunciation of one vowel is adapted in a specific context, in order to better approximate another. Vowel harmony occurs in many languages, but is not known to occur in English. However, a set of vowel harmony rules (see below) were found to be quite effective in predicting the KIT split accurately.

Our KIT rules systematically replace the phoneme $/ih/$ in OALD based on its graphemic origin and phonemic context. Pseudo-code implementing this transformation is included in the Appendix; it can be summarized with the following rules:

- A word-initial $/ih/$ is transformed to $/eh/$ if the orthography starts with eb-, em-, en- or ex-.
- $/ih/$ is retained in any of the following situations:
    1. It is the first or last phoneme of a word.
    2. It occurs within a closed syllable followed by phonemes such as $/sh/$, $/zh/$, $/ch/$, or phoneme sequences such as $/n\ sh/$.
    3. The corresponding grapheme that produced $/ih/$ is "i" or "y", and $/ih/$ is preceded or followed by velar phonemes such as $/k/$, $/g/$ or $/ng/$.
    4. The vowel harmony rules (as detailed in the Appendix) indicate the presence of $/ih/$.
- Otherwise, $/ih/$ is replaced with $/ax/$.

4.2. Syllabic consonants

In OALD, the symbols $/m/$, $/n/$ and $/l/$ represent both isolated phonemes and syllabic consonants. We prefer to insert explicit schwa symbols for the latter cases. The way this is done depends on the specific grapheme-to-phoneme (G2P) alignment algorithm used, and its accuracy. We identified a number of grapheme / phoneme combinations (shown in the Appendix) that reliably predict the presence of a syllabic consonant for the algorithms that we employ (as described in [11]).

4.3. /z/ devoicing

Devoicing of $/z/$ is a known phenomenon in SSAE. While this occurs in a number of situations, whether or not it is required is often not well understood. Two fairly reliable rules were defined, and are listed in the Appendix.

4.4. /iy/ reduction

The simplest of all the rule sets, this states that $/iy/$ at the end of a multisyllabic word is reduced to $/ih/$.

4.5. Application of manual rules

In order to apply the manual rules described above, two technologies are required: a G2P alignment component and a syllabification component. The G2P alignment is implemented using a version of the Viterbi algorithm which inserts both phoneme and grapheme nulls [11]. The syllabification component implements the syllabication rules defined in [12].

5. Accent mapping using decision trees

It has recently been found that the decision-tree techniques that are usually applied to G2P conversion can also be applied to P2P conversion, leading to improved pronunciation accuracies [7]. The application of decision-trees to G2P conversion is reviewed in the following section, after which the extension to P2P is described.

5.1. Grapheme-to-phoneme conversion

Decision trees have established themselves as a successful framework within which to perform G2P conversion [13]. Here the aim is to learn the mapping between the orthography and the pronunciation of a word based on a set of known examples. The mapping obtained in this way can then be used to derive pronunciations for new words.

For G2P, the graphemes and their context form the input and the phonemes of the pronunciation the output of the conversion process. Decision trees are therefore based on a one-to-one correspondence between graphemes (including context) and phonemes. Pronunciations are generated by sequentially passing the graphemes through the tree, and then concatenating the output phonemes. Discrepancies in length and alignment between the graphemes and phonemes are dealt with by inserting nulls into the phoneme string (for letters not corresponding to a phoneme), and by combining pairs of phonemes into pseudo phonemes for the few cases where one grapheme corresponds to two phonemes [14].

Each node of the decision tree is associated with a true/false question regarding the input. The tree is traversed from the root by recursively using the answer of each node's question. Each leaf node is associated with an output phoneme, which constitutes the tree's output [15].

