Porting an English Spoken Dialogue System to Xhosa and Zulu

Thembekile Malaza*, Luvuyo Martins*, Justus Roux* and Thomas Niesler‡

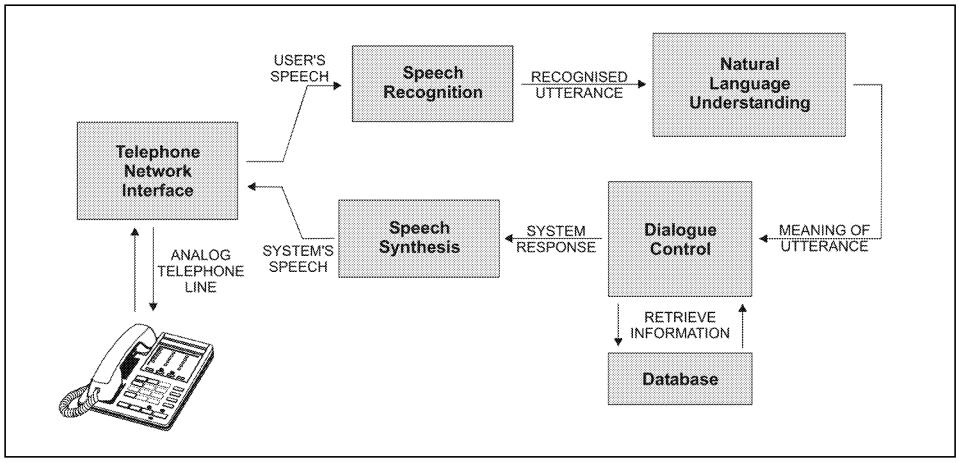
*Centre for Language and Speech Technology
‡Department of Electronic Engineering
University of Stellenbosch, Private Bag X1, Matieland 7602
malaza@sun.ac.za, lmartins@sun.ac.za, jcr@sun.ac.za, trn@dsp.sun.ac.za

The development of Xhosa and Zulu spoken dialogue systems, based on a working English prototype, is described. While direct translations of the system prompts and recognition grammars from English to Xhosa and Zulu, respectively, were appropriate in some cases, language- and culturally-specific factors often needed to be taken into account to yield an effective dialogue in the target language. This paper identifies the most important issues that were encountered in porting the dialogue system from English to Xhosa and Zulu. Issues such as code mixing, tone and politeness receive attention. The performances of the final Xhosa and Zulu prototypes are compared with the performance of the English prototype.

Introduction

A **Spoken Dialogue System** is a machine designed to maintain a dialogue with a human user using natural speech. The aim of the dialogue is often to provide the user with information that he or she requires, or to effect some transaction on the user's behalf. The following figure illustrates the components of such a dialogue system.

Figure 1: Architecture of a spoken dialogue system



System operation commences when the user contacts the system by telephone. The user's speech is transcribed into a sequence of words by the speech recognition component. These words are then interpreted by the natural language understanding component in order to determine the meaning of the user's utterance. Based on this extracted meaning, the dialogue controller decides on the most appropriate next action, which may include querying a database in order to retrieve information requested by the user. Finally, the system responds to the user by means of the speech synthesis component. This cycle continues until the transaction has been completed.

In South Africa, the development of spoken dialogue systems has so far occurred exclusively in English. The indigenous African languages have received little or no attention in this regard, despite the fact that they are more

widely spoken. This paper reports on experiences in porting an English spoken dialogue system to both Xhosa and Zulu. Particular attention is paid to the ways in which the final Xhosa and Zulu prototypes differ from naive literal translations of the English system. Finally, the performance of the three systems is compared.

The work presented in this paper is partly based on the recently completed African Speech Technology (AST) project (Roux et al., 2004). The AST initiative was aimed at promoting the technological development of South African languages, with one of its outcomes the development of English, Zulu and Xhosa prototype spoken dialogue systems. A hotel-reservation dialogue application was chosen as an easily understood and concrete example with which to carry out some experiments. The system allows a user to make a (fictitious) hotel reservation over the telephone and is designed to interact with the user in much the same way as a human receptionist would.

