Consistency of Cross-lingual Pronunciation of South African Personal Names

Mpho Kgampe and Marelie H. Davel
Human Language Technology Competency Area
CSIR Meraka Institute
Pretoria, South Africa
Email: mkgampe@csir.co.za,mdavel@csir.co.za

Abstract—**We investigate the consistency with which speakers with different language profiles are able to pronounce personal names from Afrikaans, English, Setswana and isiZulu. We gather data in a controlled research study and analyse cross-lingual pronunciation effects. We find that speakers with a similar primary language tend to agree on the 'correct' pronunciation of a name originating from their own language community, and that the ability of speakers from other language communities to approximate this pronunciation is highly dependent on the speaker-word language pair. We also find that there are systematic ways in which names are 'mis-pronounced' by different language communities: understanding such systematicity could be important when extending electronic pronunciation dictionaries (used in spoken dialogue systems) with the most important variants that occur in practice, in order to increase the accuracy of name recognition.**

I. Introduction

One of the core components of a spoken dialogue system (SDS) is its pronunciation predictor: a component that is able to predict how a word will be pronounced based on its orthography. Producing accurate pronunciations of proper names is a challenging task, and even more so if the SDS needs to function in a multilingual or cross-lingual environment. For example, a South African call routing system should be able to handle cross-lingual pronunciations of proper names (when a name originating from one language is produced by a speaker with a different language background). When producing proper names cross-lingually, a speaker may have limited or no proficiency of the language the name originates from, and may be unfamiliar with the name itself, resulting in seemingly unpredictable pronunciations.

In this study we investigate the consistency with which speakers with a specific language profile are able to pronounce personal names from other languages. Specifically we aim to determine the extent in which speakers with a specific language profile agree on the 'correct' pronunciation of a name originating from their own language community, and the extent in which speakers from other language communities are able to approximate this pronunciation. We also seek to determine whether there are systematic ways in which these names are 'mis-pronounced' by different language communities. This is important, since if such systematic patterns can be identified, these can be used to extend electronic pronunciation dictionaries (used in SDSs) with the most important variants that occur in practice, and so increase the accuracy of name recognition.

We focus on first language speakers of English, Afrikaans, isiZulu and Setswana and personal names originating from language communities speaking these four languages. We elicit controlled responses from 20 subjects between the ages of 20 and 45. In order to analyse detail effects more thoroughly, a study more focussed on specific pronunciation phenomena would be required: our aim here is to determine whether there is sufficient consistency in cross-lingual name pronunciation to warrant further investigation. In this study, we aim to answer the following specific questions with regard to the pronunciation of personal names in the four languages mentioned:

- How easily can the 'correct' pronunciation of a name be identified?
- What percentage of speakers produce the correct pronunciation?
- Are cross-lingual pronunciations less accurate than within-language pronunciations?
- Does there seem to be systematicity in the errors made when producing cross-lingual pronunciations?

The paper is structured as follows: Section II provides a brief overview of the proper name recognition task, and describes related work. Section III describes our approach to studying cross-lingual proper name pronunciation and defines the terms used in the rest of the paper. Section IV provides a detailed description of the experimental process, and Section V contains our analysis and results. In Section VI our results are summarised and further work discussed.

II. Background

Most speech technology systems use two mechanisms for pronunciation prediction: explicit pronunciation dictionaries containing lists of word-pronunciation pairs and letter-to-sound converters used to predict out-of-vocabulary words. While this combined approach tends to work well for general words, personal name pronunciation is particularly difficult for a number of reasons: spelling of names may be irregular, there is an inexhaustible supply of new names, and many names are of a cross-lingual nature (for example, English names that were originally considered French).

While the number of proper names exceeds the number of any other type of word occuring in practice [1], existing pronunciation dictionaries typically contain limited entries of this category. For example, the Jones English pronunciation dictionary [2] contains 56 300 words, of which only 13 927 are proper names. A wealth of proper name pronunciations are contained in commercial name databases which can easily contain 200 000 to 500 000 entries [3]. Additonal databases are also becoming available for research purposes [4]. Related work indicates that the language of origin of a proper name is important in being able to predict its pronunciation accurately [5], and that language-specific letter-to-sound predictors can play a role in predicting additional variants for proper name recognition [6]. We are not aware of any studies related to the cross-lingual pronunciation of personal names for speech technology purposes relating specifically to any of the South African languages.