Decision trees are trained recursively. For each new node, the available training data is split according to all possible questions. The question which results in the greatest information gain is then chosen. Information gain is the difference between a node's information entropy and the weighted entropy of its children [16], where information entropy is given by:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

For a node t with entropy $i(t)$, children t_L and t_R with respective entropies $i(t_L)$ and $i(t_R)$, and the proportion of node t's data associated with each child given by p_L and p_R such that both are non-negative and sum to 1, the entropy gain is [15]:

$$\Delta i = i(t) - p_L i(t_L) - p_R i(t_R) \qquad (2)$$

When choosing the question with the largest Δi, $i(t)$ can be omitted with no loss of generality as it remains constant for all questions at a given node. Furthermore p_k can be approximated by N_{t_k}/N_t for a child node t_k and $1/N_t$ is constant for all questions. Finding the maximum of Equation 3 therefore allows the optimal question to be found [17].

$$\Delta i = \sum_{\forall children\ k} \ \sum_{\forall phonemes\ p} N_{t_k,p} \log \left(\frac{N_{t_k,p}}{N_{t_k}} \right) \qquad (3)$$

For G2P conversion, decision trees using a context of 2 graphemes to the left, 3 to the right, and the 3 most recently generated phonemes were trained. These parameters were found to give optimal performance in preliminary experiments. G2P conversion took place from right to left. Clusters, used to find questions relating to groups of phonemes or graphemes, were automatically determined using the algorithm described in [17]. Trees were grown to their maximum size, and then pruned using a held-out dataset to improve generalisation [15].

5.2. Phoneme-to-phoneme conversion

G2P algorithms are extremely useful in situations where a certain number of pronunciations are available in the desired accent, and the requirement is to determine the pronunciations of a set of new words.

However when pronunciations are available not only in the desired (target) accent, but also in a different (source) accent, decision trees can be used to learn the mapping between the pronunciations directly. These rules can be used to map the pronunciations of all words in the source dictionary not already in the target dictionary. This is particularly useful when the number of pronunciations available in the source accent is larger than the number available in the target accent. It has been shown that additional pronunciations derived in this way are more accurate than pronunciations derived by G2P [7].

Since this process is highly analogous to G2P conversion, it is referred to as phoneme-to-phoneme (P2P) conversion. The phonemes of the source accent play the role of the graphemes in G2P, and the phones of the target accent the role of the phonemes. While there are relatively few cases in G2P of a single grapheme corresponding to multiple phonemes, there are considerably more cases where a single source phoneme corresponds to multiple target phonemes. This was dealt with by allowing the decision tree to use a larger number of pseudo phonemes (double phonemes treated as a single symbol).

6. Comparative analysis

In this section we begin by comparing the SSAE pronunciations derived from British English using the manually-created rules described in section 4, with the pronunciations determined by a human expert in SAEDICT. Subsequently, we will compare the accuracies of pronunciations derived using G2P and P2P algorithms with those obtained using the manually-derived rules.

6.1. Direct comparison

Table 1 compares the pronunciations present in the OALD dictionary directly with those present in SAEDICT. The results were obtained using 10-fold cross validation and the bootstrap confidence interval method [18]. This has been done to ensure consistency with the later G2P and P2P experiments, where the same approach will be applied. Phoneme accuracies are obtained by aligning corresponding phoneme strings, and applying Equation 4, where N_c, N_i and N_t are the numbers of correct, inserted and total phones respectively.

$$Acc = \frac{N_c - N_i}{N_t} \qquad (4)$$

Table 1 shows that 92.50% of the phonemes in the OALD and SAEDICT pronunciations correspond, and that 62.67% of the words in OALD and SAEDICT have the same pronunciation. Table 1 also compares the pronunciations obtained by applying the rules described in Section 4 to OALD, with the pronunciations in SAEDICT.

The results in Table 1 indicate that the application of the hand-crafted rules leads to an improvement in the correspondence with SAEDICT, with an additional 6% of words corresponding exactly in terms of pronunciation.

	Phoneme	Word
OALD vs. SAEDICT	$92.50 \pm 0.40\%$	62.67%
OALD-R vs. SAEDICT	$93.89 \pm 0.38\%$	68.33%

Table 1: Comparison of the OALD and OALD-derived (OALD-R) pronunciations with the corresponding pronunciations in SAEDICT.

A more detailed view of the above analysis can be obtained by reviewing the confusion statistics between OALD-R and SAEDICT. Of the 16 996 words, 11 485 agree with regard to pronunciation. Of the 5 511 remaining discrepancies, 4 183 are due to a single phoneme pair that differs per word.