Dialogue system overview

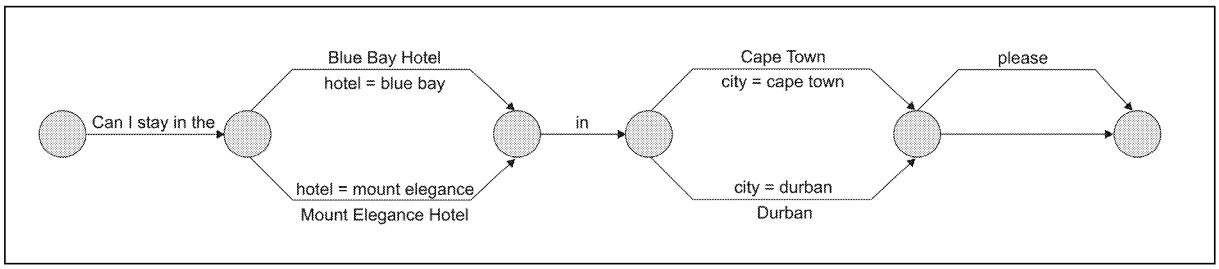
The technical details regarding the structure and operation of the dialogue system are not the focus of this paper, and have been described elsewhere (Tait et al., 2005). This section is intended only as a general overview.

The dialogue system is implemented as a state machine where each state has a well-defined function in the spoken interaction. For example, there is a state that determines the number of rooms in the reservation, and another to determine the date of arrival. In total, there are forty two states in this particular dialogue. In each state, the following sequence of steps is carried out:

1. The system communicates with the user by synthesising a **system prompt** associated with the current state.
2. The system listens for the user's reply by means of the speech recognizer. Speech recognition errors may occur, when the recognized utterance does not exactly correspond to the user's speech.
3. The system interprets the recognized utterance by means of the natural language understanding unit.
4. Based on the result of the natural language understanding, the dialogue controller determines which dialogue state should be entered next, and operation branches to step 1 above unless the dialogue is complete or the user has hung up.

The speech recognition component of the system is based on the open-source HTK hidden Markov model-based decoder (Young et al., 1999). Each dialogue state has a unique associated finite-state grammar that is used as a language model during speech recognition. The natural language understanding (NLU) component of the system is also finite-state based (Niesler & Roux, 2001) and shares the grammar used during the speech recognition search. The finite-state grammars used by the recognizer are augmented with semantic tags, which identify the meaning of the utterance. For example, the following heavily simplified finite-state grammar would allow exactly eight different utterances to be recognized. The values of the variables 'hotel' and 'city' correspond to the semantic tags and contain the result of the natural language understanding process.

Figure 2: An illustrative NLU grammar



The dialogue control decisions make use of these tags in order to determine the next dialogue state. Finally, the system employs limited-domain concatenative speech synthesis based on the Festival open source speech synthesis engine (Black & Lenzo, 2000).

The hotel-reservation system operates by successively prompting the user for information items such as the city involved, the name of the hotel, arrival and departure dates, the type of room, and finally the user's credit card details. It maintains a database of hotel-specific information, from which it determines room availability and pricing, and in which it stores a new booking.

Development of the English system

The English system was developed first. An initial prototype was iteratively refined by observing its performance and the reactions of users during **Wizard-of-Oz tests** (Dybkjaer et al., 1993) and subsequently during full system trials. Defects in the system's behaviour or capabilities were observed during the tests, after which appropriate improvements were made and further tests were performed. In this way the quality of the system was gradually improved. In particular, the system prompts were frequently re-crafted in order to reduce any ambiguity experienced by the user.

In order to ensure robustness to speech recognition errors without overly lengthening the dialogue, an **implicit confirmation** strategy was employed wherever possible. Instead of verifying each information item with the user before proceeding, the information is incorporated into the next prompt. The user can then interrupt the system by objecting if he/she has been misunderstood. Failure to object is interpreted as a confirmation by the system.

