Variable-Span Out-of-Vocabulary Named Entity Detection

Wei Chen, Sankaranarayanan Ananthakrishnan, Rohit Prasad, and Prem Natarajan

Speech, Language and Multimedia Business Unit
Raytheon BBN Technologies
Cambridge, MA 02138, U.S.A.
{wchen,sanantha,rprasad,pnataraj}@bbn.com

Abstract

Out-of-vocabulary named entities (OOV NEs) are always misrecognized by fixed-vocabulary automatic speech recognition (ASR) systems. This has a negative impact on downstream applications such as language understanding and machine translation (MT). Automatic detection of OOV NEs in ASR hypotheses can help mitigate this problem by triggering the use of alternative approaches to acquire and process these NEs. State-of-the-art OOV NE detection typically involves tagging ASR-hypothesized words using a sequence model, such as conditional random fields (CRF), in conjunction with a variety of contextual and ASR-derived features. In this paper, we propose a novel variable-span tagging approach for detecting OOV NEs. Instead of tagging individual words in ASR hypotheses, we directly tag longer spans of consecutive words. The proposed approach outperforms a state-of-the-art CRF tagger on two distinct held-out test sets with different OOV NE distributions. On a 5.1K-word test set rich in OOV NEs, our method achieves 56.1% detection rate at 10% false alarm rate (vs. 52.1% for the CRF detector). On a 39.4K-word test set with a natural distribution of OOV NEs, we obtain 73.0% detection rate at 10% false alarm rate (vs. 69.5% for the CRF detector). In all cases, OOV NEs are completely unobserved in our training data.

Index Terms: named entity detection, OOV detection, automatic speech recognition

1. Introduction

Named entities (NEs) play a crucial, information-bearing role in spoken language [1]. Because they often refer to key persons, locations, organizations, etc., the inability to understand or communicate NEs usually has a deleterious effect on spoken interactions, including human-machine interactions (e.g. dialogue systems) [2] as well as machine-mediated human-human communication (e.g, conversational spoken language translation (CSLT) systems) [3].

NEs pose a significant problem for traditional spoken language systems based on automatic speech recognition (ASR) systems. The sheer number and variety of NEs makes it impossible to ensure full coverage by a typical fixed-vocabulary ASR. Consequently, out-of-vocabulary named entities (OOV NEs) are guaranteed to be misrecognized by ASR, which usually substitutes or inserts a sequence of phonetically similar in-vocabulary words that best match the linguistic context of the hypothesis. This causes significant problems for downstream applications. In a CSLT system, for instance, misrecognized OOV NEs in the source language often produce meaningless target translations, which can lead to miscommunication and/or stalling of the conversation. Figure 1 shows an example of a misrecognized OOV NE.

The OOV NE problem can be somewhat mitigated by automatically identifying portions of the ASR hypotheses that might correspond to OOV NEs. Knowledge that the spoken utterance contains an OOV NE in a specific location can be leveraged to trigger alternative strategies to acquire and process the NE. For example, we previously leveraged such capability in an English-to-Iraqi CSLT system to splice the speech segment corresponding to detected OOV NEs directly into the target language speech [4]. This enabled our system to perform cross-lingual transfer of critical NEs even though they were unable to be translated via the traditional ASR/MT route due to guaranteed recognition failure.

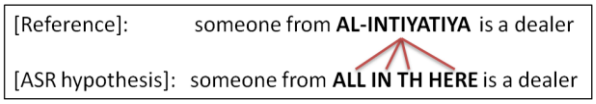

Figure 1. *Example ASR error caused by an OOV NE*

There is a large body of existing literature on OOV detection [e.g., 5, 6-10] as well as NE tagging [e.g., 11, 12-18] in speech, but significantly less prior work focusing on the compound OOV NE detection problem. Existing approaches typically combine information from both OOV detection and NE tagging. For example, Parada et al. [19] used an OOV detector to identify ASR-hypothesized words likely to be caused by OOVs while recognizing NEs in speech. Kumar et al. [20] improved OOV NE detection by applying ASR confidence scores (word posterior probabilities) to signal possibly misrecognized words. Because these existing approaches are adapted from the NE tagging literature, they share the common characteristic of tagging each individual word in the ASR hypothesis using a point-classifier such as maximum-entropy (maxent), or sequence tagger such as conditional random fields (CRF) [21]. State-of-the-art performance has been achieved with CRF-based tagging [19].

