The predictability of name pronunciation errors in four South African languages

Mpho Kgampe
Human Language Technologies Research Group
Meraka Institute, CSIR, Pretoria, South Africa
Email: mkgampe@csir.co.za

Marelie H. Davel
Multilingual Speech Technologies
North-West University, Vanderbijlpark, South Africa
Email: marelie.davel@gmail.com

***Abstract*—Personal names are often pronounced in very different ways depending on the language background of the speaker. We seek to determine whether some of these pronunciations 'errors' are systematic and if so, in which ways. Specifically, we analyze some of the the typical errors made by speakers from four South African languages (Setswana, English, isiZulu) when producing names from the same four languages. We compare these results with the pronunciations generated by four language-specific grapheme-to-phoneme (G2P) predictors trained on generic words from the four languages. We find that the G2P predictors are able to predict at least some of the typical errors humans make and, in fact, that these errors are slightly more predictable than the correct pronunciations themselves.**

I. Background

Various speech applications such as voice search, directory assistance or call routing systems rely on accurate name recognition in order to function optimally. For example, for a call to be routed to a recipient using a speech-driven call routing system, the system should be able to recognize the recipient's name as pronounced by the speaker initiating the call. Should the speaker pronounce the name in an unexpected way, recognition will not be possible and call routing will fail. In order for the call to be routed successfully, the various ways in which the name can be pronounced must therefore be predicted correctly.

Name pronunciation prediction is typically accomplished through a combination of grapheme-to-phoneme (G2P) conversion and dictionary look-up of known names. Personal names that do not match G2P conventions (and are deemed exceptions) quickly become very large [1]. Since personal names is such a large category of words, relying on lists of exceptions can become impractical. At the same time, extracting automated rules for these pronunciations is also difficult: the names themselves are diverse, and their pronunciations can differ substantially [2], [3]. In addition, many names are of a cross-lingual nature (for example English names originating from French) and may therefore incorporate a mix of spelling/pronunciation conventions.

Related work over a range of languages has shown that the language of origin is a relevant feature to consider when predicting pronunciations [2], [4]: more accurate predictions are possible if the original source language and the language of the speaker are both taken into account during pronunciation prediction. It has also been shown that a combination of G2P and phoneme-to-phoneme (P2P) conversion can improve the accuracy of proper name pronunciation [3], [5], and that G2P predictors can play a role in predicting additional variants for proper name recognition [6].

Personal name pronunciation prediction for South African names specifically, has received less attention. While first generation G2P converters exist for generic words in all South Africa's official languages [7], these were not developed to deal with proper names. In [8], a small data set of names in four South African languages (Afrikaans, English, isiZulu, Setswana) were analyzed for consistency when pronounced by speakers of the same four languages. Name pronunciations were recorded and transcribed, and based on the transcribed phone strings, the most probable within-language (L1) and cross-language (L2) pronunciations were determined. These were then used to analyze the consistency with which either of the two pronunciations were produced by speakers with the same language profile. It was found that within-language consistency was very high, while cross-language pronunciations were quite diverse. However, it was also found that speakers were more likely to produce the most probable cross-language pronunciation (the typical pronunciation of, for example, a Setswana name by an English speaker) than the pronunciation deemed to be L1 correct (the typical pronunciation of the same name by a Setswana speaker).

In this paper, the data from [8] is used to analyse the systematic differences between within-language and cross-language pronunciations. The aim is to better understand the reasons for the systematicity observed earlier. What are the types of differences between within-language and cross-language pronunciations? Does a small number of differences account for the systematicity of pronunciation errors, or are many (but consistent) differences observed? In addition, we are interested in understanding how these pronunciation errors relate to the typical errors produced by G2P systems.

The data set we use is fairly small (20 samples of each of 40 names, that is, 800 audio samples in total). In order to train a G2P/P2P system, a much larger data set would be required. In the current work, we therefore first aim to understand the types of pronunciation errors made by both human participants and existing G2P systems. We hope that this will inform the development of an efficient G2P/P2P approach, once a larger data set has been collected.