For example, assume that the previous dialogue state had determined that the booking has to be made in Johannesburg. Explicit confirmation that the city had been correctly recognized would be carried by a dedicated prompt such as:

(1) Did you say Johannesburg?

Instead, implicit confirmation embeds the city name in the next dialogue prompt, for example

(2) And where in Johannesburg would you like to stay?

The elimination of explicit confirmation states allows the total length of the dialogue to be strongly reduced, leading to greater user satisfaction.

We have also adopted a **system-directed** dialogue strategy in our prototypes. This means that the system actively maintains the initiative during the dialogue, while the role of the user is reactive. In this way the natural language understanding process is simplified, since it is not necessary for the system to guess the user's intention. Rather, the user is assumed always to react to the most recent prompt. For example, a relatively vague prompt such as:

(3) What type of room would you like?

may be expected to result in a large variety of replies. However a more specific reformulation such as:

(4) Would you like a single, double or twin room?

makes it extremely likely that the users reply will explicitly contain the words 'single', 'double' or 'twin'. This greatly simplifies the complexity and the reliability of both the speech recognition as well as the natural language understanding phases of the dialogue process.

Development of the Xhosa and Zulu systems

Both the Xhosa and the Zulu systems are based on the same dialogue state-machine structure used by the English system. However the associated system prompts and recognition grammars are language-specific and hence must differ. As a starting point, direct literal translations of the English prompts and grammars to, respectively, Xhosa and Zulu were used. In many cases these simplistic translations were unnatural, confusing or impractical, and extensive language-specific refinement was necessary in order to yield the final prototype systems. This section describes the nature of two of these refinements related to code mixing and politeness, respectively.

Code mixing

Code mixing is the process by which English words or phrases are embedded into Xhosa and Zulu sentences. It is accepted practice and occurs frequently in both languages. The understanding and appropriate incorporation of code mixing into system prompts and recognition grammars were integral to the successful development of Xhosa and Zulu prototype dialogue systems.

Mixed codes when dealing with numbers

Numeric input or confirmation was required or supplied by our systems at various points in the dialogue. For example, the user was asked to state the number of rooms in the booking, and to supply credit card details. The system was required to repeat these numbers as well as the total value of the booking for confirmation purposes.

When dealing with numerical quantities, it is particularly common for Xhosa and Zulu speakers to switch to English. This normally occurs because the English is more concise. For example, the English prompt:

(5) One room costs two hundred and fifty Rands.

could be formulated purely in Zulu and Xhosa, with its literal translation respectively as:

(6a) *Ikamelo libiza amakhulu amabili namashumi amahlanu amarandi.* (Zulu)

'One room costs hundreds that are two and tens that are five of Rands.'

(6b) *Igumbi elinye lixabisa amakhulu amabini anamashumi amahlanu eerandi.* (Xhosa)

'One room costs hundreds that are two that have tens that are five of Rands.'

However, an alternative using mixed codes is usually preferred. This includes the use of a vowel prefixed to the English word / phrase:

(7a) *Ikamelo libiza u-two hundred and fifty Rands.* (Zulu)

(7b) *Igumbi elinye lixabisa i-two hundred and fifty Rands.* (Xhosa)

It is clear that the alternatives using mixed codes are considerably more concise. The difference in length becomes more pronounced as the amount increases, i.e. thousands, tens of thousands etc.

The Xhosa and Zulu systems employed mixed codes in all system prompts dealing with numerical items. However, since the user might legitimately state amounts or numbers in either English (i.e. using a mixed code) or in Xhosa or Zulu, both alternatives were included in the recognition grammars. This allowed users to state numeric information both with and without using a mixed code. This however increases the complexity of the recognition grammars and hence the difficulty of the speech recognition process.

Once the final set of system trials had been completed, the results were used to determine the extent to which it was necessary to cater for code mixing in our dialogue.

Table 1 below reflects the actual prevalence of code mixing by, respectively, seventy-nine Xhosa speaking callers and ninety Zulu-speaking callers during the final system trials:

Table 1: Prevalence of code mixing when citing numeric items

	Xhosa system (n=79)	**Zulu system (n=90)**
Mixed-codes for Credit Card details	100%	100%
Mixed codes in number of rooms	0%	5%

Table 1 indicates that, in the final system trials credit card details were always supplied using mixed codes. The number of rooms in a booking, however, was normally given in Xhosa or Zulu respectively, although some mixed codes were recorded for the Zulu system.

