A First South African Corpus of Multilingual Code-switched Soap Opera Speech

Ewald van der Westhuizen, Thomas Niesler
Department of Electrical and Electronic Engineering
Stellenbosch University, South Africa
ewaldvdw@sun.ac.za, trn@sun.ac.za

Abstract

We introduce a speech corpus containing multilingual code-switching compiled from South African soap operas. The corpus contains English, isiZulu, isiXhosa, Setswana and Sesotho speech, paired into four language-balanced subcorpora containing English-isiZulu, English-isiXhosa, English-Setswana and English-Sesotho. In total, the corpus contains 14.3 hours of annotated and segmented speech. The soap opera speech is typically fast, spontaneous and may express emotion, with a speech rate that is between 1.22 and 1.83 times higher than prompted speech in the same languages. Among the 10343 code-switched utterances in the corpus, 19207 intrasentential language switches are observed. Insertional code-switching with English words is observed to be most frequent. Intraword code-switching, where English words are supplemented with Bantu affixes in an effort to conform to Bantu phonology, is also observed. Most bigrams containing code-switching occur only once, making up between 64% and 92% of such bigrams in each subcorpus.

Keywords: code-switching, spontaneous speech, South African languages, isiZulu, isiXhosa, Setswana, Sesotho

1. Introduction

South Africa has 11 official languages and since the majority of South Africans are multilingual, code-switching occurs commonly and spontaneously. Code-switching is the phenomenon of using more than one language within the same conversation or utterance (Van Dulm, 2007). Language switches may even occur mid-word, as examples in our data will demonstrate.

In an effort to be representative and accommodate a large viewer base, conversations in South African soap operas are multilingual and exhibit code-switching. Soap operas are therefore an interesting source of multilingual code-switched speech data. The corpus we describe in this paper displays various types of code-switching. Intersentential code-switching occurs when the language alternates between utterances of a conversation. For example, a speaker may utter a sentence in English while the following sentence is spoken in isiZulu. Intrasentential code-switching occurs when the language alternates within a single utterance. For example, an utterance may start in English and mid-sentence switch to isiZulu. Intrasentential code-switching can further be divided into alternational and insertional switching. Alternational code-switching occurs when an utterance starts with a phrase in one language and switches to a phrase in another language and where the phrases conform to the grammar of the respective languages. Insertional code-switching occurs when a language element from a secondary language is embedded into the structure of a matrix language. Code-switching can also occur at morpheme-level within word boundaries (intraword). Examples of intraword switching from our corpus include the joining of Bantu language class affixes to English stems.

Currently our corpus contains 14.3 hours of language-balanced speech compiled from soap opera broadcasts. Four Bantu languages are paired with English to yield subcorpora containing English-isiZulu, English-isiXhosa, English-Setswana and English-Sesotho code-switched speech. IsiZulu and isiXhosa are part of the Nguni language family, while Setswana and Sesotho are part of the Sotho-Tswana language family. Hence these five languages represent an interesting basis for studies into acoustic and language modelling for code-switched speech. The corpus is planned to be made available for research use.

2. Background

A number of code-switched corpora have already been described in the literature. Below we summarise a few notable examples. The list is however not intended to be exhaustive.

1. The SEAME corpus from Nanyang Technological University, Singapore, and Universiti Sains Malaysia, contains 63 hours of spontaneous Mandarin-English code-switched conversational and interview speech uttered by Malaysian and Singaporean speakers (Vu et al., 2012; Dong et al., 2010; Adel et al., 2015).
2. The HKUST Mandarin-English Corpus from the Hong Kong University of Science and Technology (Li et al., 2012; Li and Fung, 2013) consists of code-switched spontaneous speech from meetings and interviews and comprises 5 hours of transcribed and 15 hours of untranscribed speech.
3. The CECOS Chinese-English Corpus was compiled at the National Cheng Kung University in Taiwan (Shen et al., 2011). It contains 12 hours of speech collected from 77 speakers uttering prompted code-switch sentences.
4. The CUMIX Cantonese-English speech corpus (Chan et al., 2005) was compiled at The Chinese University of Hong Kong. It contains 17 hours of code-switched speech read by 80 speakers.
5. A small English-Spanish Corpus was compiled at the University of Texas. The corpus contains 40 minutes of transcribed spontaneous conversations with a vocabulary of 1516 words (Franco and Solorio, 2007).

Figure 1: A screenshot of ELAN media annotation tool. Only four annotation tiers are shown due to limited space.