However, as shown in Figure 1, we observe that ASR often breaks OOV NEs down into multiple in-vocabulary words. Thus, a drawback of the *single-word tagging* approach is that it fails to capture contextual information that can help identify a span of consecutively misrecognized words corresponding to an OOV NE. We address the above shortcoming of existing OOV NE tagging approaches by focusing on variable-length word spans as atomic units for OOV NE detection. Thus, our training and inference instances consist of multi-word spans in the ASR hypothesis. We use a simple maxent classifier for tagging each span, and collate the predictions to obtain word-level decisions. We compare this approach to a state-of-the-art CRF-based system using a rich set of contextual and ASR-derived features.

The remainder of this paper is organized as follows. Section 2 describes our training corpus and evaluation sets. Section 3 describes various baseline OOV NE detectors, including a state-of-the-art CRF-based system. In Section 4, we present the novel variable-span OOV NE detection approach. Experimental results are summarized in Section 5.

2. Training and Evaluation Data

OOV NE detection operates on ASR hypotheses. We use the BBN Byblos ASR system to transcribe speech automatically. The system uses a multi-pass decoding strategy in which models of increasing complexity are used in successive passes to obtain gradually refined recognition hypotheses [22]. The acoustic model (AM) was trained on approximately 200 hours of manually transcribed conversational English speech, which consisted of 129K utterances segmented by sentence boundaries. These speech data came from the English side of the DARPA TransTac two-way spoken dialogue collections covering various domains, including force protection (e.g., checkpoint, reconnaissance, and patrol), medical diagnosis, aid, maintenance, infrastructure, and others [23]. We trained the language model (LM) on 6M sentences with 60M words, drawn from both the TransTac domain and out-of-domain sources. We obtained 11% word error rate (WER) on a held-out test set of 3,138 utterances (38K words). Besides 1-best hypotheses, the decoder also generates confusion networks (word graphs) from the decoding lattice.

2.1. Training Corpus

A subset of the reference transcriptions of the 200 hours of English speech was manually annotated with NE labels. We employed the jack-knifing technique to generate a training corpus for the OOV NE detector. We divided the 200 hours of English speech with reference transcriptions into ten equal partitions. Each partition was decoded with an LM that left out transcriptions for that partition (ten different LMs were trained, one for each partition). The global baseline AM was used for decoding all partitions. OOV NE errors were elicited in each partition by excluding labeled NEs in that partition from the corresponding decoding lexicon and LM. Word-level reference OOV NE labels were generated for the entire training set (combining all jack-knife partitions) by automatically aligning the ASR hypotheses with corresponding reference transcriptions (i.e., we labeled a word in the ASR hypothesis as an OOV NE error if it was aligned to an OOV NE in the corresponding reference transcription).

Of the 129K ASR-decoded English utterances in the jack-knifed training set, only 6,235 utterances contained OOV NEs (a total of 7,765 NE tokens, corresponding to 0.5% of all reference tokens in the corpus). Regardless of the approach used, initial experiments indicated that models trained only on the instances containing OOV NEs outperform those trained on the entire corpus. Thus, we trained all OOV NE detectors on the subset of 6,235 instances containing these errors. The first row of Table 1 summarizes the ground-truth labeled corpus used for training the OOV NE detectors.

Since errors caused by OOV NEs are often "bursty", we used the so-called *BI* encoding for the reference labels - the first word in an OOV NE error was labeled *BOOVNE* for "beginning of error caused by an OOV NE", and all the following consecutive errors were labeled *OOVNE*. This encoding is particularly beneficial for sequence taggers such as CRF. For maxent-based point-classification, we collapsed these tags to a single *OOVNE* label.