III. Approach

A. Overview

In order to better understand the consistency with which personal names are produced, we gather data from first language speakers in a controlled environment. We elicit responses based on names presented visually, record these and transcribe the responses manually. The transcribed phoneme strings form the basis for further statistical analysis.

B. Terms and definitions

In order to contextualise our approach, it is useful to define a number of terms explicitly. We have already alluded to the difficulty of deciding what the 'correct' pronunciation of a name is. We therefore provide our definition of this and other terms in Table I.

C. Variables considered

We anticipate that the factors mentioned above will all have an effect when considering cross-lingual proper name pronunciation.

The specific variables that we consider during our experimental design are listed in Table II.

In the current set of experiments, the following variables are either controlled, independant or measured (dependant):

- **Controlled**: Age of speaker, level of education of speaker and name ambiguity.
- **Independant**: Primary language of speaker, relevant pronunciation languages and language of origin of name.
- **Measured**: Speaker's pronunciation of name, speaker's knowledge of language of origin of name, speaker's familiarity with name, correct pronunciation and L2 correct pronunciation.

IV. Experimental process

In this section we describe the process followed during data collection, with regard to name selection (Section IV-A), respondent selection (Section IV-B), recording of responses (Section IV-C) and transcription of responses (Section IV-D).

TABLE I
Definition of terms.

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 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.
Name familiarity	An individual has encountered the name before, specifically by hearing or speaking it.
Name ambiguity	A name has more than one correct pronunciation (such as the name *Jean* which can be pronounced as either / d_0Z i: n / or / d_0Z A: n /).

TABLE II
Variables considered.

Type	Variable
Speaker	Primary language of speaker
	Additional languages of speaker (and level of exposure per language)
	Age of speaker
	Level of education of speaker
Name	Language of origin of name
	Name ambiguity
	Relevant pronunciation languages
Name & pronunciation language	Correct pronunciation
	L2 correct pronunciation (per L2)
Speaker & name	Speakers's knowledge of the language of origin of name
	Speaker's familiarity with name
	Speaker & Pronunciation of name per pronunciation language

A. Name selection

Twenty full names (each consisting of first name and surname) were selected from known lists of names on the Internet. Of the resulting forty names, ten names originated from each language forming part of the study. Names selected were mostly of writers, poets and other individuals involved in literature, both from South Africa and Botswana. Once the lists were constituted, these were verified by primary language speakers. The names included some that were group specific, such as *Khulile Nxumalo*, which is a general Nguni name and not necessarily only an isiZulu name. First names and surnames were separated and randomly recombined within a language group in order to prevent known pronunciations of known names (such as *Elisabeth Eybers*) influencing results.

B. Respondent selection

We select 20 respondents, five each a primary language speaker of English, Setswana, isiZulu and Afrikaans, all between the ages of 20 and 45, all with at least a high school qualification, and all currently residing in the Gauteng area. We only select respondents who have a single primary language. For each respondent, the following information is captured (after the recording session has been completed): primary language, competency in additional languages, highest qualification and place obtained.

C. Recording

Each recording session consists of a number of sections.

1) Natural pronunciations: Each respondent is asked to pronounce a set of name-surname pairs as naturally as possible. Word pairs are presented one at a time. Respondents pronounce the name (and are allowed to repeat the name if necessary) and continue to the next name only when ready.

2) Language of origin: Each respondent is asked to provide their opinion as to the language of origin of a name pair, by selecting either a specific language, one of two language groups, or 'uncertain'.

3) Forced pronunciation language: Each respondent is asked to pronounce a set of name-surname pairs using a specific pronunciation language. Specifically, respondents are asked: *If you thought X was a Y name, how would you pronounce it?* Not all possibilities are recorded: only some names and some pronunciation languages are selected for this section. (A total of 90 words are recorded per respondent.)

4) Name familiarity: Once all responses have been completed, each respondent is asked to indicate whether he/she was familiar with a name prior to the study.

D. Transcription

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. The phoneme set used for the transcriptions is a based on the Lwazi phoneme sets [7], one of which exists for each of the eleven official languages of South Africa. The individual phoneme sets for the four languages studied are combined into a single extended phoneme set. This set is simplified in only one way: plosives that have 3 versions (for example, the standard /k/, aspirated /k_h/ and ejective /k_>/) are reduced to two versions, an aspirated and non-aspirated one.