6. A Frisian-Dutch corpus of radio broadcast speech was compiled at Radboud University, Nijmegen, containing 18.5 hours of speech with code-switching (Yılmaz et al., 2016).
7. A corpus of Sepedi-English code-switched speech was compiled by the South African CSIR (Modipa et al., 2013). It contains 10 hours of prompted speech, sourced from radio broadcasts and read by 20 Sepedi speakers.

The list above demonstrates that data collection for code-switched speech has focussed mostly on English with Asian languages. Recently this has been extended by a European language pair (Frisian-Dutch). Only one corpus containing an African language could be found. Our corpus extends this field by providing spontaneous data in several African languages, some of which are related.

3. Data Collection

The corpus is compiled from digital video recordings of 626 South African soap opera episodes. Mono audio sampled at 32kHz and coded as 16-bit PCM was extracted from the original source videos for each episode. The ELAN media annotation tool (Wittenburg et al., 2006) was used to segment the audio into segments corresponding to sentences. Figure 1 shows how ELAN displays the audio waveform and annotation tiers of a transcribed code-switched utterance. Sentences containing code-switching were further subdivided into segments corresponding to a single language. Such language segments were transcribed and tagged with a language label. Hence the following annotation tiers were used:

- monolingual sentence text,
- monolingual sentence language,
- code-switch segment text,
- code-switch segment language, and
- speaker name or ID.

Comparisons between the soap opera scripts and the uttered speech showed that the actors display a strong tendency to ad-lib. Therefore the speech can be regarded as spontaneous. Actors who code-switch usually do so between English and their Bantu mother tongue. Nevertheless, many utterances do contain more than two languages. These have been excluded from the current corpus, which focusses on code-switching in language pairs.

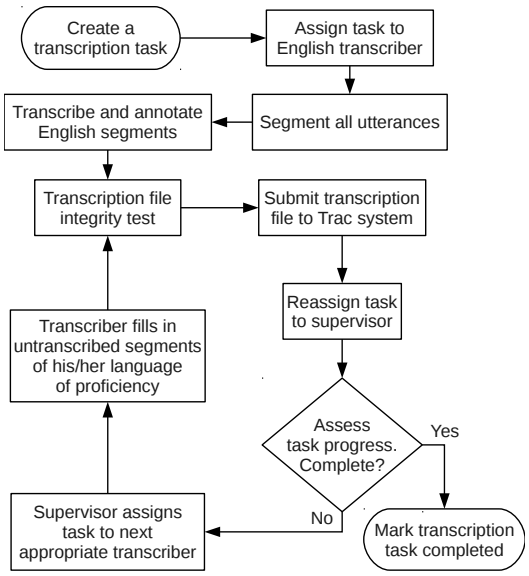

Figure 2: Flowchart depicting the transcription process for a single soap opera episode.

All speech was transcribed by fluent bilingual speakers. The Trac software project management, wiki and bug tracking system (Edgewall Software, 2012) was used to manage the tasks assigned to the team of transcribers. The flowchart in Figure 2 shows the steps involved in transcribing the speech of a single episode. A transcription task is created and first assigned to the principal transcriber. The principal transcriber performs the initial utterance segmentation and English transcription. Before a transcription file is submitted to the repository, a transcriber has to perform a transcription error and spell check. The transcription error check reports on common errors that occurred during transcription. For example, if transcription text was provided, but the segment language or speaker name was omitted, an error is flagged. After submission of the transcription file, the transcription task is reassigned to the supervisor to assess overall progress for the episode. If incomplete utterance segments still exist, the task is assigned to another transcriber with proficiency in the relevant languages. This continues until the supervisor confirms the episode to be fully transcribed and marks the task as complete. During this process, an episode is passed to at least one transcriber of each of the five languages, each annotating speech in their language of proficiency.

No overlapping speech was transcribed. Contracted forms of words were transcribed as closely to the speaker's actual pronunciation as possible. Contracted pronunciations are widespread in the Bantu languages where postlexical deletion regularly occurs in fast spoken spontaneous speech (Van der Westhuizen and Niesler, 2016). Using the example shown in Figure 1, the word *kukhona* is transcribed as *k'khon'*, with the apostrophes indicating the deletion of the *u* and *a* in the pronunciation.

