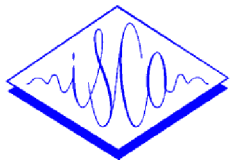
ISCA Archive
http://www.isca-speech.org/archive

SLTU-2014, St. Petersburg, Russia, 14-16 May 2014



FEATURES FOR FACTORED LANGUAGE MODELS FOR CODE-SWITCHING SPEECH

Heike Adel[1,2], *Katrin Kirchhoff*[2], *Dominic Telaar*[1], *Ngoc Thang Vu*[1], *Tim Schlippe*[1], *Tanja Schultz*[1]

[1]Cognitive Systems Lab, Karlsruhe Institute of Technology (KIT)
[2]Department of Electrical Engineering, University of Washington (UW)
heike.adel@student.kit.edu

ABSTRACT

This paper presents investigations of features which can be used to predict Code-Switching speech. For this task, factored language models are applied and implemented into a state-of-the-art decoder. Different possible factors, such as words, part-of-speech tags, Brown word clusters, open class words and open class word clusters are explored. We find that Brown word clusters, part-of-speech tags and open-class words are most effective at reducing the perplexity of factored language models on the Mandarin-English Code-Switching corpus SEAME. In decoding experiments, the model containing Brown word clusters and part-of-speech tags and the model also including open class word clusters yield the best mixed error rate results. In summary, the factored language models can reduce the perplexity on the SEAME evaluation set by up to 10.8% relative and the mixed error rate by up to 3.4% relative.

Index Terms— language modeling, factored language models, Code-Switching speech

1. INTRODUCTION

The term Code-Switching (CS) denotes speech which contains more than one language. Speakers switch their language while they are talking. This phenomenon mainly appears in multilingual communities, such as immigrant communities. However, it increasingly occurs in former monolingual cultures due to globalization. In many countries, speakers switch between their native language and English within their utterances. This is a challenge for speech recognition systems which are typically monolingual. While there have been promising approaches to handle Code-Switching in the field of acoustic modeling, language modeling is still a great challenge. The main reason is a shortage of training data. Whereas about 60h of training data might be sufficient for the estimation of acoustic models, the transcriptions of these data are not enough to build reliable language models. Hence, the Code-Switching task can be regarded as under-resourced, especially in the context of language modeling.
This paper explores the use of factored language models to reduce this difficulty. The main advantages of factored language models compared to traditional n-gram approaches are the integration of features and the possibility of generalized backoff. They can improve language model results especially for languages with small training data sets. Hence, they are applied to the Code-Switching task in this study. The main contributions of this paper are the integration of factored language models into a dynamic decoder and the investigation of different features for the task of Code-Switching. The language models are evaluated in terms of perplexity and mixed error rate which is a combination of word error rate and character error rate.

2. RELATED WORK

This section describes previous work in the field of Code-Switching, language modeling for Code-Switching and factored language models. Furthermore, a study of obtaining vector representations for words is presented since they will be used to create additional features in this paper.

2.1. Code-Switching

In [1, 2, 3], it is observed that Code-Switching occurs at positions in an utterance where it does not violate the syntactical rules of the involved languages. On the one hand, Code-Switching can be regarded as a speaker dependent phenomenon [4, 5]. On the other hand, particular CS patterns are shared across speakers [6]. It can be observed that part-of-speech (POS) tags may predict Code-Switching points more reliably than words themselves. The authors of [7] predict CS points using several linguistic features, such as word form, language ID, POS tags or the position of the word relative to the phrase. The best result is obtained by combining the features. The authors of [8] compare four different kinds of n-gram language models to predict Code-Switching. They discover that clustering all foreign words into their POS classes leads to the best performance.
In [9], we adapted recurrent neural network language models to Code-Switching by adding features to the input vector and factorizing the output vector into language classes. These models do not only reduce the perplexities but also the mixed error rates when they are applied to rescore n-best lists.