2.2. Evaluation Sets

We evaluated OOV NE detection on two different configurations. In the first configuration, the held-out development and test set utterances were deliberately constructed to be rich in previously unseen OOV NEs. These sets also contained other challenges for ASR, including OOV non-names, phonetically confusable words (homophones), mispronunciations, etc.

In the second configuration, we evaluated OOV NE detection performance on held-out development and test sets drawn from the DARPA TransTac collection. These sets resemble the ASR training corpus and contain a more natural distribution of OOV NEs as observed during the course of real conversations in the domains of interest. ASR hypotheses of both evaluation sets were hand-annotated with OOV NE tags.

The last four rows of Table 1 summarize the development and test sets used in both configurations. As expected, the OOV NE rate, and consequently ASR word error rate (WER) on the problem-rich sets corresponding to the first configuration are quite high. On the other hand, the sets with natural distribution of NEs have significantly lower OOV NE rate and ASR WER. We emphasize that, in both configurations, our evaluation framework mirrors a real-world scenario where OOV NEs are not simulated but are completely unknown to and previously unobserved by both ASR model training as well as the OOV NE detection training process.

Table 1. *Training & evaluation datasets*.

Corpus	*Size (words)*	*OOV NE rate*	*WER*
`Training`	86.9K	8.9%	24.0%
`Dev-Rich`	23.6K	1.6%	26.7%
`Test-Rich`	5.1K	5.7%	41.5%
`Dev-Natural`	45.3K	0.7%	13.6%
`Test-Natural`	39.4K	0.4%	13.3%

3. Baseline OOV NE Detection

In order to establish a baseline for comparison to our proposed approach, we implemented a CRF-based OOV NE tagger similar to the system recently shown by Parada et al. [19] to perform well on this task. The baseline CRF OOV NE detector was trained using CRF++ [24]. CRF models the sequential characteristics of OOV NE errors in ASR hypotheses (e.g, OOV NE errors are likely to occur in bursts). We incorporated features commonly used for both OOV detection and NE recognition [e.g., 19, 20]. The following features were computed for every ASR hypothesized word:

Current Word identity: The current ASR hypothesized word.

Part-of-speech (POS): We automatically tagged the ASR hypotheses with their parts-of-speech using the Stanford tagger [25].

ASR confidence: Word posterior probabilities (WPP) computed from acoustic and language model scores using the forward-backward algorithm on a word lattice. Higher WPP implies greater ASR "confidence" in the hypothesized word.

Language model salience: We used the negative LM log-likelihood of the candidate word given its *n*-gram context applied during ASR decoding. High perplexity implies mismatched linguistic context.

Word vs. graphone decoding disagreement: We computed the phonetic edit distance between hypotheses generated by word and sub-word ASR systems. We used graphones [26] as sub-word units due to their robustness (lower error rate)

compared to phoneme decoding. Graphones were automatically learned and inferred from letter-to-phoneme alignment obtained using standard statistical machine translation (SMT) phrase extraction techniques [27]. Maximum length of graphemes was set to three letters.

ASR error posterior: Posterior scores generated by a CRF-based ASR error detector combing all the features described above, as well as additional features including the density of the word confusion network (i.e., the number of competing hypothesis words in a given slot), phonetic acoustic model scores, word boundary detector posterior, and homophone indicator features [28]. At 10% false positive rate, the ASR error detection rate is 68.3% for the test set rich in OOV NEs, and 61.2% for the test set with natural OOV NE distribution.

Context: All the above features from the previous and the following *n* words. We tuned *n*, ranging from 0 to 3, on the held-out development data sets. The highest OOV NE detection rate was achieved when *n* is set to 2.

Because CRF++ is designed to run with categorical features, we discretized all real-valued features into 100 bins, each containing roughly the same number of instances. The number of bins is chosen empirically, and may require tuning if used on a different corpus, task, or domain. We tuned various model parameters, including the CRF order (first and second), context size of neighboring features, and the extent of overfitting, on the held-out development sets to maximize detection rates. For both the *rich* and the *natural* OOV NE test sets, the best development-set performance was achieved using a second order CRF to model dependency between successive labels, a context of five words (i.e., features from the two prior, current, and two successive hypothesized words), and 0.1 overfitting. This is the CRF-ASR baseline.