All transcriptions are made in lower case, except for proper nouns and the English word 'I'. Since the language label for a proper noun can be ambiguous, the following rule was applied. Proper nouns are assigned the language label of

English-isiZulu												
set	engtok	zultok	tottok	engtyp	zultyp	tottyp	utt.cnt	emdur	zmdur	ecdur	zsdur	dur
train	28033	24350	52383	3608	6765	10373	8381	1.55h	1.55h	45.86m	56.99m	4.81h
dev	838	734	1572	415	443	858	225	0.0	0.0	4.01m	3.96m	8m
test	2459	3199	5658	871	1420	2291	768	0.0	0.0	12.76m	17.85m	30.4m
Total	31330	28283	59613	3842	7425	11269	9374	1.55h	1.55h	1.04h	1.31h	5.45h
English-isiXhosa												
set	engtok	xhotok	tottok	engtyp	xhotyp	tottyp	utt.cnt	emdur	xmdur	ecdur	xcdur	dur
train	20332	12223	32555	2630	5108	7738	6996	65.22m	53.55m	18.04m	23.73m	160.54m
dev	1155	1149	2304	484	764	1248	554	2.86m	6.48m	2.21m	2.13m	13.68m
test	1151	1501	2652	500	899	1399	391	0.0	0.0	5.56m	8.78m	14.34m
Total	22638	14873	37511	2829	5997	8826	7941	68.08m	60.03m	25.81m	34.64m	3.143h
English-Setswana												
set	engtok	tsntok	tottok	engtyp	tsntyp	tottyp	utt.cnt	emdur	tmdur	ecdur	tcdur	dur
train	16155	19570	35725	2360	1450	3810	5290	40.4m	30.96m	34.37m	34.01m	139.74m
dev	1168	2539	3707	513	539	1052	517	0.76m	4.26m	4.54m	4.27m	13.83m
test	1960	2979	4939	728	526	1254	496	0.0	0.0	8.87m	8.96m	17.83m
Total	19283	25088	44371	2606	1627	4233	6303	41.16m	35.22m	47.78m	47.24m	2.86h
English-Sesotho												
set	engtok	sottok	tottok	engtyp	sottyp	tottyp	utt.cnt	emdur	smdur	ecdur	scdur	dur
train	15372	19825	35197	2253	2086	4339	5659	49.34m	35.32m	23.02m	34.04m	141.72m
dev	840	2227	3067	436	614	1050	475	1.09m	5.05m	3.03m	3.59m	12.77m
test	1789	2265	4054	658	535	1193	429	0.0	0.0	7.80m	7.74m	15.54m
Total	18001	24317	42318	2518	2437	4955	6563	50.43m	40.37m	33.85m	45.37m	2.83h

Table 1: Subcorpora statistics. (train; Training set; dev: Development set; test: Evaluation test set; eng: English; zul: isiZulu; xho: isiXhosa; tsn: Setswana; sot: Sesotho; tot: Total; tok: tokens; typ: types; utt.cnt: Utterance count; emdur: Duration of English monolingual sentences; zmdur: Duration of isiZulu monolingual sentences; xmdur: Duration of isiXhosa monolingual sentences; tmdur: Duration of Setswana monolingual sentences; smdur: Duration of Sesotho monolingual sentences; ecdur: Duration of English code-switched segments; zcdur: Duration of isiZulu code-switched segments; xcdur: Duration of isiXhosa code-switched segments; tcdur: Duration of Setswana code-switched segments; scdur: Duration of Sesotho code-switched segments; dur: Total speech duration.)

English-isiZulu utt.cnt			English-isiXhosa utt.cnt			English-Setswana utt.cnt			English-Sesotho utt.cnt		
em	zm	ezcs	em	xm	excs	em	tm	etcs	em	sm	escs
1226	4364	3784	3372	2877	1692	1790	1827	2686	2493	1889	2181

Table 2: Monolingual and code-switched utterance counts for each full subcorpus. (utt.cnt: utterance count; em: English monolingual; zm: isiZulu monolingual; ezcs: English-isiZulu code-switched; xm: isiXhosa monolingual; excs: English-isiXhosa code-switched; tm: Setswana monolingual; etcs: English-Setswana code-switched; sm: Sesotho monolingual; escs: English-Sesotho code-switched)

the preceding word, e.g.:
`w1_en w2_en` **`name_en`** `w1_zu w2_zu`,
where `_en` and `_zu` are the English and isiZulu language labels, respectively.

