

LANGUAGE STYLE AND DOMAIN ADAPTATION FOR CROSS-LANGUAGE SLU PORTING

Evgeny A. Stepanov, Ilya Kashkarev, Ali Orkan Bayer, Giuseppe Riccardi, Arindam Ghosh

Signals and Interactive Systems Lab
Department of Information Engineering and Computer Science
University of Trento, Italy

ABSTRACT

Automatic cross-language Spoken Language Understanding porting is plagued by two limitations. First, SLU are usually trained on limited domain corpora. Second, language pair resources (e.g. aligned corpora) are scarce or unmatched in style (e.g. news vs. conversation). We present experiments on automatic style adaptation of the input for the translation systems and their output for SLU. We approach the problem of scarce aligned data by adapting the available parallel data to the target domain using limited in-domain and larger web crawled close-to-domain corpora. SLU performance is optimized by re-ranking its output with Recurrent Neural Network-based joint language model. We evaluate end-to-end SLU porting on close and distant language pairs: Spanish - Italian and Turkish - Italian; and achieve significant improvements both in translation quality and SLU performance.

Index Terms— Spoken Language Understanding, Statistical Machine Translation, Domain Adaptation

1. INTRODUCTION

With respect to the direction and the object of translation, the approaches to Spoken Language Understanding (SLU) porting via Statistical Machine Translation (SMT) can be grouped under two categories: Test-on-Source and Test-on-Target. In the *Test-on-Source* approach the direction of translation is from a language the system is being ported to (target language) to the language of the existing SDS (source language). The object of translation is user utterances in the target language. Consequently, SLU of the existing system is “extended” via SMT to cover a new language, and the success depends on the quality of machine translation. In the *Test-on-Target* approach (also referred to as Train-on-Target) the direction of translation is the opposite, i.e. from the source language to the target language. The object of translation is the data used to train the source SLU, and new language understanding components are trained. Thus, the success also relies on the accurate transfer of annotation.

This research is partially funded by the FP7 PortDial project n. 296170.

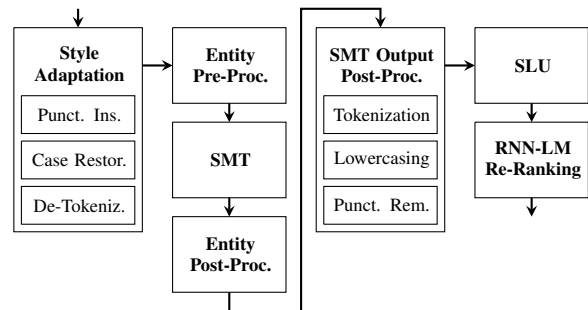


Fig. 1. Test-on-Source Spoken Language Understanding pipeline based on Statistical Machine Translation.

In the literature, the Test-on-Source approach is credited as having better performance (e.g. [1, 2, 3, 4]). Moreover, the procedure is simpler to implement, since it does not require porting of annotation. Additional techniques such as statistical post-editing and ‘smeared’ SLU training proposed by [3] make this approach even more appealing.

Both approaches to SLU porting are plagued by two limitations: (1) SLU are usually trained on limited domain corpora, and (2) parallel corpora are scarce. Majority of the research on SLU porting make use of in-domain parallel corpora (usually manually translated) to train SMT systems, and experiment on close or resource-rich language pairs. This paper, on the other hand, presents Test-on-Source SLU porting via SMT using off-the-shelf general-domain system and a system trained on out-of-domain data. We present experiments on language style adaptation for off-the-shelf SMT systems and domain adaptation for the SMT systems trained on out-of-domain data. The corpora used for domain adaptation are *in-domain* corpus used to train the source language SLU, and *close-to-domain* web crawled corpus. Both language style and domain adaptation take place in the SMT pipeline. The semantic parses of the translation hypotheses are further re-ranked with in-domain Recurrent Neural Network-based joint language model [5] in the source language. Fig. 1 presents the overall architecture of the process.

We evaluate end-to-end SLU system porting on both close and distant language pairs: Spanish - Italian and Turkish - Italian; and achieve significant improvements both in translation quality and SLU performance.

The paper is structured as follows: we first describe corpora and SMT systems used throughout the paper in Section 2. Then we present language style adaptation for translation of speech transcriptions (Section 3) and domain adaptation for SMT trained on out-of-domain corpora (Section 4). In Section 5 we describe our SLU model and RNN-LM based re-ranking. Section 6 provides concluding remarks.

2. CORPORA AND SMT SYSTEMS

In this section we briefly describe the corpora used to train and adapt SMT systems, as well as SMT systems themselves. For the close-to-domain LUNA Web corpus, we additionally describe the crawling and cleaning processes. Finally, we report SMT performances of these baseline systems.

2.1. Corpora

The **in-domain** corpus used throughout the paper is *Italian LUNA Corpus* [6]. The corpus is a collection of 723 human-machine dialogs (approximately 4K user turns) in the hardware/software help desk domain, annotated at several levels including concept attribute-value annotation that is necessary to train SLU models. *Multilingual LUNA Corpus* is the translation of Italian LUNA Corpus to the target languages Spanish and Turkish via professional translation services. The translated corpus consists of text only (i.e. annotations have not been transferred). The corpus is used to train in-domain SMT systems: Spanish - Italian and Turkish - Italian. The translations of development and test sets of the corpus, in both Spanish and Turkish, are used as an input for the SMT systems, and the evaluation is done against Italian reference.

The **close-to-domain** *LUNA Web Corpus* was crawled from the web. Starting from the original Italian LUNA corpus, rules for query construction and sentence selection were manually created. The first 100 query results returned by Google Search were downloaded. The downloaded documents were automatically sentence split, and the same hand-crafted rules were used to extract sentences close to the LUNA domain. The resulting set of 80K crawled sentences was semi-automatically cleaned for encoding and spelling issues to result in a 50K sentence close-to-domain corpus.