The paper is structured as follows: Section II describes the approach we follow in analyzing pronunciation errors; Section III contains our results; and the paper is concluded in Section IV with a summary of results and main findings.

II. Approach

To investigate the systematic errors speakers make when pronouncing personal names cross-lingually, we base our analysis on a corpus of name pronunciations gathered earlier. We analyze the differences in pronunciations, and compare the results from the human respondents with those obtained from four language-specific G2P converters. The main terms used in the rest of the paper coincide with those used in [8] and are listed in Table I.

Term	Definition
Primary language of speaker	The main language a speaker acquired while growing up: the language used by the speaker as a child to communicate with his/her primary care givers and other members of his/her immediate family. While a speaker typically has only one primary language, a speaker from a truly multilingual home may have more. Also referred to as L1.
Additional language(s) of speaker	Any language(s) spoken in addition to the primary language. Also referred to as L2.
Language community	A group of speakers sharing the same primary language.
Language of origin of name	The primary language spoken by the majority of people in the language community where the name was first used.
Pronunciation language of name	The language in which a name is pronounced. For example, the name *Elizabeth* may be pronounced as / i l @ z @ b @ T / in English and / E l i s a b E t / in Afrikaans (using X-SAMPA notation, as in the rest of this paper).
Correct pronunciation	The pronunciation(s) used by the majority of speakers from a specific language community, when producing a name from that language community.
L2 correct pronunciation	The pronunciation(s) used by the majority of speakers from a specific language (L2) community, when producing a name from another language (L1) community.

TABLE I
Definition of main terms used in this study, from [8].

A. Data

As main resource, we use the pronunciation corpus gathered in [8]. This corpus contains pronunciations by 20 primary language speakers of English, Afrikaans, isiZulu and Setswana of 40 personal names originating from language communities speaking these four languages. All speakers were between the ages of 20 and 45. Data was gathered in a controlled environment. Responses were elicited based on names presented visually, the pronunciations were recorded and manually transcribed. That is, recordings are phonemically transcribed by two individual transcribers, who listen to the recorded words one by one, and manually annotate each word with its perceived pronunciation. Where the transcribers disagree on a pronunciation, the disagreement is discussed and consensus reached. The third person then reviews the transcription and changes are made wherever necessary. These transcribed phoneme strings form the basis for further statistical analysis.

In addition, we use existing G2P converters to generate predicted pronunciations for all four languages. The Afrikaans converter was developed from an approximately 24k-word dictionary [9], the English converter was developed based on a large British English dictionary and a manually constructed P2P rule set [10], and the Setswana and isiZulu dictionaries from small 5k-word dictionaries [7].

B. Pronunciation alignment

When comparing two pronunciations, these are considered similar only when all phonemes match. If any of the phonemes differ, the pronunciations will not match and each specific phoneme that does not match its counterpart is considered to be a 'phoneme change'. Analysis consists of first aligning each target and observed pronunciation on a phoneme-by-phoneme basis and then analyzing the phoneme changes over groups of speakers. For example (using X-SAMPA notation), after aligning the pronunciations:

$$/\,\{\ d\ @\ m\,/ \rightarrow /\,a\ d\ a\ m\,/ \tag{1}$$

the following alignment is obtained:

$$/\{/ \rightarrow /a/$$
$$/d/ \rightarrow /d/$$
$$/@/ \rightarrow /a/$$
$$/m/ \rightarrow /m/$$

and from this alignment, the following changes are identified:

$$/\{/ \rightarrow /a/ \quad /@/ \rightarrow /a/$$

This alignment must also take possible differences in the number of phonemes into account. For example, aligning:

$$/\,d\ @\ r\ k\,/ \rightarrow\ /\,d\ E\ k\,/ \tag{2}$$

will result in:

$$/d/ \rightarrow /d/$$
$$/@/ \rightarrow /E/$$
$$/r/deleted$$
$$/k/ \rightarrow k$$

Automated alignment is performed using dynamic programming. This process maps the two phoneme strings to one another allowing insertions, deletions and substitutions. The dynamic programming process produces the alignment with the lowest cost, where the least number of insertions, deletions or substitutions are required to change one of the phoneme strings into the other.