Many short English function words, such as *but*, *why* and *if* occur very frequently in Bantu speech. It is tedious and time-consuming to mark all instances of such words as English. It could even be argued that such common English words have been adopted as loanwords by the Bantu languages. As a time-saving strategy during annotation, these words were either:

- labelled with the language they are embedded in, e.g. **`but_zu`** `w1_zu w2_zu w1_en w2_en`, or
- labelled as English when adjacent to an English word, `w1_zu w2_zu` **`but_en`** `w1_en w2_en`.

Since the set of English function words in question is closed, these words can easily and unambiguously be identified in the transcriptions.

4. Data Analysis

Tables 1 and 2 show statistics for the current version of our corpus. For each subcorpus, the tables show the speech durations, word token, type and utterance counts as well as the divisions into training, development and test sets. Since the transcription process is on-going, these statistics are subject to change in future and will be updated accordingly.
Table 3 shows the number of times the language switches in the corpus. Switching to English occurs slightly more often than switching from English to another language. We can consequently reason that, for our data, there is a slightly higher probability of an imminent switch when the current language of use is not English.

Subcorpus	Switch.dir	Count
engzul	eng→zul	3099
	zul→eng	3717
engxho	eng→xho	1195
	xho→eng	1479
engtsn	eng→tsn	2600
	tsn→eng	2728
engsot	eng→sot	2109
	sot→eng	2280
Total		19207

Table 3: Number of intrasentential language switches for full subcorpora. (Switch.dir: direction of language switch)

seg.len	EZ		EX		ET		ES	
	eng	zul	eng	xho	eng	tsn	eng	sot
1	2776	1859	1030	774	2019	766	1707	680
2	782	1502	294	655	644	804	500	674
3	388	988	163	419	379	631	285	553
4	280	534	118	226	265	555	196	447
5	214	292	101	113	177	378	131	300
6	168	186	96	69	132	308	83	255
7	138	90	69	37	102	213	55	151
8	84	48	46	19	70	126	49	110
9	51	27	36	9	53	101	25	88
> 10	144	49	79	13	118	173	51	229

Table 4: Segment lengths (seg.len) in number of words for monolingual segments of code-switched sentences against frequency count per language across each of the full subcorpora. (EZ: English-isiZulu; EX: English-isiXhosa; ET: English-Setswana; ES: English-Sesotho)

Next, we consider the length of monolingual segments in code-switched sentences. Segment lengths are measured in terms of number of word tokens. Table 4 tabulates the frequency of segments of a particular length. For example, there are 2776 instances of single-word English segments among the 3784 code-switched sentences in the English-isiZulu subcorpus. Single-word English segments are the most frequent within each subcorpus and typically represent insertional code-switching. English segments consisting of two or more words are much less frequent. The insertion of an English word into a Bantu matrix language is therefore the most common form of code-switching in our corpus. For the Bantu languages, there is a more gradual decline in frequency as the segment length increases.

Table 5 shows the training set occurrence counts of bigrams containing code-switching for the four subcorpora. The majority of code-switched bigrams occur only once. This is least pronounced for eng→tsn, where 1210 of 1878 (64%) such bigrams occur only once, while it is most pronounced for eng→xho, where 719 of the 784 (92%) occur only once. The most frequent eng→zul code-switched bigram occurs 31 times. These figures emphasise the sparsity of the data with respect to code-switch events. This will pose a challenge in terms of their effective modelling. In particular, since the majority of code-switch bigrams occur only once in the training data, it will in general not be possible to model code-switch events occurring in new and unseen data directly from training data examples (Van der Westhuizen and Niesler, 2017).

From English to isiZulu			
Trigger	Count	Target	Count
and	128	-a	179
no	94	ukuthi	158
so	92	-e	70
but	73	mina	52
sure	70	yakho	44
if	70	i-	43
for	43	-i	39
then	42	yami	38
why	41	wena	37
okay	39	nje	37
From isiZulu to English			
Trigger	Count	Target	Count
i-	465	and	98
u-	205	right	88
ama-	145	for	66
ukuthi	119	I	61
e-	116	sure	60
ngi-	83	you	57
le	64	but	39
ne-	55	so	33
uku-	39	understand	31
kwi-	39	or	30

Table 6: Token counts for English and isiZulu trigger and target types.