2.2. Factored language modeling

Due to the possibility of integrating various features into factored language models, it is possible to handle rich morphology in languages like Arabic [10, 11]. In [12], the authors explore the perplexities of different factored language models on the Arabic CallHome corpus. Different features are investigated, such as words, morphological classes, word stems, word roots and vowel patterns. The authors of [11] also work on developing language models for Arabic. They use the GALE corpus and discover that a combination of morphological decomposition and factored language modeling yields the best results.
We performed initial experiments with factored language models for Code-Switching speech in [13]. It is shown that they outperform n-gram language models. Especially for the case of backoff to 2-grams, they prove their superior quality. The best performance is achieved by combining their estimates with recurrent neural network probabilities.

2.3. Semantic word representations

In [14], the authors explore the linguistic information in the word representation learned by a recurrent neural network. They discover that the network is able to capture both syntactic and semantic regularities. For example, the relationship of the vectors for "man" and "king" is the same as the relationship of the vectors for "woman" and "queen". In this paper, these word representations will be used to derive features for factored language models.

3. FACTORED LANGUAGE MODELING

Factored language models (FLMs) consider vectors of features (e.g. words, morphological classes, word stems or clusters) [12, 15]. The following equation expresses that a word w_t is regarded as a collection of factors $f_t^1, f_t^2, ...f_t^K$:

$$w_t \equiv \left\{f_t^1, f_t^2, ...f_t^K\right\} = f_t^{1:K} \quad (1)$$

Hence, not only words are used to predict the next word but a sequence of factors. The advantage of regarding factor histories instead of word histories is that usually there are fewer different factors than different words. Hence, the coverage of factor histories in training texts may be greater than that of equally long word histories. This is especially important for short training texts. Nevertheless, it is unlikely to see all factor history combinations. In case of unseen histories, generalized backoff is performed. Some of the factors in the history are dropped. For each omitted factor, a backoff result is calculated. If there is more than one result, the probabilities are combined, for instance using their average, their sum, or their product.

4. FEATURES FOR CODE-SWITCHING SPEECH

One design choice of factored language models is the selection of appropriate features. This section describes the features used in this paper.

4.1. Part-of-speech tags

Based on the results of previous studies [9, 13], part-of-speech tags are used as features for Code-Switching. To obtain part-of-speech tags of the mixed-speech training text, the tagging process as described in [9, 16] is applied. First, it selects a matrix and an embedded language [17]. In the corpus used in this paper, Mandarin is the matrix language and English the embedded language. Second, language islands are extracted. A language island is a sequence of three or more consecutive words of the embedded language. These language islands are passed to a monolingual English part-of-speech tagger [18] while all the remaining text is passed to a Mandarin part-of-speech tagger [19]. The idea is to provide the taggers with as much context as possible. Since the single English words in the Mandarin segments are falsely tagged as nouns in most cases, a post-processing step is applied. All English words which are not language islands are selected and their tags are replaced by the tags provided by the English part-of-speech tagger.

4.2. Brown word clusters

Due to missing references, the accuracy of the part-of-speech tagging process cannot be evaluated. The tags may contain errors and, therefore, the prediction of Code-Switching events based on these features may not be optimal. Therefore, the unsupervised clustering method by Brown et al. [20] is applied as an alternative to POS tagging. It assigns words to classes based on their distributions in a training text. Hence, its results may be more robust than part-of-speech tags in the case of Code-Switching. In initial experiments, Brown clusters with different numbers of classes are evaluated as factors in factored language models. The best perplexity results on the development set are achieved with 70 classes. Therefore, a cluster size of 70 is used in the following experiments.

4.3. Open class words

In addition to syntactical and distributional features, semantic information could be used in the factored language models. Since the Code-Switching corpus used in this paper contains no topic assignments, open class words are used to provide additional semantic information. Typically, words can be categorized into closed class words (function words) and open class words (content words). Closed class words specify grammatic relations rather than semantic meaning. Examples are conjugations, prepositions and determiners. The class of those words is called closed since their number is finite and

typically no new words are added to them. On the other hand, open class words express meaning, such as ideas, concepts or attributes. Their class is called open since it can be extended with new words, such as "Bollywood". It contains, for example, nouns, verbs, adjectives and adverbs [21]. In this study, the last preceeding open class word is determined for each word and added as a feature.