C. Statistical analysis

We are interested in analyzing the type and frequency of differences between the L1 correct (or target) pronunciation and each pronunciation as produced by each speaker (the observed pronunciations); between the G2P system and the L1 correct pronunciation; and between each of the observed pronunciations and the G2P system.

When analyzing the data, we count the number of times a phoneme changes to another phoneme. We also look at the speaker languages where these changes occur. By analyzing the distribution of changes, it is possible to determine whether there are a small number of changes that occur repeatedly, or whether a large number of different changes are observed. Since the frequencies of phonemes observed in the target pronunciations are quite different, that is, some occur much more frequently and others less, both the exact number of times a change occurs and the percentage of times it occurs are of interest. While the percentage provides a better indication of the importance of the change, this may seem artificially high where the number of times the phoneme was observed in the target pronunciation is low.

Due to the experimental protocol, no target phoneme is observed less than 5 times in target pronunciations. The most frequently observed phoneme /U/, occurs 40 times. When considering systematic changes, only changes that occur more than 40% of the time for a single speaker language are listed. Confidence in the validity of the changes observed increases as the number of times a target phoneme observed increases. This latter value is therefore displayed whenever percentages are listed, as an indication of the possible validity of the observed change. In addition, if a change with regard to one language occurs with an average of 40% or more, additional languages where the same change is observed are also listed. These values may then be less than 40%.

As the size of the study sample is restricted, it is not possible to draw detailed conclusions with regard to all the cross-lingual phenomena that occur. The purpose of the analysis is therefore to provide an indication of the types and frequency of cross-lingual changes, within the scope of the experimental data available.

D. G2P prediction

We predict the pronunciation of each of the words in the combined word list (of 40 names) using the G2P rules of each of the languages individually. This results in 4 G2P-predicted pronunciations of every word. These pronunciations are analyzed further in the next section.

III. Analysis and Results

We first investigate the distribution of the various differences between target and observed pronunciations. We then list the systematic differences taking place per speaker language. We focus on changes that are statistically significant but then also list separately some changes which are not statistically significant but may be linguistically plausible. Finally, we consider the difference between the G2P predicted pronunciations and both the L1 correct pronunciation and the typical errors respondents made.

A. Distribution of differences

In this section, the distribution of observed differences are considered per speaker language. By analyzing the distribution of changes, it is possible to determine whether there is a small number of changes that occur repeatedly or whether a large number of different changes are observed. As before, both the exact number of times a change occurs and the percentage of times it occurs are considered. Phoneme changes are considered across the different name languages but per individual speaker language.

By comparing the phoneme changes that occur per language for Afrikaans, English and isiZulu we find fairly similar trends. This includes a single change that dominates the others, a few changes that occur very frequently and a long tail of changes occurring only once or twice. Setswana has a smoother distribution with less very frequent changes and less changes overall.

When the same information is viewed in terms of percentages (the number of times a change occurs as a percentage of the number of times the target phoneme was observed) all languages behave almost similarly, with a small set of changes occurring 40% or more of the time, and a long tail of changes with low percentage values. This is shown in Figure 1.

B. Types of differences observed

This section investigates the main systematic changes observed, additional changes that are of linguistic interest and the changes observed on a name/speaker language pair bases.

1) Systematic differences: Table II lists the systematic differences between target and observed pronunciation deemed to be significant when speakers pronounce names within language and cross-lingually. Here we illustrate the number of times a specific phoneme occurs, that is, the correct phoneme and the phoneme the correct one is changed to. These changes occur within specific speaker languages and the percentage averages that occur are listed.