Table 2 lists the main categories of such single-phoneme confusions. A possible cause for these differences is provided for the main categories, determined through manual inspection and random sampling. Here 'low' indicates that the cause is applicable in less than 10%, 'medium' in 10-40%, and high in more than 40% of cases. While most discrepancies relate to two pronunciations of a word that are both valid, the analysis also reveals errors in both OALD-R and SAEDICT.

Phoneme pair	Number occur.	Both acceptable	OALD-R error	SAEDICT error
ih:ax	1 406	medium	low	high
0:r	886	high	–	–
ax:ih	547	high	medium	–
s:z	366	high	low	low
z:s	229	high	low	low
iy:ih	224	high	medium	medium
aa:ax	75	high	medium	–
ae:ax	52	medium	high	–
aa:ao	36	high	–	–
eh:ax	27	high	–	–
0:p	24	high	–	low
ih:eh	22	high	medium	–
TOTAL	3894	-	-	-

Table 2: Examples of the categories of discrepancies between OALD-R and SAEDICT.

Table 2 shows that the number of occurrences per category decrease rapidly, with the last 30 categories containing only 1 occurrence each. While systematic errors can be found by analysing the main categories, smaller discrepancies and human error become more clearly visible when the tail is analysed. For example, when the last 40 categories are analysed, 16% and 27% of the words identfied contain errors in OALD-R and SAEDICT, respectively.

From an analysis of the confusion statistics obtained when comparing SAEDICT with both OALD and OALD-R, the following are observed:

- There is a considerable improvement in the correspondence of */ih/* and */iy/* vowels from OALD to OALD-R. Of all */ih/* vowels, 87% in OALD-R match those in SAEDICT, whereas this figure is only 73% for OALD. Similarly, 93% of OALD-R */iy/* vowels align with */iy/* in SAEDICT, compared to only 42% for OALD.
- It appears that */z/*'s which have been devoiced do not correspond reliably in OALD-R and SAEDICT. Almost as many */s/* phones in SAEDICT align to */z/* in OALD-R as in OALD (7% and 8% of */s/* phones respectively). The additional */s/* phones in OALD-R (i.e. those that have been devoiced by the rules) usually align with */z/* phones in SAEDICT. Examples of devoicing present in SAEDICT (relative to OALD) but not corrected in OALD-R include *holds* and *levels*, while examples of words devoiced in OALD-R that are voiced in SAEDICT include *lunches* and *acknowledges*.
- The majority of schwas inserted into OALD by the rules (for syllabic consonants) map to schwas in SAEDICT. The only other significant vowel to which these inserted schwas align is */ih/*.
- Both OALD-R and SAEDICT contain a number of errors that can be identified by analysing discrepancies between the two dictionaries.

6.2. Comparison with G2P and P2P

In order to evaluate the effectiveness of the G2P and P2P approaches in deriving SSAE pronunciations, these algorithms were applied to OALD, OALD-R and SAEDICT using 10-fold cross validation and the same data splits used to calculate the results in Section 6.1. For each of the 90:10 train/test segmentations:

- G2P decision trees were trained on the 90% of SAEDICT pronunciations reserved for training, and the accuracy determined on the remaining 10% test words. The same process was repeated for OALD and OALD-R.
- P2P decision trees were trained on the 90% of OALD and corresponding SAEDICT pronunciations. These decision trees were used to map the OALD pronunciations of the remaining 10% test words, and the result was compared with the corresponding 10% SAEDICT pronunciations to determine accuracy (OALD $\rightarrow$ SAEDICT). Note that in the P2P experiments, SAEDICT was viewed as the 'correct' pronunciation. This process was repeated for OALD-R (OALD-R $\rightarrow$ SAEDICT).

The results of the above processes are presented in Tables 3 and 4. Table 3 indicates that OALD and OLAD-R appear to have a greater internal consistency, resulting in higher G2P scores.