4.4. Open class word clusters

Finally, the open class words are clustered in order to summarize them into classes similar to topics. In order to create semantic classes, recurrent neural networks are applied (see Section 2.3). They are trained on monolingual English and Mandarin Gigaword text data[1] to increase the available training data. Afterwards, the word representation vectors are extracted and clustered into k classes using the Graclus implementation [22] of spectral clustering [23]. Given a similarity graph of the vectors, it performs a multilevel algorithm which consists of three steps: coarsening, base-clustering and refinement. In the first step, the similarity graph is transformed into graphs with a smaller number of nodes at different levels until $5 \cdot k$ nodes remain at the lowest level. At each level, nodes are combined to so-called supernodes depending on the weight of the edge between them and on their degree. In the second step, bisection clustering is performed on the coarsened graph. Finally, the graph is uncoarsened again. At each level, the lower level clustering results are used as initialization for weighted kernel k-means clustering.
If monolingual recurrent neural networks are trained as described above, clustering their word vectors results in monolingual English and monolingual Mandarin classes. In order to create bilingual clusters, a bilingual text is created by concatenating English and Mandarin text lines. During training of the recurrent neural network language model, the network is reset after each line. Hence, word vectors for both languages are combined in a single network. However, the reset ensures that the languages are not mixed since the English and Mandarin lines may not depend on each other. After training, the word vectors are extracted and clustered using Graclus. This results in bilingual word classes which will be called "BL spectral clusters" in the following. Another possibility to obtain bilingual classes is the usage of the Code-Switching corpus as training text for the recurrent neural network. Both approaches are compared in the following experiments.
In order to evaluate the usefulness of semantic classes in contrast to distribution based classes, the open class words of the monolingual texts and the Code-Switching corpus are clustered with the Brown method as well. To be able to distinguish them from the Brown clusters described in Section 4.2, they will be referred to as "oc Brown clusters".

[1]English Gigaword: LDC2011T07, Mandarin Gigaword: LDC2011T13

4.4.1. Examplary clustering results

In the following, some clustering results for English words are provided. The Brown clustering method assigns words like "sponsored", "granted" and "funded" to the same class but it also clusters words like "girlfriend" and "shack". The spectral clustering method, on the other hand, clusters "girlfriend" with "grandfather", "grandmother", "daddy", "nephew" and "aunt". It further finds semantic similarities like "gym", "swim", "ski" and "skiing" and it assigns "championships", "elephants", "olympics" and "stadium" to the same class. However, it also results in classes containing words like "death", "died", "confirmed" and "apartment". For those words, the Brown clustering methods detects no reasonable classes, either. Hence, both methods contain both meaningful and arbitrary-looking classes. This could be improved by using more training data.

5. EXPERIMENTS

This section first describes the data corpus used in this paper. Afterwards, results of an initial analysis of the different features are presented. Finally, perplexity results and mixed error rate results of the different FLMs are shown.

5.1. Data corpus

SEAME (South East Asia Mandarin-English) is a conversational Mandarin-English CS speech corpus recorded from Singaporean and Malaysian speakers by [24]. It was used for the research project "Code-Switch" jointly performed by Nanyang Technological University (NTU) and Karlsruhe Institute of Technology (KIT). The recordings consist of spontanously spoken interviews and conversations of about 63 hours. For this task, we deleted all hesitations and divided the transcribed words into four categories: English words, Mandarin words, particles (Singaporean and Malaysian discourse particles) and others (other languages). These categories are used as language information in the language models. The average number of CS points between Mandarin and English is 2.6 per utterance. The duration of monolingual segments is quite short: The average duration of English and Mandarin segments is only 0.67 seconds and 0.81 seconds respectively. In total, the corpus contains 9,210 unique English and 7,471 unique Mandarin words. We divided the corpus into three disjoint sets (training, development and test set) and assigned the data based on several criteria (gender, speaking style, ratio of Singaporean and Malaysian speakers, ratio of the four categories, and the duration in each set). Table 1 lists the statistics of the corpus.