TABLE II
Systematic differences in the way L2 speakers approximate the L1 correct pronunciation

no. of phons	correct phoneme	changed to	Speaker language			
			A	E	S	Z
40	U	O	70%	68%	18%	75%
25	I	E	52%	56%	12%	56%
10	{	a	20%	10%	60%	70%
10	i:	i	-	-	60%	70%
10	p_h	p	70%	40%	-	-
10	p_h	f	10%	40%		
5	tl_>	t	-	80%	-	-
5	kx	x	-	-	-	60%
5	ts_h	tS	-	-	-	60%
5	Q	O	-	-	100%	80%
5	Oi	i	-	-	20%	60%
5	ai	i	-	-	40%	-
5	ai	@i	20%	40%	40%	100%
5	ai	E	52%	56%	12%	56%
5	r\	r	40%	-	100%	60%
5	J	j	60%	-	40%	-
5	|\g_0	|\	20%	20%	60%	-
5	|\g_0	d_0Z	-	40%	-	-

Considering the alveolar approximant /r\/ which is an English phoneme, we see that Setswana speakers changed it to an alveolar thrill /r/ 100% of the time, isiZulu speakers changed it 60% of the time and Afrikaans speakers with an average of

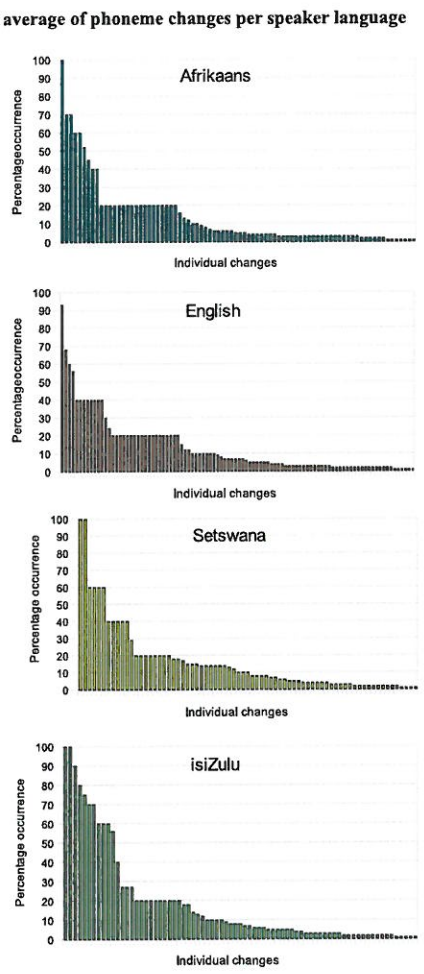

Fig. 1. *Average percentage of phoneme changes*

40%. English speakers are familiar with pronouncing /r\/; this is evident from the absence of any changes made within the speaker language (English).

The same can be said for the long vowel /i:/ which is changed by an average of 60% of isiZulu speakers and by an average of 70% of Setswana speakers. The diphthongs /Oi/ and /ai/ are changed by isiZulu speakers by an average 60% and 100% respectively (/ai/ change to more than one phoneme). Setswana speakers also changed both phonemes, however, with lower averages. /Oi/ was changed 20% and /ai/ 40% of the time. The long vowel /i:/ and the diphthongs given as examples are of English origin, hence it is plausible that changes are made by isiZulu and Setswana speakers and not by English or Afrikaans speakers.

In [8] it was found that many of the pronunciation discrepancies in Setswana were caused by the Setswana seven-vowel system. In this experiment, E/I and O/U are often mispronounced: /U/ is changed to /O/ and /I/ changed to /E/ by speakers of all languages (including a minority of Setswana speakers). The four vowels present pronunciation variation not only cross-lingually, as is mainly the case with other languages, but within language as well. These results may be influenced by the difficulty in transcribing these vowels accurately. The phonemic transcription becomes a difficult task when more that one language is involved, the transcriber may not be a primary language speaker of other languages.

It is also observed that the voiced dental click sound /|\g_0/, which is an Nguni phoneme, is changed: Firstly, to a different click by Setswana speakers with an average of 60% and Afrikaans and English speakers with an average of 20% each; and secondly, to an affricate /d_0Z/ by English speakers with an average of 40%. Other observations relate to the stops /p_h/, /tl_>/ and /x/, the affricates /kx/ and /ts_h/ the palatal nasal /J/ and the vowels (see Table II).