A code-switched bigram consists of two tokens. We define the first token as the trigger and the second as the target. Tables 6 to 9 show the most frequent word types serving either as a trigger or target in the subcorpora. Triggers in the Bantu languages include prefixes which join to English stems. Conversely, Bantu targets include suffixes similarly joining to English stems. Such affixes serve to supplement the pronunciation of an English word to conform to Bantu phonology and was also observed by (Modipa et al., 2013). An atypical trigger is the Sesotho suffix '-a' (Table 9) which precedes a switch to English. Examples of such cases are shown in the following two sentence fragments:

`u_so interview_en` **`-a_so`** `a_en while_en ago_en`

and

`o_so na_so push_en` **`-a_so`** `for_en this_en.`

In both cases the Sesotho suffix `-a` follows an English stem and precedes a switch to an English word, sandwiching the suffix between two English units.

Finally, Table 10 shows the speech rates calculated from aligned phone-level transcriptions for our corpus as well as for prompted speech in the same languages (Barnard et al., 2014). We see that the spontaneous soap opera speech is notably faster than the prompted speech in all cases.

5. Conclusion

We have compiled a 14.3 hour corpus of spontaneous multilingual code-switched speech from South African soap opera broadcasts. The corpus contains four language-balanced subcorpora: English-isiZulu, English-isiXhosa, English-Setswana and English-Sesotho. Many interesting

Bigram occurrence count	1	2	3	4	5	6	7	8	9	10	11	13	14	15	26	31	Total # bgtok
eng→zul #bgs	1838	87	26	8	7	3	3	0	1	0	0	0	0	0	0	1	2236
zul→eng #bgs	2064	163	38	14	9	6	1	0	2	1	0	2	0	1	1	0	2743
eng→xho #bgs	719	18	7	2	0	0	0	0	0	0	0	0	0	0	0	0	784
xho→eng #bgs	916	35	4	2	1	0	0	0	0	0	0	0	0	0	0	0	1011
eng→tsn #bgs	1210	137	49	16	10	5	5	2	2	2	0	0	1	0	0	0	1878
tsn→eng #bgs	1527	132	22	5	5	3	3	0	0	1	0	0	0	0	0	0	1951
eng→sot #bgs	1083	98	23	17	12	1	4	0	1	2	0	1	0	0	0	0	1552
sot→eng #bgs	1333	108	27	7	1	5	1	1	0	0	1	0	0	0	0	0	1719

Table 5: Occurrence counts of code-switch bigrams for the subcorpora training sets. (#bgs: number of bigrams; bgtok: bigram tokens.)

From English to isiXhosa			
Trigger	Count	Target	Count
and	56	ukuba	34
so	32	yakho	27
mean	25	yam	22
look	24	wena	21
but	23	yakhe	16
right	22	kodwa	15
for	16	-a	15
no	15	i-	13
well	14	-e	12
then	14	ngoku	11
From isiXhosa to English			
Trigger	Count	Target	Count
i-	157	and	46
u-	44	right	41
ne-	30	I	31
nge-	22	you	30
la-	22	is	24
kwi-	21	so	21
le	19	family	21
e-	19	for	19
yi-	18	but	17
ukuba	16	please	14

Table 7: Token counts for English and isiXhosa trigger and target types.

From English to Setswana			
Trigger	Count	Target	Count
so	70	-a	234
if	62	ke	193
why	60	o	182
well	51	ya	178
and	51	-e	169
then	49	go	157
like	47	e	106
know	39	a	87
mean	38	le	61
for	38	nna	55
From Setswana to English			
Trigger	Count	Target	Count
di	209	you	61
le	178	and	52
ke	174	I	47
re	123	for	44
o	115	like	38
ka	79	right	32
e	75	a	29
mo	67	sure	28
ko	65	the	27
go	65	but	25

Table 8: Token counts for English and Setswana trigger and target types.

examples of code-switching are observed, with English word insertions the most frequent. However, the distribution of code-switch examples is sparse, which could make modelling of unseen code-switch events challenging. Intraword code-switching, where Bantu affixes are joined to English stems in an effort to change English words to conform with Bantu phonology, is also often observed. The corpus is useful in the study of code-switched automatic speech recognition, discourse and dialogue analysis. The corpus is planned to be made available for research use.

6. Acknowledgements

We would like to thank e.tv and Yula Quinn at Rhythm City for assistance with data compilation. We pay tribute to Dumi Masilela (29) who tragically passed away during the course of this work. He portrayed the character Sifiso Ngema.

7. Bibliographical References

Adel, H., Vu, N. T., Kirchhoff, K., Telaar, D., and Schultz, T. (2015). Syntactic and semantic features for code-switching factored language models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(3):431–440.

Barnard, E., Davel, M. H., Van Heerden, C., De Wet, F., and Badenhorst, J. (2014). The NCHLT speech corpus of the South African languages. In *Proceedings of the 4th International Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2014)*, St Petersburg, Russia, 14-16 May 2014.