5.2. Feature analysis

Before the features described in Section 4 are used for language modeling, their Code-Switching rates are evaluated.

Table 1. Statistics of the SEAME corpus

	Training set	Development set	Evaluation set
# Speakers	139	8	8
Duration(hrs)	59.2	2.1	1.5
# Utterances	48,040	1,943	1,018
# Token	525,168	23,776	11,294

The Code-Switching rate for a feature is calculated by dividing its frequency in front of Code-Switching points by its frequency in the entire text.

$$\text{CS rate} = \frac{\text{frequency in front of CS point}}{\text{total frequency}} \quad (2)$$

The Code-Switching rate can be regarded as the probability that the feature triggers a Code-Switching event. Table 2 shows the range of the Code-Switching rates for the different features.

Table 2. Overview of the maximum Code-Switching rates of different potential trigger features

Feature	Man → En CS	En → Man CS
Words	≤ **53.43%**	≤ 56.25%
Part-of-speech tags	≤ 43.13%	≤ 47.78%
Brown word cluster	≤ 52.73%	≤ **72.67%**
Open class words	≤ 33.33%	≤ 54.53%
Open class clusters	≤ 34.44%	≤ 56.66%

In average, the Brown word clusters seem to be the most promising features for the prediction of Code-Switching points.

5.3. Perplexity experiments

For each feature and several feature combinations, a factored language model is built and evaluated. The backoff parameters are manually optimized on the SEAME development set. Table 3 presents the perplexity results of the different language models. The language models contain words and the features mentioned in the table as conditioning factors. The time steps of the factors range from one to two time steps in the past. This has been chosen based on experimental results.

Consistent with the feature analysis results, the Brown word clusters lead to the greatest perplexity reductions. The best perplexities are obtained by combining them with part-of-speech tags and open class words. While clustering seems to help when applied to all words, it does not improve the performance when it is applied to open class words only.
In the following part, the different clustering methods for open class words as described in Section 4 are evaluated in detail. For this purpose, factored language models are built using words, part-of-speech tags, Brown word clusters and open class word clusters as factors. They are then evaluated in terms of perplexity on the SEAME development set. Table 4

Table 3. Summary: PPLs of different FLMs
"oc" abbreviates the term "open class"

Model	PPL dev	PPL eval
Baseline (3-gram)	268.39	282.86
POS	260.70	267.86
POS + LID	257.62	264.20
Brown clusters	257.17	265.50
Brown clusters + POS	249.00	255.34
Brown clusters + LID	260.39	268.71
Brown clusters + POS + LID	251.39	259.05
Oc words	278.12	281.31
Oc words + Brown clusters + POS	**247.18**	**252.37**
Oc clusters + Brown clusters + POS	247.24	252.60

presents the results of these experiments. For each method, different clustering sizes are tested. If the SEAME training text is used for clustering, the classes are called CS classes. Otherwise, they are referred to as EN- and MAN-classes.

Table 4. PPL results using different open class word clustering methods

Model	PPL dev	PPL eval
Oc words (unclustered, about 16k oc words)	**247.18**	**252.37**
Oc Brown cluster 400 CS classes	250.06	255.08
Oc Brown cluster 2000 CS classes	248.07	253.01
Oc Brown cluster 4000 CS classes	247.52	252.44
Oc Brown cluster 6000 CS classes	247.47	252.53
Oc Brown cluster 200 EN + 200 MAN classes	251.42	256.71
Oc Brown cluster 1000 EN + 1000 MAN classes	248.40	253.78
Oc Brown cluster 2000 EN + 2000 MAN classes	247.89	252.84
Oc Brown cluster 3000 EN + 3000 MAN classes	247.56	252.61
Spectral cluster 1000 CS classes	251.61	255.87
Spectral cluster 2000 CS classes	250.53	254.65
Spectral cluster 3000 CS classes	249.97	254.65
Spectral cluster 1000 EN + 1000 MAN classes	248.93	254.07
Spectral cluster 2000 EN + 2000 MAN classes	248.31	252.94
Spectral cluster 3000 EN + 3000 MAN classes	248.02	252.69
BL spectral cluster 500 classes	248.12	253.35
BL spectral cluster 800 classes	247.24	252.60