2) Other differences: Table III lists additional changes that are of interest. These occur with lower percentages (averages of 20% or lower) but were predictable from a linguistic perspective. The currently available data set does not confirm or deny the validity of these changes: that would require a larger study to ascertain.

TABLE III
Examples of changes observed that are not statistically significant but linguistically plausible.

no. of phons	correct phoneme	changed to	Speaker language			
			A	E	S	Z
60	E	I	7%	10%	9%	2%
60	E	i	4%	5%	-	-
60	E	@	-	4%	-	-
58	@	a	-	-	15%	12%
58	@	i	-	-	3%	8%
58	@	E	-	-	10%	4%
25	I	@	16%	12%	%	%
20	x	g	20%	15%	-	-
10	i:	@	-	-	20%	-
10	i:	I	-	-	20%	-
5	|\g_0	!\	20%	-	-	-
5	r\	j	20%	-		20%
5	tl_>	K	20%	-	-	-

It is interesting to note the following:

- The phoneme /@/ changes into more than one phoneme. For example it changes to /a/, /i/ and /E/. This is an indication of the difficulty of pronouncing the phoneme, unfamiliarity with it and/or an error made when pronouncing it. This is also evident for the phoneme /i:/.
- In the speaker language Afrikaans, 20% of speakers changed the Sotho-Tswana phoneme /tl_>/ to the Nguni fricative /K/. This is likely to occur when a speaker is not familiar with the name language's orthography and sound, leading to the confusion of the two name languages concerned.
- Afrikaans and isiZulu speakers changed /r\/ to /j/ 20% of the time. Some speakers changed it to a phoneme /r/

as previously indicated in Table II.

- 20% of Afrikaans speakers resorted to the dental click /!\/ as they were aware that /|\g_0/ is a click sound, however, they were not sure which one and/or did not know how this particular one is pronounced.
- The phoneme /x/ changes to /g/ because orthographically the two phonemes are similar. The speaker may confuse one with the other.
- The difficulty of pronouncing the Setswana /I/ is again evident, with Afrikaans and English speakers changing it to /@/. The phoneme was changed to /E/ in both speaker languages in Table II.

TABLE IV
Examples of systematic cross-lingual pronunciation errors observed

Name language \ Speaker language	A	E	S	Z
A	-	ai→@i r→ r\ O→Q	ai → i a → E @ → i	@u→O x→g @→E
E	O→u Q→O r\ →r	-	@ → a Q → O r\ →r	{ → E Oi → O j i r\ →r
S	U→O I→i/E p_h→p	p_h→f ts_h→tS I→@/E	-	kx → x U → O I → E
Z	J → j O → u a → A:	E→i a→{ O→u	a → E |\g_0 → |\ E→I	-

3) A language-pair perspective: A larger percentage of similar transformations are observed across different name languages if the speaker language is kept constant, than if the speaker language is varied and the name language kept constant. While we therefore focus our attention on differences that occur per speaker language, it is also interesting to view some of these changes from a language pair perspective, as shown in Table IV. Observing the speaker language Afrikaans, we find that /O/ changes to /u/ when the name language is English, and also when the name language is Setswana. In the speaker language Setswana, we observe the same changes with /a/ changing to /E/ when the name languages are Afrikaans and isiZulu, and also with the speaker language Setswana. For example, /@/ changes to /i/ when the name language is Afrikaans and it changes to /a/ when the name language is English. /@/ is changed to different phonemes as observed in Table III.

C. G2P analysis

We now consider the extent to which automated G2P converters create similar or different errors to those observed in human respondents. We use two measures of pronunciation similarity: phoneme correctness and phoneme accuracy. In both cases the evaluated and reference pronunciations are aligned with one another on a per phoneme basis. When measuring phoneme correctness we calculate the number of aligned phonemes that match exactly and express that number as a percentage of the phonemes in the aligned pronunciation; when measuring phoneme accuracy we first subtract insertions before calculating this percentage.

1) G2P accuracy: As a first step we compare phone recognition accuracy of the G2P converters with that of the human respondents. Table V lists phoneme recognition correctness and accuracy in three different situations: (a) when comparing speaker pronunciations with the L1 correct pronunciation, (b) when comparing G2P prediction with the L1 correct pronunciation and (c) when comparing speaker pronunciations with the G2P prediction.