In addition to this strong baseline system, we also implemented a maxent-based word-level detector trained on the above OOV NE corpus (MAXENT-ASR); a CRF-based word-level NE detector trained on ground truth transcriptions of the OOV NE corpus (CRF-REFERENCE); and a maxent-based word-level NE detector trained on ground truth transcriptions of the OOV NE corpus (MAXENT-REFERENCE). We used the same feature set and training utterances for building all detectors.

4. Variable-Span OOV NE Detection

As the example in Figure 1 demonstrates, ASR misrecognition of OOV NEs always generates a contiguous sequence of one or more incorrectly hypothesized words. Instead of tagging each misrecognized word individually, it makes intuitive sense to tag the entire *span* as corresponding to an OOV NE. By treating the span as an atomic unit for labeling OOV NEs, we are able to better capture contextual information that indicates the presence of a NE in the utterance. Because we are now classifying individual spans rather than tagging word sequences, we employ a simpler point-wise maxent classifier trained on the same feature set used by the CRF baseline.

Labeled spans corresponding to OOV NE errors are available in the training corpus. We begin by identifying all OOV NE and non-OOV NE regions in the training instances. For the example ASR hypothesis shown in Figure 1, "*someone from*" and "*is a dealer*" are two non-OOV NE regions separated by an OOV NE region "*all in th here*". For each region, we find all spans of consecutive words up to a specified maximum length and assign to each span the label of the containing region. Table 2 illustrates span-level training instances for the above example, up to span length of 3.

We extract features by considering each span as an atomic unit. Thus, all span-internal words and corresponding POS tags are mapped to "current word identity" features. Words and corresponding POS tags to the left and right of the span are considered its "lexical context". For instance, the immediate left context of the OOV NE span "*all_in_th*" is "*from*", whereas the immediate right context of the OOV NE span "*in_th_there*" is "*is*". Span-level real-valued features (e.g, "ASR confidence", "LM salience", "ASR error posterior", etc.) are computed as simple arithmetic averages of the corresponding feature values over all words in the span. We trained a maxent classifier with the L-BFGS algorithm. Maxent Gaussian priors were tuned to be 0.05 on both development sets.

Table 2. *Example training instances of variable span lengths.*

L	***Non-OOVNE*** *someone from*	***OOVNE*** *all in th here*	***Non-OOVNE*** *is a dealer*
1	*someone*, *from*	*all*, *in*, *th*, *here*	*is*, *a*, *dealer*
2	*someone_from*	*all_in*, *in_th*, *th_here*	*is_a*, *a_dealer*
3	*n/a*	*all_in_th*, *in_th_here*	*is_a_dealer*

Labeled regions are unavailable during inference on development/test sets. Therefore, we iteratively tag each span in the ASR hypotheses with the maxent classifier, up to the maximum span length specified in training. In the first pass, we tag all single-word spans; in the second pass, we tag all two-word spans; and so on. For each span, the classifier returns the posterior probability of that span corresponding to an OOV NE. We aggregate classifier posteriors for all spans covering a candidate hypothesis word to decide if it corresponds to an OOV NE misrecognition.

We observed that the variable-span maxent classifier tends to overestimate OOV NE posteriors for longer spans, possibly because they trigger more "current word identity" feature functions. To alleviate this issue, we applied span length-specific scaling factors to the estimated OOV NE posteriors. The final score for each hypothesized word is computed as the sum of scaled posteriors over all the spans covering that word.

We tuned the maximum span length and span length-specific scaling factors on the development sets. The optimal maximum span length was determined to be three words. Optimal scaling factors for single-word, two-word, and three-word spans were 0.2, 0.5, and 0.3 for the OOV NE-rich configuration; 0.4, 0.5, and 0.1 for the configuration with natural distribution of OOV NEs. These scaling factors were applied to posteriors estimated on test-set spans to obtain word-level decisions. For reporting purposes, we refer to this approach as SPANMAXENT-ASR.