The credit card details consist of a sixteen-digit card number (which is prompted for in four groups of four digits), a four-digit expiry date and a three-digit security code. The number of rooms, while unrestricted in principle, in practice always consisted only of one or two digits. The results indicate that such 'simple' numerical quantities are almost always specified in Xhosa or Zulu. It could be argued that this is the case because one- or two-digit numbers in English are not significantly shorter, and hence speakers prefer not to switch code. For example, the sentence:

(8) I will need two rooms

could be uttered purely as

(9a) *Ngizodinga amakamelo amabili.* (Zulu)

(9b) *Ndiza kufuna amagumbi amabini.* (Xhosa)

or, alternatively, though rarely, using a mixed code:

(10a) *Ngizodinga amakamelo ayi-two.* (Zulu)

(10b) *Ndiza kufuna amagumbi ayi-two.* (Xhosa)

A further contributing factor could be the grammatical and intonational nature of the original language, where the rhythmic pattern of the phrase /*amakamelo amabili*/ will probably feel more natural than with /*amakamelo ayi-two*/. However, it could also be generation related, as older speakers are less inclined to switch codes, but this was not investigated in this study.

Mixed codes in dates

When requesting or confirming a date, the day and the name of the month must be considered. For example, a Zulu caller specified his arrival date as:

(11) *Ngomhla weshumi kuNcwaba.* 'On the 10th of August.'

but it would also have been acceptable to use a mixed code in both Zulu and Xhosa, for example:

(12a) *Ngomhla ka-ten ku-August.* (Zulu) 'On the 10th of August.'

(12b) *Ngomhla we-ten ku-August.* (Xhosa) 'On the 10th of August.'

The systems made provision for both alternatives in the recognition and understanding grammars. When investigating the results of the system trials, we found that both were indeed used, as summarized in Table 2.

Table 2: Prevalence of code mixing when citing dates

	Xhosa system (n=79)	**Zulu system (n=90)**
Mixed-codes for numbers in dates	48%	62%
Mixed codes for month names	100%	95%

From this table it is clear that, when referring to numbers in dates, Xhosa and Zulu users tended to use mixed codes in 48% and 62% of the cases respectively. However, it appears that both Xhosa and Zulu month names have largely fallen into disuse as 100% of Xhosa and 95% of Zulu users used a mixed code.

Mixed codes in place names

A town or city is often known by more than one name, and very often these names may be used alternatively in a particular dialogue. For example, the city of Johannesburg may be casually referred to as 'Joburg' or even 'Jozi' by English speakers. However, the variety of alternative names given to a place is substantially larger in Xhosa and Zulu. For example, the city of Durban is most often referred to as 'eThekwini', but may also be referred to as:

(13) *eGagasini* 'at the place with waves'
eMdubane 'slang for at Durban'
kwelikaBhanana 'at the land of banana'

The same goes for Johannesburg, which is commonly referred to as 'eRhawutini' but could also be referred to as:

(14) *eGoli* 'at the place of gold'
kweleNduduma 'at the place of hills'
eMshishi 'at the place of business'

Although such alternatives were allowed in the recognition grammars, it was found that, as in the English system, the user replies were constrained to the city names used in the system prompt. For example, when the Zulu system prompted the user:

(15) *Kuqala ngazise igama ledolobha ozothanda ukuhlala kulo.*
Sinamahhotela kulamadolobha alandelayo: eGoli, eKapa naseThekwini.
'First, please give me the name of the town in which you would like to stay.
We have hotels in Johannesburg, Cape Town and Durban.'

the user would reply with 'eThekwini', using the form in the prompt and not any of the other possible alternatives listed above.