From the results it is clear that the standard G2P prediction is not very accurate – obtaining between 52% and 69% accuracy – compared with 85% to 92% accuracy observed among human respondents. It should be noted that the method whereby the correct pronunciation is determined (by majority vote) favours the latter measurement. Still, it is very interesting to note that the G2P converter is better at predicting the errors respondents make (59% to 73%) than at predicting the actual L1 correct pronunciation.

TABLE V
Phone recognition accuracies when comparing (a) speaker pronunciations with L1 correct (b) G2P predictions with L1 correct pronunciations and (c) speaker pronunciations with G2P predictions

	language	correctness	accuracy
(a)	A	92.4	91.8
	E	89.7	89.2
	S	91.6	91.6
	Z	85.8	84.8
(b)	A	59.8	51.9
	E	54.0	50.2
	S	73.6	69.0
	Z	70.7	62.8
(c)	A	64.9	58.6
	E	62.2	60.6
	S	78.0	73.1
	Z	81.7	71.8

2) Comparing G2P predictions and L1 correct pronunciation: We investigate the differences between G2P-predicted pronunciations and the L1 correct pronunciation. Not all phoneme changes are listed, rather, we focus on the phonemes that were analysed in Section III-B1. Observing Table VI, we see that, for example, /p_h/ was changed to /f/ by the English G2P rules and to /p h\/ by Afrikaans G2P rules. Table II illustrates that both Afrikaans and English speakers changed /p_h/ to /f/; the phoneme did not change to /p h\/ at any stage, but rather to /p/ in both speaker languages.

Another example is the phoneme /|\g_0/, which was consistently changed to /|\/ by Afrikaans, English and Setswana speakers. As the target phoneme does not occur in any of these languages, the G2P systems fail completely, generating phonemes such as /s/ (Afrikaans and English) and /k/ (Setswana). /U/ was changed to a different phoneme by the G2P rules of different languages, as seen in Table VI.

3) Comparing G2P predictions and speaker errors: Table VII lists the differences observed between the G2P rules and speakers when pronouncing cross-lingual personal names. Not all possible occurrences are listed, only those that are of

TABLE VI
Differences between G2P and the L1 correct pronunciation

no. of phons	correct phoneme	changed to	Speaker language			
			A	E	S	Z
8	U	O	50%	-	100%	88%
8	U	u@	-	50%	-	-
8	U	u	38%	-	-	-
9	E	@	67%	34%	-	-
12	@	a	9%	-	-	-
12	@	r	17%	-	17%	-
12	@	i	-	-	34%	34%
12	@	E	-	9%	25%	25%
1	ai	@i	100%	-	-	-
1	ai	@	-	100%	-	-
1	ai	i	-	-	100%	100%
1	Oi	a	100%	-	-	-
1	Oi	i	-	-	100%	100%
5	I	@	60%	-	-	-
5	I	@i	20%	-	-	-
5	I	{	20%	-	-	-
5	I	E	-	40%	100%	100%
5	I	i		20%	-	-
5	I	i:	-	20%	-	-
1	kx	x	100%	-	-	-
1	kx	g	-	100%	-	-
4	x	k	25%	-	-	-
2	i:	@	50%	-	-	-
2	i:	i@	50%	-	-	-
2	i:	E	-	-	50%	50%
2	i:	i	-	-	50%	50%
1	|\g_0	s	100%	100%	-	-
1	|\g_0	k	-	-	100%	-
1	r\	r	100%	-	100%	100%
1	tl_>	t l	100%	100%	-	100%
2	p_h	p h\	100%	-	-	-
2	p_h	f	-	100%		