Chan, J. Y. C., Ching, P. C., and Lee, T. (2005). Development of a Cantonese-English Code-Mixing Speech Corpus. In *Proceedings of Interspeech 2005*, Lisbon, Portugal, 4-8 September 2005.

Dong, M., Chan, P., Cen, L., Li, H., Teo, J., and Kua, P. J. (2010). SEAME: A Mandarin-English Code-Switching

From English to Sesotho			
Trigger	Count	Target	Count
if	50	-a	226
so	40	-e	127
maybe	36	ke	117
know	36	u	98
why	34	ya	88
well	29	ha	70
then	29	e	69
for	28	hore	67
mean	26	o	63
because	24	ea	59
From Sesotho to English			
Trigger	Count	Target	Count
di	175	and	47
ke	128	right	34
le	123	something	33
ka	78	you	30
u	75	I	27
mo	71	the	25
-a	62	like	25
ho	60	maybe	22
o	57	for	22
e	56	so	19

Table 9: Token counts for English and Sesotho trigger and target types.

Corpus	Language	Rate (ph/s)	
		Soap opera spontaneous	NCHLT prompted
EZ	English	13.09	10.62
	isiZulu	16.50	9.04
EX	English	15.38	–
	isXhosa	19.98	8.77
ET	English	13.37	–
	Setswana	16.29	10.39
ES	English	13.24	–
	Sesotho	16.03	9.86

Table 10: Speech rate in phones per second (ph/s) of spontaneous soap opera speech and prompted speech for English and isiZulu. (EZ: English-isiZulu; EX: English-isiXhosa; ET: English-Setswana; ES: English-Sesotho)

Speech Corpus in South-East Asia. In *Proceedings of Interspeech 2010*, Makuhari, Japan, 26-30 September 2010.

Edgewall Software. (2012). Trac: Integrated SCM & Project Management. `https://trac.edgewall.org/`, Version 0.12.3. Release date: 2012-02-06.

Franco, J. C. and Solorio, T. (2007). Baby-steps towards building a Spanglish language model. In Gelbukh A., editor, *Computational Linguistics and Intelligent Text Processing. CICLing 2007. Lecture Notes in Computer Science, vol 4394*, Berlin, Heidelberg. Springer.

Li, Y. and Fung, P. (2013). Improved mixed language speech recognition using asymmetric acoustic model and language model with code-switch inversion constraints. In *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, 26-31 May 2013.

Li, Y., Yu, Y., and Fung, P. (2012). A Mandarin-English Code-Switching Corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey, 23-25 May 2012.

Modipa, T. I., Davel, M. H., and de Wet, F. (2013). Implications of Sepedi/English code switching for ASR systems. In *Proceedings of the Twenty-Fourth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2013)*, Johannesburg, South Africa, 3 December 2013.

Shen, H.-P., Wu, C.-H., Yang, Y.-T., and Hsu, C.-S. (2011). CECOS: A Chinese-English code-switching speech database. In *Proceedings of the 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA 2011)*, Hsinchu City, Taiwan, 26-28 October 2011.

Van der Westhuizen, E. and Niesler, T. R. (2016). The effect of postlexical deletion on automatic speech recognition in fast spontaneously spoken Zulu. In *Proceedings of Interspeech 2016*, San Francisco, CA, 8-12 September 2016.

Van der Westhuizen, E. and Niesler, T. R. (2017). Synthesising isiZulu-English code-switch bigrams using word embeddings. In *Proceedings of Interspeech 2017*, Stockholm, Sweden, 20-24 August 2017.

Van Dulm, O. (2007). *The grammar of English-Afrikaans code switching: A feature checking account.* LOT, Utrecht, The Netherlands.

Vu, N. T., Lyu, D.-C., Weiner, J., Telaar, D., Schlippe, T., Blaicher, F., Chng, E.-S., Schultz, T., and Li, H. (2012). A first speech recognition system for Mandarin-English code-switch conversational speech. In *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*, Kyoto, Japan, 25-30 March 2012.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 24-26 May 2006.

Yılmaz, E., Andringa, M., Kingma, S., Dijkstra, J., Van der Kuip, F., Van de Velde, H., Kampstra, F., Algra, J., Van den Heuvel, H., and Van Leeuwen, D. (2016). A longitudinal bilingual Frisian-Dutch radio broadcast database designed for code-switching research. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, 23-28 May 2016.