	Phoneme	Word
SAEDICT	$89.26 \pm 0.57\%$	56.29%
OALD	$89.87 \pm 0.58\%$	59.64%
OALD-R	$90.33 \pm 0.56\%$	60.86%

Table 3: G2P accuracies for SAEDICT, OALD and OALD-R.

By comparing Tables 1 and 4 it is clear that the hand-crafted rules result in higher accuracy than the application of G2P. Finally, it is also clear that automatic P2P rules provide the best performance, regardless of which dictionary is used as a source. However, Table 4 shows that there is no significant difference between OALD or OALD-R as source dictionary.

7. Conclusion

In this paper, manual and data-driven approaches to pronunciation dictionary development were evaluated. Firstly, the application of a small set of manually developed rules to convert a British English pronunciation dictionary to Standard South African English was contrasted with full manual development

	Phoneme	Word
G2P: SAEDICT	$89.26 \pm 0.57\%$	56.29%
P2P: OALD $\rightarrow$ SAEDICT	$96.03 \pm 0.33\%$	79.20%
P2P: OALD-R $\rightarrow$ SAEDICT	$95.81 \pm 0.33\%$	78.13%

Table 4: Accuracies for automatic rule-based derivations.

of a dictionary in the latter accent. Secondly, P2P rules extracted from various dictionary pairs were compared with regard to accuracy. Three main conclusions were reached:

1. When the British English pronunciation of the word in question is known, P2P learning provides a significantly more accurate mechanism for pronunciation prediction than G2P learning[1].
2. The set of manually developed rules applied to OALD approximated the manually developed dictionary suprisingly well, with a large number of discrepancies due to convention and acceptable pronunciation variations.
3. While OALD-R contains more errors than SAEDICT, both dictionaries can be improved through a detailed discrepancy analysis, as demonstrated in Table 2.

Further work related to this topic includes verifying SAEDICT based on the results obtained in this paper, and experimenting with additional G2P and P2P learning techniques. Specific techniques currently being considered include the use of a more flexible set of decision tree questions, and the possible application of the Default&Refine algorithm, previously shown to be effective when applied to similar tasks [19].

8. References

[1] R.I. Damper, Y. March, M.J. Adamson, and K. Gustafson, "Evaluating the pronunciation component of text-to-speech systems for English: a performance comparison of different approaches," *Computer Speech and Language*, vol. 13, pp. 155–176, April 1999.

[2] M.Davel and O.Martirosian, "Pronunciation dictionary development in resource-scarce environments," in *Proceedings of Interspeech*, Brighton, UK, Sept 2009, pp. 2851–2854.

[3] Korin Richmond, Robert A.J. Clark, and Sue Fitt, "Robust LTS rules with the Combilex speech technology lexicon," in *Proceedings of Interspeech*, Brighton, UK, Sept 2009, pp. 1295–1298.

[4] Ian Bekker, *The Vowels of South African English*, Ph.D. thesis, North-West University, Potchefstroom, South Africa, 2009.

[5] BEEP, "The British English Example Pronunciation (beep) dictionary," `ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries`.

[6] R. Mitten, "Computer-usable version of Oxford Advanced Learner's Dictionary of Current English," Tech. Rep., Oxford Text Archive, 1992.

[7] L. Loots and T.R. Niesler, "Data-driven phonetic comparison and conversion between South African, British and American English pronunciations," in *Proc. Interspeech*, Brighton, England, 2009.

[8] T.R. Niesler, P.H. Louw, and J.C. Roux, "Phonetic analysis of Afrikaans, English, Xhosa and Zulu using South African speech databases," *Southern African Linguistics and Applied Language Studies*, vol. 23, no. 4, pp. 459–474, 2005.

[9] O.M. Martirosian, *Adapting a pronunciation dictionary to SSAE for Automatic Speech Recognition*, Ph.D. thesis, North-West University, Potchefstroom, South Africa, 2009.

[10] Lwazi, "The Lwazi resource site," `http://www.meraka.org.za/lwazi`.

[11] M. Davel and E. Barnard, "The efficient creation of pronunciation dictionaries: machine learning factors in bootstrapping," in *Proceedings of Interspeech*, Jeju, Korea, 2004, pp. 2781–2784.