Use of prefixes in mixed codes

In order to embed English numbers and other words into Xhosa and Zulu sentences in a grammatically acceptable way, 'prefixes' must be prepended. The 'prefixes' consist of single vowel elements, such as *u-* or *i-*, or two or three syllable entities that could represent copulative verbs or relative copulative forms (Malcolm, 1966:196–230). The point is, however, that these languages, although related, display certain idiosyncrasies that have to be taken into account when developing the grammars to cover all potential responses of the user.

Consider the following responses:

(16) *Ngizodinga amakamelo angu-five / ayi-five.* (Zulu)
Ndiza kufuna amagumbi ayi-five. (Xhosa)
'I will need five rooms'; literally 'rooms that are five.'

The copulatives *angu- / ayi-* relate the number 'five' to the object 'amakamelo' (rooms) in Zulu in the mixed code, while Xhosa will only allow the *ayi-* form.

(17) *Ngizodinga ikamelo eliyi-one.* (Zulu)
Ndiza kufuna igumbi eliyi-one. (Xhosa)
'I will need one room.'

In (16) and (17) above grammatical rules related to plurality and singularity, respectively, govern the use of the copulative prefixes.

The omission of the prefix is not acceptable, in either Xhosa or Zulu. For example, the following renditions will be regarded as incorrect:

(18a) **Ngizodinga amakamelo five.* (Zulu)
(18b) **Ndiza kufuna amagumbi five.* (Xhosa)
*'I will need five rooms.'

When dealing with money amounts, it would appear that the use of a particular prefix in Xhosa is more constrained than in Zulu where two versions are possible (*u-* and *i-*). Consider the following examples

(19a) *Ikamelo libiza u-/ i-two hundred and fifty Rands.* (Zulu)

(19b) *Igumbi lixabisa i-two hundred and fifty Rands.* (Xhosa)

'The room costs two hundred and fifty Rands.'

Prefixes must also be employed when citing dates, however, a variety of options seems possible. For example, the English statement:

(20) I will be arriving on the second of July.

can be formulated in Zulu as

(21a) *Ngizofika mhla zizimbili kuNtulikazi.*

(21b) *Ngizofika mhla zizimbili kuJulayi.*

(21c) *Ngizofika ngomhla ka-two kuJulayi.*

(21d) *Ngizofika ngezi-five zikaJuly.*

(21e) *Ngizofika ngezi-five kuJuly.*

All these varieties need to be anticipated in a systematic manner in the grammar definitions.

When English proper names are incorporated into Zulu and Xhosa, prefixes must again be used, either to denote grammatical class or the location – these forms are fortunately very predictable. Consider the following examples, valid in both Xhosa and Zulu:

(22a) *iOcean View Hotel* – grammatical class 9: 'Ocean View Hotel'

(22b) *eOcean View Hotel* – location: 'at the Ocean View Hotel'

For the credit card types the English names are used either in a copulative sense (cf yi-Master Card) or as a statement of grammatical class (cf i-Visa):

(23) *Ingabe nhloboni yekhadi lokuthenga ngesikweletu onalo – yi-Master Card, i-Visa, i-Diners' Club?* (Zulu)

'What type of credit card do you have – is it a Master Card, Visa, Diners' Club?'

Apart from grammatical constraints, the use of some of these prefixes in mixed codes seems to be arbitrary depending on the user's language background and experience within the Nguni group of languages. This phenomenon poses a particular challenge for developers of automated systems since this (as yet undescribed) variation must nevertheless be accounted for.

Mixed codes to circumvent lexical, semantic or grammatical language differences

Code switching may not be merely a matter of convenience. It may happen, for example, that appropriate translations of English words do not exist. Cumbersome and sometimes misleading or confusing descriptions would have to be used instead. For example, the English dialogue system's opening prompt is:

(24) Central Hotel Reservations; good morning / good afternoon / good evening!

The time reference here was linked to the time set on the system so as to present the appropriate prompt at the appropriate time of the day or night. Since this distinction does not exist in Xhosa, no translation took place, and instead the following prompt was created:

(25) *Wamkelekile, kuse-Central Hotel Reservations apha.*

'Welcome, this is Central Hotel Reservations.'