The results show that none of the clustering methods is able to improve the model using unclustered open class words. Furthermore, the more classes are used, the better are the perplexity results. More classes mean less words per class and, therefore, approximate the unclustered words. An explanation for this could be the increased branching factor after a cluster compared to the branching factor after a word. Within the different clustering methods, the bilingual spectral clusters perform the best, followed by the oc Brown clusters built with the Code-Switching corpus. While the clusters based on the Code-Switching corpus contain all possible words of the development and evaluation set (apart from out-of-vocabulary words), the clusters built from the monolingual data cannot cover all words of the Code-Switching vocabulary. For example, the BL spectral 800 classes do not contain about 2200 of 3000 open class words of the development set and 1300 of 2000 of the evaluation set. This could be another reason why the clusters do not reduce the perplexity values achieved

by the model with the unclustered open class words. In the following decoding experiments, the bilingual spectral cluster with 800 classes is used as open class word clusters since it provided the best perplexity results.

5.4. Decoding experiments

5.4.1. Description of the ASR system

For decoding, the state-of-the-art decoder BioKIT is used. It is a dynamic decoder. The acoustic model is speaker independent and applies a fully-continuous 3-state left-to-right HMM. The emission probabilities are modeled with bottleneck features [25]. The phone set contains all English and Mandarin phones including tags for continuous speech (+noise+, +breath+, +laugh+) and an additional phone +particle+ to model Singaporean and Malaysian particles. For context dependent acoustic modeling, the decision tree splitting process is stopped at 3,500 quintphones. Then, merge-and-split training is applied followed by three iterations of Viterbi training. To obtain a dictionary, the CMU English [26] and Mandarin pronunciation dictionaries [27] are merged into one bilingual pronunciation dictionary. The number of dictionary entries is 56k. Additionally, we apply several rules from [28] which might delete or change a phone to generate pronunciation variants for Singaporean English. As a baseline language model, a trigram language model is built from the SEAME training transcriptions using the SRILM toolkit [29]. This model is interpolated with two monolingual language models that has been created from 350k English sentences from NIST and 400k Mandarin sentences from the GALE project which has been collected from online newspapers. More information about the baseline decoding system can be found in [30]. As an evaluation measure, we use the mixed error rate [30]. It is a combination of word error rates for English segments and character error rates for Mandarin segments. Due to this, the performance can be compared across different segmentations of Mandarin.

5.4.2. Integration of FLMs into the decoding process

In order to perform decoding experiments with factored language models, the BioKIT decoder is extended to support such language models and generalized backoff. The best results are obtained when using a traditional n-gram language model for language model lookahead and combining the factored language model probabilities and the n-gram model probabilities at every word end. For combination, linear interpolation is applied. Figure 1 shows mixed error rate results on the development set for different interpolation weights using the factored language model with words and part-of-speech tags as factors. These weight experiments are performed on a subset of the development set (on about 20% of all sentences) in order to reduce computational efforts. They result in a factored language model weight of 0.55.