Cross-lingual phonemic transcription is not a simple task: features that are phonemic in one language (differnt realisations change the meaning of the word) and not in another, are often not clearly realised in the second language. For example, the duration of vowels play a phonemic role in the English vowel pairs: /u/ and /u:/, /i/ and /i:/, and /O/ and /O:/. Since duration is not phonemic in isiZulu, a word such as 'zulu' would typically be transcribed as /z u l u/ in monolingual transcriptions, irrespective of the length of the vowels. Also, such vowels are often not clearly either an /u/ or an /u:/ but can occur anywhere on the continuim in between. How should they be transcribed? The same issue is encountered with regard to aspiration (phonemic for certain Setswana and isiZulu sounds, but not in English or Afrikaans) and with regard to the 7-vowel system used in Setswana, rather than the 5-vowel system of the other 3 languages. In order to deal with such issues consistently, we transcribe sounds that occur in between two valid phonemes with the version closest to the phoneme most frequently realised by the speaker. (For example, an English speaker trying to produce a Sepedi /I/, but not realising it accurately, would be transcribed as either an /E/ or an /i/, depending on which one of the latter to vowels is closest.)

Where a respondent produced more than one pronunciation for a single question, only the second pronunciation is used during analysis. Where the transcribers disagree on a pronunciation, the disagreement is discussed and consensus reached.

V. Analysis and Results

In total, 2 600 words were recorded and transcribed. 130 words were recorded per respondent: 40 natural pronunciations and 90 pronunciations using a forced pronunciation language. The 20 respondents analysed consisted of 5 Afrikaans speakers, 5 English speakers, 5 Setswana speakers and 5 isiZulu speakers. While additional information is contained within the data gathered, we focus our analysis in this section on answering the specific questions posed in Section I.

A. Within-language consistency

We first evaluate the consistency with which speakers from a specific language community produce pronunciations for names originating from that community. We find that the correct pronunciation is fairly easily identifiable, and that agreement among speakers is high, as shown in Table III. Here we calculate the percentage of L1 speakers who agree on a single pronuniciation of a specific L1 word (when producing a natural pronunciation), and average over all L1 words. Only pronunciations that are fully in agreement are counted. (If a single phoneme is different, the full pronunciation is deemed to be different.)

TABLE III
Percentage of L1 speakers producing the correct L1 pronunciation, averaged over all L1 words.

Language	**% correct**
A	78%
E	78%
S	68%
Z	92%

The isiZulu pronunciations were found to be most consistent, and Setswana pronunciations least. Most of the Setswana discrepancies relate to different choices made by speakers when pronouncing the 'e' letter as either /E/ or /I/, and the 'o' letter as either /O/ or /U/. English differences mainly relate to different choices when pronouncing /O/ or /Q/, as well as the use of different combinations of /a/, /A:/, /{/ and /@/. Afrikaans differences in pronunciation is mostly caused by speakers producing either an English or Afrikaans version of a specific name (irrespective of the fact that both the speaker and the name are Afrikaans).

B. L2 proficiency in approximating L1 pronunciations

Once a correct pronunciation has been identified, we can determine the proficiency of L2 speakers in approaching this pronunciation. We find that the ability to produce cross-lingual pronunciations differ substantially depending on the L1-L2 language pair, as shown in Table IV, and is generally quite poor. Here we measure the percentage of L2 speakers who produce the exact correct pronunciation identified above (again, only when producing natural pronunciations), and average over all words (per language). For example, the table indicates that Setswana speakers produce isiZulu names very well (names are correctly produced 62% of the time), while English names are only pronounced correctly 32% of the time by the same Setswana speakers.

TABLE IV
Percentage of L2 speakers producing the correct L1 pronunciation, averaged over names.

		Speaker language			
		A	**E**	**S**	**Z**
Name language	**A**	-	38%	24%	34%
	E	56%	-	32%	34%
	S	18%	10%	-	22%
	Z	44%	32%	62%	-

C. L2 consistency

As could be seen from Table IV, some of the language pairs (such as Setswana speakers of English names) indicate low proficiency in producing the L1 correct pronunciation. For these pairs, how consistent are the pronunciations within a language community, or are the mistakes ad hoc? We investigate this by calculating the percentage of L2 speakers who agree on a single pronuniciation of a specific L1 word (when producing a natural pronunciation), and average over all L1 words. These results, as listed in Table V, show that there is consistency in the errors being made. In fact, by comparing Tables IV and V, it can be seen that speakers more consistently produce the L2 correct pronunciation, than the actual (L1) correct pronunciation.