Apart from evading the time reference, the words 'central' and 'reservations' were not translated to Xhosa, even though suitable Xhosa words do exist, since these would appear unnatural in the context of the English word 'hotel'.

Even if literal translations of English words were possible, the correct meaning could not be preserved in some cases. For example, the English prompt:

(26a) The Summit City, a three star hotel, or the Gold Reef City, a five star hotel.

could be translated literally into Xhosa as

(26b) *I-Summit City, ezinkwenkwezi ezintathu, okanye I-Gold Reef ezinkwenkwezi ezintlanu.*

However, in Xhosa there is no clear link between 'iinkwenkwezi' (stars) and the system of hotel grading; therefore it seemed necessary to resort to code mixing within the prompt as:

(26c) *KwiSummit City eyi-three star okanye kwi-Gold Reef City eyi-five star.*

This sentence is both culturally acceptable and semantically accurate.

Finally, even if direct translations were possible, lexical differences could render unacceptable results. For example, when confirming the duration of stay, the English system states the number of nights involved; however, in both Xhosa and Zulu the noun 'ubusuku' (night) belongs to class 14, which does not have a plural form. The word for day ('usuku'), on the other hand, does have a plural form. Therefore the system confirms the duration of stay in days instead of nights.

Tone and politeness

Feedback from users testing the dialogue systems appears to indicate that in some cases it is inappropriate to be too direct when requesting information in Xhosa or in Zulu. For example, the English system requests the city name by prompting:

(27) First, tell me in which city you would like to stay?

This formulation has been found to be effective because it elicits short and therefore easily processed replies. It has not been found to be curt or rude by any of the English users; it could be regarded as a polite question. However the direct translation into Xhosa:

(28a) *Kuqala ndixelele ukuba kweyiphi idolophu ongathanda ukuhlala kuyo?*

which means literally,

(28b) First tell me (that) it is in which town that you want stay (in)?

was described as impolite by some Xhosa users since it sets a demanding tone. Xhosa culture dictates a more polite approach when dealing with persons with whom one is not familiar. A more appropriate alternative was found to be:

(29a) *Nceda, kuqala ndixelele igama ledolophu ongathanda ukuhlala kuyo?*

which means literally

(29b) Please, first would you tell me the name of the city you would like to stay in?

This more polite form was found to be more acceptable to Xhosa users.

System performance

The performance of the English, Xhosa and Zulu dialogue systems was evaluated by requesting a set of volunteers to call the system and make a reservation. The volunteers were recruited by the research team and consisted of University staff and students, each of whom was presented with an information sheet describing the operation of the system before being asked to place a call. The calls were subsequently analysed and all user utterances annotated manually in order to determine the accuracy of the speech recognition as well as the performance of the dialogue system as a whole.

Table 3 summarizes the overall performance of each system. Although the average number of utterances made per call is approximately the same for each system, the average duration of the calls made to the Xhosa and Zulu systems is much longer than for the English systems. Xhosa and Zulu speakers showed a tendency to use longer sentences in their replies than their English counterparts.

Table 3: Overall system performance during usability testing

System	English	Xhosa	Zulu
Total calls	77	76	90
Average call duration (sec)	267	391	411
Recorded utterances	2326	2246	2405
Average utterances per call	30.2	29.6	26.7
Successful reservations	60 (78%)	53 (70%)	61 (68%)
Unsuccessful calls	17 (22%)	23 (30%)	29 (32%)

The last row of Table 3 shows the number of users who experienced difficulty in using the system, lost patience and prematurely terminated their calls. Analysis of the call logs indicate that in approximately one fifth of these cases problems were caused by loud and frequent background noises which led to severe speech recognition errors. In the remaining cases failure was mainly due to persistently erroneous recognition of the speaker in question.

Table 4 summarizes the accuracy of the speech recognition and understanding processes for a selection of the dialogue states. It indicates the percentage of all the utterances that were processed in each of these states for which the correct information was determined from the user's speech.