5. Experimental Results and Discussion

We compared the proposed SPANMAXENT-ASR approach to the state-of-the-art CRF-based detector (CRF-ASR) and with the three other approaches (MAXENT-ASR, CRF-REFERENCE, MAXENT-REFERENCE) described in Section 3. In each case, systems were evaluated on the unseen test sets after parameter tuning on the corresponding development sets.

We note that even though SPANMAXENT-ASR performs inference on variable-length spans, we evaluate its detection performance at the word level in order to be consistent with the competing methods that operate at this level.

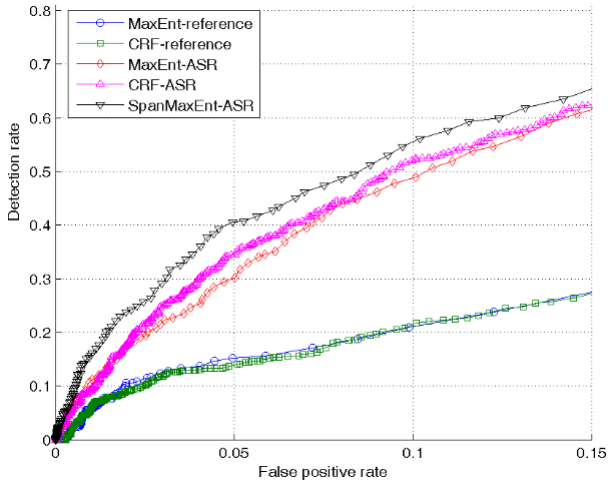

Figure 2. Comparison of detection performance on OOV NE-rich test set (Test-Rich).

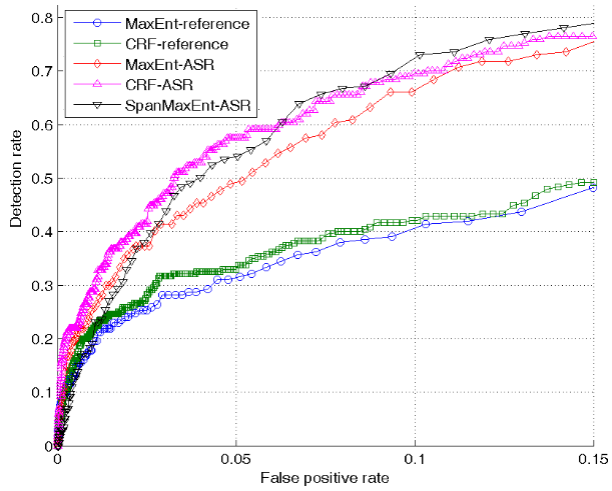

Figure 3. Comparison of detection performance on naturally distributed OOV NEs (Test-Natural).

Figures 2 and 3 show the receiver operating characteristic (ROC) curves for all five approaches on the rich and natural OOV NE distribution configurations, respectively. We focus on a low false positive rate range of 0-15%, where we typically operate the OOV NE detector. On the OOV NE-rich test set (Test-Rich), SPANMAXENT-ASR strongly outperforms the baseline CRF-ASR over the entire region of interest. On the test set with naturally distributed OOV NEs (Test-Natural), SPANMAXENT-ASR beats CRF-ASR over more than half the region of interest, but appears to be less robust on the low-false-alarm-rate region from 0-6%. MAXENT-ASR is a close third in all cases, performing slightly worse than CRF-ASR as expected. The last two methods, viz. CRF-REFERENCE and MAXENT-REFERENCE, trained on ground truth transcriptions of the NE-tagged training corpus, perform significantly worse than the top three, which are trained on jack-knifed ASR hypotheses of the training set. This is due to mismatch between training and testing conditions; NEs and their context are always observed during training, but are significantly corrupted during testing.