D. Systematic effects

From Table V it can be seen that some of the pronunciation effects are indeed systematic within speaker communities. While a detailed analysis of these effects are outside the scope of this paper, we list examples of systematic differences observed in Table VI.

TABLE V
Percentage of L2 speakers producing a consistent L2 pronunciation, averaged over names.

		Speaker language			
		A	**E**	**S**	**Z**
Name language	**A**	-	54%	48%	56%
	E	62%	-	68%	50%
	S	54%	42%	-	54%
	Z	52%	48%	68%	-

TABLE VI
Examples of systematic cross-lingual pronunciation errors observed.

		Speaker language			
		A	**E**	**S**	**Z**
Name language	**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	-

E. Utilising 'forced' pronunciations

How natural are the pronunciations generated when humans are forced to pronounce a word in a different pronunciation language, and can this also assist us in uncovering systematicity in pronunciation errors? In order to determine this, we investigate the extent in which the pronunciations of speakers of language X forcing themselves to pronounce a word in language Y approximate the type of errors made by language Y speakers, when producing words in language X. Table VII indicates that there is also significant consistency in the way in which speaker generate forced pronunciations, and that the concept of a 'pronunciation language' helps to guide how words are pronounced. Interestingly, many of the systematic changes (some examples of which are given in Table VI) were also observed in the forced pronunciations. These observations will form the basis for further statistical analysis of the data.

TABLE VII
Consistency of forced pronunciation per pronunciation language, word language and speaker language

Pron language	**Word language**	**Speaker language**			
		A	**E**	**S**	**Z**
A	E	66%	54%	70%	44%
E	A	56%	48%	50%	48%
S	Z	54%	38%	60%	46%
Z	S	58%	36%	66%	46%

VI. Conclusion

In this paper we analysed the consistency with which speakers with a specific primary language are able to pronounce personal names from Afrikaans, English, Setswana and isiZulu. We performed a controlled research study using 20 subjects (2 600 recorded words) and evaluated the consistency with which pronunciations were produced.

We find that speakers with a similar language profile tend to agree on the correct pronunciation of a name originating from their own language community, and that the ability of speakers from other language communities to approximate this pronunciation is highly dependent on the language pair. We also find that the concept of a 'L2 correct pronunciation' is very strong, with L2 speakers more consistently producing this pronunciation, than the 'correct' L1 pronunciation. A number of systematic substitutions are identified (some examples of which are shown in Table 6).

In future work we aim to verify the accuracy of our transcripts automatically. As cross-lingual transcription is a difficult task and prone to human error, we would like to use automated classification techniques to assist us in flagging transcription errors. A more detailed analysis of systematic pronunciation variation (on phoneme level rather than word level) would then be possible. Such a detailed analysis of systematic substitutions could be useful to determine whether extending electronic pronunciation dictionaries with cross-lingual variants could increase the accuracy of spoken name recognition in multilingual environments.

Acknowledgment

The authors would like to thank Mpho Raborife and Etienne Barnard, who kindly assisted with transcription and transcription verification; Gerhard van Huyssteen for his linguistic advice; and the many respondents who donated their time towards this study.

References

[1] M. Adda-Decker and L. Lamel, "Multilingual dictionaries," in *Multilingual Speech Processing*, T. Schultz and K. Kirchhoff, Eds. Berlington, MA, USA: Academic Press, 2006, pp. 123–166.

[2] D. Jones, *An English Pronounciation Dictionary*, ser. Eleventh edition. Letchworth, Herts, Great Britain: The Temple Press, 1950.

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

[4] S. Schaden, "A database for analysis of cross-lingual pronunciation variants of European city names," in *Proc. LREC*, Portland, Oregon, United States, 2002, pp. 75–87.

[5] A. Font Llitjós and A. W. Black, "Knowledge of language origin improves pronunciation accuracy of proper names," in *Eurospeech*, 2001, pp. 1919–1922.

[6] H. van den Heuvel, B. Reveil, and J.-P. Martens, "Pronunciation-based ASR for names," 2008.

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