Table 4: Per-state system performance during usability testing

State name	English	Xhosa	Zulu
City name	85%	92%	96%
Hotel name	83%	62%	47%
Arrival date	57%	43%	52%
Departure date	71%	40%	54%
Number of single rooms	86%	27%	74%
Number of double rooms	97%	23%	62%
Credit card number	64%	61%	36%
Credit card type	76%	86%	89%
Credit card expiry date	48%	43%	42%
Credit card security code	61%	69%	17%
Overall performance	76.4%	64.2%	66.3%

Table 4 indicates that the overall performance of the Xhosa and the Zulu systems is approximately equal, and that the performance of the English system is significantly better. Furthermore, recognition of numeric amounts is generally much less reliable in Xhosa and Zulu than in English. This is especially evident when considering the number of rooms, the credit card number, the credit card security code, and also the arrival and departure dates. The city name and the credit card type, on the other hand, are recognized with comparable accuracy by all three systems. The relative degradation in performance for the Xhosa and Zulu systems with respect to the English

system is ascribed to the significantly more complex recognition grammars that were used in these two languages in order to allow for code mixing.

Summary and conclusion

The development of Xhosa and Zulu dialogue systems emanating from an English prototype was described in some detail. Initial Xhosa and Zulu systems were obtained through literal and direct translations of the English system prompts and recognition grammars. Subsequent extensive revision and refinement of these two systems were necessary in order to obtain systems that seemed more natural and could be used by mother-tongue speakers.

The integration of code mixing into both the system prompts as well as the recognition grammars was found to be pivotal to the development of systems that could be understood and used by Xhosa and Zulu users. In the prototypes, numeric quantities and dates had to be produced by using mixed codes. Users were, however, permitted to reply either using mixed codes, or pure Xhosa or Zulu. This added complexity to the speech recognition process and led to deterioration in the system's performance due to a greater prevalence of recognition errors. However, it was found that, longer digit strings, like credit card numbers, were exclusively cited in mixed code, and only shorter numbers like those used in dates, or those used to specify the number of rooms in a booking, were cited in Xhosa and Zulu. Hence the recognition grammars could be simplified for longer digit strings without sacrificing system flexibility, and probably improving overall system accuracy.

While code mixing was prevalent, and English words were often integrated into Xhosa and Zulu sentences by speakers, this process involves the use of a variety of prefixes in order to embed the English word in a grammatically acceptable way. These prefixes must also be accounted for in the recognition grammars. Since the variety of combinations that may occur is large, this increases the complexity of the recognition grammars, with further negative impact on system accuracy.

Finally, it was found that while short system prompts in a particular language may result in improved performance in that language (e.g. in English), mother-tongue speakers of Xhosa and Zulu interpreted these types of prompts as too aggressive and impolite, and insisted on more polite, albeit longer prompts. Hence the phrasing of system prompts must be crafted with this in mind.

References

Black, A.W. & Lenzo, K.A. 2000. Limited domain synthesis. *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000).* Beijing, China: 411–414.

Dybkjaer, H., Bernsen, N.O. & Dybkjaer, L. 1993. Wizard-of-Oz and the trade-off between naturalness and recogniser constraints. *Proceedings of Eurospeech 93*. Berlin, Germany:947–950.

Malcolm, D.McK. 1966. *A new Zulu manual.* Cape Town: Longmans.

Niesler, T.R. & Roux, J.C. 2001. Natural language understanding in the DACST-AST dialogue system. *Proceedings of the twelfth annual symposium of the Pattern Recognition Association of South Africa (PRASA)*. Franschhoek:134–136.

Roux, J.C., Louw, P.H. & Niesler, T.R. 2004. The African Speech Technology project: An assessment. *Proceedings of the fourth international conference on language resources and evaluation (LREC)*. Lisbon:93–96.

Tait, M., Visagie, A. & Niesler, T.R. 2005. Development of a spoken dialogue system operating in Afrikaans, South African English and Xhosa. *Transactions of the SAIEE* 96(3):224–233.

Young, S.J., Kershaw, D., Odell, J., Ollason, D., Valtchev, V. & Woodland, P.C. 1999. *The HTK book* 2.2. Entropic Cambridge Research Laboratory. Cambridge, England.