TABLE VII
Differences between G2P and observed pronunciations

no. of phons	correct phoneme	changed to	Speaker language			
			A	E	S	Z
16	O	u	44%	100%	-	-
5	O	Q	-	60%	-	-
18	a	@	-	45%	-	-
12	@	r	17%	-	9%	-
8	@	i	-	-	34%	34%
8	@	E	-	9%	25%	25%
1	ai	@i	100%	-	-	-
1	ai	@	-	100%	-	-
2	@i	i	-	-	25%	38%
2	@i	E	-	-	25%	-
1	@i	ai	-	100%	-	-
1	@i	@i	100%		-	-
1	Oi	a	100%	-	-	-
1	Oi	i	-	-	100%	-
1	Oi	Oi	-	100%	-	-
2	I	@	100%	-	-	-
6	I	E	-	-	100%	% -
3	x	g	-	-	-	100%
2	i:	@	50%	-	-	-
3	r	r\		100%	-	-
1	r\	r	100%	-	-	-

interest based on previous analysis. For example, looking at speaker language English, we find that the phoneme /r/ is changed to /r\/ in all three occurrences.

IV. Conclusion

In this paper, we investigated the pronunciation errors made by speakers when producing personal names across languages. We analysed the differences between the single L1 correct pronunciation, the various pronunciation errors respondents make, and the errors made by four G2P converters.

We observed various differences within name/speaker language pairs and found that the speaker language variable is more important than the name language variable, as many errors occur in speaker languages irrespective of name language. Interestingly, we found that G2P predictors are better at approximating the behaviour of speakers in general, than at approximating the true L1 correct

From our analysis it seems that two very different types of errors are made: one where the orthography is correctly interpreted but wrongly produced by the speaker (such as the English phoneme /r\/ produced as /r/) and one where the orthography is wrongly interpreted (such as the Setswana /p_h/ produced as /f/ by English speakers.) The former error is easy to capture in a P2P system, while the latter can be modelled using a G2P approach. This bodes well for the development of a G2P/P2P approach that is applicable to the prediction of South African personal names. In order to develop such a system we would require a significantly larger data set; development of such a corpus is currently in process [11].

References

[1] M. F. Spiegel, "Proper name pronunciations for speech technology applications," *International journal of speech technology*, no. 6, pp. 419–427, 2003.

[2] A. Font Llitjós, "Improving pronunciation accuracy of proper names with language origin classes," Master's thesis, Department of Computer Science, University of Edinburgh, 2001.

[3] H. van den Heuvel, J.-P. Martens, and N. Konings, "G2p conversion of names. what can we do (better)?" in *Interspeech*, Antwerp, Belgium, 2007, pp. 1773–1776.

[4] B. Réveil, J.-P. Martens, and B. D'hoore, "How speaker tongue and name source language affect the automatic recognition of spoken names," in *Proceedings Interspeech*, Brighton, 2009, pp. 2995–2998.

[5] N. K. Q. Yang, J.-P. Martens and H. van den Heuvel, "Development of a phoneme-to-phoneme (p2p) converter to improve the grapheme-to-phoneme (g2p) conversion of names," in *Proceedings LREC*, Genua, 2006, pp. 287–292.

[6] H. van den Heuvel, B. Réveil, and J.-P. Martens, "Pronunciation-based ASR for names," in *Proceedings Interspeech*, Brighton, 2009, pp. 2991–2994.

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

[8] M. Kgampe and M. Davel, "Consistency of cross-lingual pronunciation of South African personal names," in *Proc. Pattern Recognition Association of South Africa (PRASA)*, Stellenbosch, South Africa, 2010, pp. 123–127.

[9] D. M. H and F. de Wet, "Verifying pronunciation dictionaries using conflict analysis," in *Proc. Interspeech*, Makuhari, Japan, Sept. 2010, pp. 1898–1901.

[10] L. Loots, . D. M, E. Barnard, and T. Niesler, "Comparing manually-developed and data-driven rules for p2p learning," in *Proc. Pattern Recognition Association of South Africa (PRASA)*, Stellenbosch, South Africa, Nov 2009, pp. 127–300.

[11] O. Giwa, M. H. Davel, and E. Barnard, "A Southern African corpus for multilingual name pronunciation," in *Proc. Pattern Recognition Association of South Africa (PRASA) (accepted for publication)*, Johannesburg, South Africa, Nov 2011.