Fig. 1. Mixed error rate results for different FLM weights

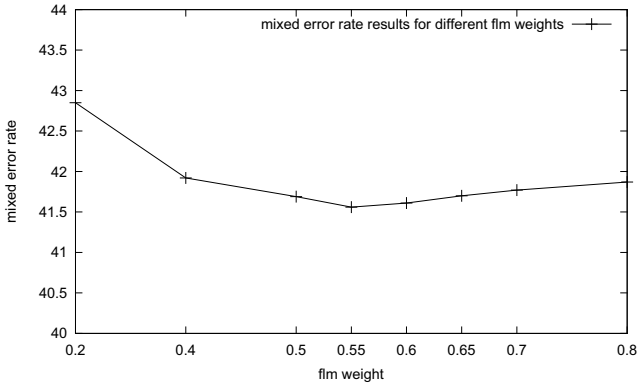


Table 5 presents the mixed error rate results for decoding with the different factored language models. For each factor combination, the factored language model with the least perplexity on the development set is used. The factored language models are used with a language model weight of 60 while for the baseline 3-gram model, the language model weight is set to 50. This has been chosen based on experimental results.

Table 5. Mixed error rate results for the different FLMs when the FLMs are interpolated with the baseline 3gram using an FLM interpolation weight of 0.55

Model	MER dev	MER eval
Decoder-baseline 3-gram	39.96%	34.31%
POS	39.47%	33.46%
POS+LID	39.66%	33.30%
Brown clusters	39.45%	33.93%
Brown clusters + POS	**39.30%**	33.60%
Brown clusters + POS + LID	39.39%	33.16%
Oc words + Brown clusters + POS	39.33%	**33.15%**
ML spectral 800 oc clusters + Brown cl + POS	**39.30%**	33.16%

To be able to better interpret the results and improvements provided by the factored language models, an analysis is performed. The mixed error rates of the baseline model and the FLM Brown clusters + POS are compared in detail. The results show that the FLM improves the recognition both for monolingual segments and for Code-Switching points. In particular, the mixed error rate for English segments is reduced from 59.40% to 57.52%. The mixed error rate for Mandarin segments is changed from 36.48% to 36.12%. Especially, the number of Mandarin insertions is improved and less English words are confused with Mandarin words. While the baseline model recognizes 1724 words at Code-Switching points correctly, the decoding with the FLM leads to 1737 correct words at Code-Switching points.

6. CONCLUSION

This paper described our investigations of features for Code-Switching language modeling. They were tested as factors in factored language models and evaluated in terms of language model perplexity and mixed error rate. The combination of words, Brown word clusters, part-of-speech tags and open class words yielded the best perplexity results on the SEAME development set. The corresponding factored language model reduced the perplexity by 10.8% relative on the evaluation set. Clusters of open class words were also investigated as factors but did not yield any improvement in terms of perplexity over individual open class words. In order to perform decoding experiments, factored language models were integrated into the state-of-the-art dynamic decoder BioKIT. They were used in combination with a traditional 3-gram language model. The factored language model containing Brown word clusters and part-of-speech tags as factors and the model also including open class word clusters achieved the best mixed error rate results. The mixed error rates could be improved by up to 3.4% relative.

7. REFERENCES

[1] S. Poplack, *Syntactic structure and social function of code-switching*, vol. 2, Centro de Estudios Puertorriqueños,[City University of New York], 1978.

[2] E.G. Bokamba, "Are there syntactic constraints on code-mixing?," *World Englishes*, vol. 8, no. 3, pp. 277–292, 1989.

[3] P. Muysken, *Bilingual speech: A typology of code-mixing*, vol. 11, Cambridge University Press, 2000.

[4] P. Auer, "From codeswitching via language mixing to fused lects toward a dynamic typology of bilingual speech," *International Journal of Bilingualism*, vol. 3, no. 4, pp. 309–332, 1999.

[5] N.T. Vu, H. Adel, and T. Schultz, "An investigation of code-switching attitude dependent language modeling," in *Proc. of SLSP*, 2013.

[6] S. Poplack, "Sometimes I'll start a sentence in spanish y termino en español: toward a typology of code-switching," *Linguistics*, vol. 18, no. 7-8, pp. 581–618, 1980.