[12] T.A. Hall, "English syllabification as the interaction of markedness constraints," *Studia Linguistica*, vol. 60, pp. 1–33, 2006.

[13] K.-S.Han and G.-L. Chen, "Letter-to-sound for small-footprint multilingual tts engine," in *Proc. ICSLP*, Jeju, Korea, 2004.

[14] R.I. Damper, Y. Marchand, J.D. Marsters, and A. Bazin, "Aligning letters and phonemes for speech synthesis," in *Proc. 5th ISCA Speech Synthesis Workshop*, Pittsburg, USA, 2004.

[15] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*, Wadsworth & Brooks,, 1984.

[16] J. Suontausta and J. Häkkinen, "Decision tree based text-to-phoneme mapping for speech recognition," in *Proc. ICSLP*, Beijing, China, 2000.

[17] A.K. Kienappel and R. Kneser, "Designing very compact decision trees for grapheme-to-phoneme transcription," in *Proc. Eurospeech*, Aalborg, Denmark, 2001.

[18] M. Bisani and H. Ney, "Bootstrap estimates for confidence intervals in ASR performance evaluation," in *Proc. ICASSP*, Montreal, Canada, 2004.

[19] M. Davel and E. Barnard, "Pronunciation prediction with Default&Refine," *Computer Speech and Language*, vol. 22, pp. 374–393, 2008.

[1] For out-of-vocabulary words, whether British English G2P followed by British to South African English P2P is also more beneficial than SSAE G2P was not established.

9. Appendix: Manual adaptation rules

This appendix contains pseudo code describing the manually developed adaptation rules.

9.1. The KIT split

```
Assume the values of all current /ih/s
are unknown. Set to /?/.

IF word starts with ’eb-’, ’em-’, ’en-’
                                 or ’ex-’:
THEN change initial /?/ to /eh/

FOR all remaining /?/s in word:
  IF any of the following hold:
# velar lifting
    /?/ following after an /h/
      and generated by an ’i’ or ’y’
    /?/ before or after a /k/,/ng/ or /g/
      and generated by an ’i’ or ’y’
#palato-alveolar lifting
    /?/ preceding a /sh/, /zh/, /ch/, /jh/
    /?/ preceding a /n sh/, /n zh/, /n ch/
      or /n jh/
      and within a syllable
#word position lifting
    /?/ at end of word
    /?/ at beginning of word
    /?/ generated by word ending with
      ’-ies’, ’-ied’, ’-eys’
  THEN change /?/ to /ih/

#vowel harmony rules
FOR all remaining /?/s in word,
  starting at end of word,
  and working backwards:
  IF open syllable:
    IF any of the following hold:
      next vowel after /?/ is /ax/,
        and not adjacent
      next vowel after /?/ is /ey/
        and not adjacent
        and not first syllable in word:
      next vowel after /?/ is /ay/
          and not adjacent
          and not first syllable in word:
    THEN change /?/ to /ax/
    ELSE:
      change /?/ to /ih/

#default rule
FOR all remaining /?/s in word:
  Change /?/ to /ax/
```

9.2. Syllabic consonants

Note that the rules related to syllabic consonants are nfluenced by the specific g-to-p alignment algorithm used (and the accuracy of alignments obtained).

```
syllabic consonants = ’m’,’l’ and ’n’

IF word contains a phonemic null
  before or after syllabic consonant
  and null aligned to a vowel
  and any of the following patterns occur:
    /z 0+ m/
    /[d t k f v s z sh th zh] 0+ n/
    /[b d p t k g f v s z sh jh th m n] 0+ l/
    /[b d g s z t k p f]) l 0/
  and pattern at end of word
    or followed by consonants only
THEN
  insert a schwa
    before the syllabic consonant
```

9.3. /z/ devoicing

```
IF /[s sh jh ch] ax z/
  at end of word
  and orthography does not end
    on ’rs’ or ’res’
THEN
  change /z/ to /s/

IF /[aa uw iy] dh z/
  at end of word
THEN
  change /dh z/ to /th s/
```