Table 3 summarizes, for both test configurations, the OOV NE detection rate at a fixed false alarm rate of 10% for all five methods compared in this work. At this operating point, the detection rate of SPANMAXENT-ASR is significantly superior to the CRF-ASR baseline by 4.0% absolute (7.7% relative) on Test-Rich ($p < 0.001$ according to a two-tailed paired permutation test with 10,000 random permutations); whereas on Test-Natural, it outperforms CRF-ASR marginally by 3.5% absolute (5.0% relative, $p = 0.056$). MAXENT-ASR trails CRF-ASR by 2-3%, while the traditional clean-text NE taggers lag considerably.

Table 3. *Detection rates of the five approaches on the two test configurations at a fixed false positive rate of 10%.*

Method	*Test–Rich*	*Test–Natural*
MAXENT-REFERENCE	21.0%	41.4%
CRF-REFERENCE	21.8%	42.1%
MAXENT-ASR	48.9%	67.2%
CRF-ASR	52.1%	69.5%
SPANMAXENT-ASR	56.1%	73.0%

6. Conclusions

Automatic detection of OOV NEs is critical for many ASR-driven spoken language systems, because it can serve to trigger alternative strategies for acquiring and processing NEs that are guaranteed to be misrecognized by ASR. Existing approaches view OOV NE detection as a word-sequence labeling problem, similar to clean-text NE tagging. The best results in the literature have been obtained using CRF-based sequence tagging. However, these approaches tend to overlook important contextual information for OOV NE error spans, thereby producing sub-optimal results.

In this paper, we recognized that OOV NE detection is a span-labeling problem. With this view, we proposed a novel, variable-span OOV NE tagging method to label spans of consecutive words in ASR hypotheses. We compared the proposed approach to a state-of-the-art CRF-based system, as well as three other systems, including standard clean-text NE taggers. We evaluated each method on two different configurations with different positive class distributions – one rich in OOV NEs, and the other with a much lower frequency, natural distribution of OOV NEs. We showed that the variable-span OOV NE detector outperforms the state-of-the-art CRF baseline in both configurations over a majority of the region of interest in the ROC curve. While the distribution of OOV NEs in test data can affect performance of the variable-span detector in its current implementation, our experiments clearly demonstrated the effectiveness of the span-based view for OOV NE detection in ASR hypotheses.

7. Acknowledgements

This paper is based upon work supported by the DARPA BOLT Program. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

8. References

[1] Grishman, R. and Sundheim, B., "Message Understanding Conference - 6: A Brief History," in *International Conference on Computational Linguistics*, 1996.

[2] Béchet, F., Gorin, A. L., Wright, J. H., and Tür, D. H., "Detecting and Extracting Named Entities from Spontaneous Speech in a Mixed-Initiative Spoken Dialogue Context: How May I Help You?," *Speech Communication,* vol. 42, pp. 207-225, 2004.

[3] Prasad, R., Moran, C., Choi, F., Meermeier, R., Saleem, S., Kao, C.-L., Stallard, D., and Natarajan, P., "Name Aware Speech-to-Speech Translation for English/Iraqi," In *proceedings of the IEEE Spoken Language Technology Workshop*, 2008.

[4] Prasad, R., Kumar, R., Ananthakrishnan, S., Chen, W., Hewavitharana, S., Roy, M., Choi, F., Challenner, A., Kan, E., Neelakantan, A., and Natarajan, P., "Active Error Detection and Resolution for Speech to Speech Translation," In *proceedings of the International Workshop on Spoken Language Translation*, 2012.

[5] Parada, C., Dredze, M., Filimonov, D., and Jelinek, F., "Contextual Information Improves OOV Detection in Speech," In *proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2010.

[6] Bazzi, I., "Modeling Out-of-Vocabulary Words for Robust Speech Recognition," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2002.

[7] Hayamizu, S., Itou, K., and Tanaka, K., "Detection of Unknown Words in Large Vocabulary Speech Recognition," In *proceedings of the Eurospeech*, 1993.

[8] Yazgan, A. and Saraclar, M., "Hybrid Language Models for Out of Vocabulary Word Detection in Large Vocabulary Conversational Speech Recognition," In *proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2004.