[7] T. Solorio and Y. Liu, "Learning to predict code-switching points," in *Proc. of the Conference on Empirical Methods in Natural Language Processing*. ACL, 2008, pp. 973–981.

[8] J.Y.C. Chan, PC Ching, T. Lee, and H. Cao, "Automatic speech recognition of Cantonese-English code-mixing utterances," in *Proc. of Interspeech*, 2006.

[9] H. Adel, N.T. Vu, F. Kraus, T. Schlippe, H. Li, and T. Schultz, "Recurrent neural network language modeling for code switching conversational speech," in *Proc. of ICASSP*. IEEE, 2013.

[10] K. Duh and K. Kirchhoff, "Automatic learning of language model structure," in *Proc. of the 20th international conference on Computational Linguistics*. ACL, 2004, p. 148.

[11] A. El-Desoky, R. Schlüter, and H. Ney, "A hybrid morphologically decomposed factored language models for arabic LVCSR," in *Proc. of HLT-NAACL*. ACL, 2010, pp. 701–704.

[12] K. Kirchhoff, J.A. Bilmes, and K. Duh, "Factored language models tutorial," Tech. Rep. UWEETR-2008-004, University of Washington, EE Department, 2007.

[13] H. Adel, N.T. Vu, and T. Schultz, "Combination of recurrent neural networks and factored language models for code-switching language modeling," in *Proc. of ACL*, 2013.

[14] T. Mikolov, W.-T. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proc. of HLT-NAACL*. ACL, 2013, pp. 746–751.

[15] J.A. Bilmes and K. Kirchhoff, "Factored language models and generalized parallel backoff," in *Proc. of HLT-NAACL*. ACL, 2003, pp. 4–6.

[16] T. Schultz, P. Fung, and C. Burgmer, "Detecting code-switch events based on textual features," Diploma thesis, 2009.

[17] C.M. Scotton, *Duelling languages: Grammatical structure in codeswitching*, Oxford University Press, 1997.

[18] K. Toutanova, D. Klein, C.D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. of HLT-NAACL*. ACL, 2003, pp. 173–180.

[19] H. Tseng, P. Chang, G. Andrew, D. Jurafsky, and C.D. Manning, "A conditional random field word segmenter," *fourth SIGHAN Workshop on Chinese Language Processing*, 2005.

[20] P.F. Brown, P.V. Desouza, R.L. Mercer, V.J.D. Pietra, and J.C. Lai, "Class-based n-gram models of natural language," *Computational Linguistics*, vol. 18, no. 4, pp. 467–479, 1992.

[21] V. Fromkin, *An introduction to language*, Cengage Learning, 2013.

[22] I.S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 11, pp. 1944–1957, 2007.

[23] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[24] D.C. Lyu, T.P. Tan, E.S. Chng, and H. Li, "An analysis of a Mandarin-English code-switching speech corpus: SEAME," *Age*, vol. 21, pp. 25–8, 2010.

[25] N.T. Vu, F. Metze, and T. Schultz, "Multilingual bottleneck features and its application for under-resourced languages," in *Proc. of SLTU*, 2012.

[26] "CMU pronouncation dictionary for English," `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

[27] R. Hsiao, M. Fuhs, Y.C. Tam, Q. Jin, and T. Schultz, "The CMU-InterACT 2008 mandarin transcription system," in *Proc. of ICASSP*, 2008.

[28] W. Chen, Y. Tan, E. Chng, and H. Li, "The development of a Singapore English call resource," *Oriental COCOSDA, Nepal*, 2010.

[29] A. Stolcke et al., "SRILM-an extensible language modeling toolkit," in *Proc. of SLP*, 2002, vol. 2, pp. 901–904.

[30] N.T. Vu, D.C. Lyu, J. Weiner, D. Telaar, T. Schlippe, F. Blaicher, E.S. Chng, T. Schultz, and H. Li, "A first speech recognition system for Mandarin-English code-switch conversational speech," in *Proc. of ICASSP*. IEEE, 2012, pp. 4889–4892.