[9] Bisani, M. and Ney, H., "Open Vocabulary Speech Recognition with Flat Hybrid Models," In *proceedings of the International Speech Communication Association (INTERSPEECH)*, 2005.

[10] Qin, L. and Rudnicky, A., "OOV Word Detection using Hybrid Models with Mixed Types of Fragments," In *proceedings of the International Speech Communication Association (INTERSPEECH)*, 2012.

[11] Gravano, A., Jansche, M., and Bacchiani, M., "Restoring Punctuation and Capitalization in Transcribed Speech," In *proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009.

[12] Huang, F., "Multilingual Named Entity Extraction and Translation from Text and Speech," Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh, PA, 2005.

[13] Sudoh, K., Tsukada, H., and Isozaki, H., "Incorporating Speech Recognition Confidence into Discriminative Named Entity Recognition of Speech Data," In *proceedings of the ACL*, 2006.

[14] Horlock, J. and King, S., "Named Entity Extraction from Word Lattices," In *proceedings of the Eurospeech*, 2003.

[15] Favre, B., Béchet, F., and Nocéra, P., "Robust Named Entity Extraction from Large Spoken Archives," In *proceedings of the EMNLP*, 2005.

[16] Siu, M. H., Vessenes, T., Bulyko, I., and Kimball, O., "Improved Named Entity Extraction from Conversational Speech with Language Model Adaptation," In *proceedings of the IEEE SLT Workshop*, 2010.

[17] Paass, G., Pilz, A., and Schwenninger, J., "Named Entity Recognition of Spoken Documents Using Subword Units," In *proceedings of the Intl. Conf. on Semantic Computing*, Berkely, CA, 2009.

[18] Kurata, G., Itoh, N., Nishimura, M., Sethy, A., and Ramabhadran, B., "Named Entity Recognition from Conversational Telephone Speech Leveraging Word Confusion Networks for Training and Recognition," In *proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, 2011.

[19] Parada, C., Dredze, M., and Jelinek, F., "OOV Sensitive Named Entity Detection in Speech," In *proceedings of the International Speech Communication Association (INTERSPEECH)*, 2011.

[20] Kumar, R., Prasad, R., Ananthakrishnan, S., Vembu, A. N., Stallard, D., Tsakalidis, S., and Natarajan, P., "Detecting OOV Named Entities in Conversational Speech," In *proceedings of the International Speech Communication Association (INTERSPEECH)*, 2012.

[21] Lafferty, J., McCallum, A., and Pereira, F., "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data " in *ICML*, 2001, pp. 282-289.

[22] Nguyen, L. and Schwartz, R., "Efficient 2-pass N-best Decoder," In *proceedings of the Eurospeech*, Rhodes, Greece, 1997.

[23] Stallard, D., Prasad, R., Natarajan, P., Choi, F., Saleem, S., Meermeier, R., Krstovski, K., Ananthakrishnan, S., and Devlin, J., "The BBN TransTalk Speech-to-Speech Translation System," *Speech and Language Technologies, InTech,* pp. 31-52, 2011.

[24] Kudo, T. Available: http://crfpp.googlecode.com/svn/trunk/doc/index.html

[25] Toutanova, K., Klein, D., Manning, C., and Singer, Y., "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," In *proceedings of the Human Language Technologies: The 4th Annual Conference of the North American Chapter of the Association for Computational Linguistics* 2003.

[26] Bisani, M. and Ney, H., "Investigations on Joint-Multigram Models for Grapheme-to-Phoneme Conversion," In *proceedings of the Int. Conf. on Spoken Language Processing*, Denver, CO, USA, 2002.

[27] Koehn, P., Och, F. J., and Marcu, D., "Statistical Phrase-Based Translation," In *proceedings of the NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Edmonton, Canada, 2003.

[28] Chen, W., Ananthakrishnan, S., Kumar, R., Prasad, R., and Natarajan, P., "ASR Error Detection in a Conversational Spoken Language Translation System," In *proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, 2013.