The **out-of-domain** *Europarl Parallel Corpus* [7] of the proceedings of the European Parliament is the most popular corpus in machine translation community. It encompasses 21 European languages, including the languages of interest: Spanish and Italian. Version 7 (from May 2012) of the corpus was used to create Italian-Spanish parallel corpus of approximately 1.8M sentence pairs. This parallel corpus is used to train the out-of-domain Spanish-Italian SMT system.

The Spanish sentences of Europarl are additionally used to train language models for language style adaptation experiments. For Turkish, the corpus used for language style adaptation is *Turkish Wikipedia* dump. The text was extracted and

Table 1. SMT System Baselines: *Google Translate*: General-domain Off-the-Shelf SMT; *Europarl Moses*: Out-of-domain SMT; *LUNA Moses*: In-domain SMT. Performance on LUNA Development Set for Spanish-Italian (ES-IT) and Turkish-Italian (TR-IT) is reported as 4-gram BLEU score.

SMT System	Language Pair	
	ES-IT	TR-IT
Google Translate	25.89	13.72
Europarl Moses	35.08	N/A
LUNA Moses	49.77	33.39

sentence split to result in approximately 3M sentences.

2.2. Baseline SMT Systems

Google Translate is a **general-domain** SMT system designed to provide reliable translations of text in multiple genres. It is trained on a vast variety of parallel written texts (as opposed to speech transcriptions). Since it is targeted for a wide range of languages, the translations go through English as a bridge language, i.e. a sentence in Turkish or Spanish is first translated into English and then to Italian.

Europarl Moses is an **out-of-domain** data trained mooses-based SMT system. Moses¹ is a statistical machine translation system that, given a parallel corpus, allows training translation models for any language pair automatically. The tool supports various translation models: phrase-based and tree-based, as well as factored models; and input of different level of complexity from text to ASR lattices. Here we use a phrase-based translation model on plain text. Prior to the training, Europarl corpus was pre-processed to be suitable for speech transcriptions: it was tokenized, lowercased and all punctuation was removed.

LUNA Moses is an **in-domain** data trained SMT system. Multilingual LUNA Corpus was used to train both Spanish - Italian and Turkish - Italian systems. These systems represent an upper-bound performance.

The performance of the three baseline SMT systems for Spanish - Italian and Turkish - Italian language pairs is reported in Table 1 using 4-gram BLEU score [8]. Since Europarl does not have Turkish, there is no out-of-domain SMT system for Turkish - Italian. As expected, in-domain SMT systems perform the best for both language pairs, followed by the out-of-domain SMT system for Spanish - Italian, since the training corpus was already pre-processed for speech transcriptions. Google Translate has the worst performance.

3. LANGUAGE STYLE ADAPTATION

Using off-the-shelf SMT systems like Google Translate for SLU porting has both advantages and disadvantages. The ad-

¹<http://www.statmt.org>

vantages are that the SDS developers do not require expertise in machine translation, and can obtain satisfactory translations for a wide range of language pairs without the need for parallel corpora. The disadvantages, on the other hand, are that these SMT systems are general domain, and are trained on written text, which differs in style from the spoken conversation transcriptions SLU is trained on.

In this section we describe and evaluate our approach to the problem of unmatched style (see Fig. 1). The description is organized from the SLU perspective. First, the output of SMT system is matched the conversation transcription style in the source language. Second, the input (ASR output or transcription in target language) is matched the SMT training data style, i.e. written text. Since in conversation transcription style, unlike written text, symbols and numbers are always spelled out; we apply an additional step of entity pre- and post-processing aimed at reducing the noise added by SMT.

3.1. SMT Output Post-Processing

The differences between conversation transcriptions (data for training SLU systems) and written text is that, the former has no sentence boundaries, no capitalization, no punctuation, and it is tokenized. Thus, SMT output post-processing step consists of tokenization, lowercasing, and removing all punctuation except single quotes used in contractions. All the techniques are widely applied in various NLP tasks.

The same process is applied to Europarl prior to training the SMT system (Europarl Moses baseline) to bring it closer to the conversational style.

3.2. Language Style Adaptation

The process of adapting conversation transcription style to the written text style is the direct opposite of the SMT output post-processing. Thus, the steps are: automatic punctuation insertion, automatic case restoration, and de-tokenization. De-tokenization (attaching punctuation marks and contractions to the respective tokens) is handled by language dependent rules. Case Restoration and Punctuation Insertion, on the other hand, require training of statistical models; thus, these two steps are described in more detail.

Automatic Punctuation Insertion requires language models (LM) trained on a lowercased and tokenized corpus that contains punctuation. The language model we use is a trigram back-off language model with modified Kneser-Ney discounting, trained on Spanish section of Europarl for Spanish-Italian translation and on Wikipedia for Turkish-Italian. Prior to training the LMs, both corpora were sentence split, and the beginning and end of sentence tags were inserted. To reduce noise we restrict the punctuation lexicon to a few most frequent punctuation marks: ‘.’, ‘,’, ‘?’, ‘;’. The *hidden-ngram* tool from SRILM toolkit [9], which tags a sequence of tokens with hidden events occurring between them, is used for recovering the missing punctuation.

Automatic Case Restoration requires cased corpora to train language and case models. For training these models, we use the same corpora as for automatic punctuation insertion, with the exception that the corpora remain cased. We apply *Moses recaser*, provided with the Moses translation system. The tool trains a restricted translation model to translate from lowercased to cased text. Additionally, it applies sentence initial capitalization.

3.3. Entity Processing for SMT

Dialogs, like any data, often contain named entities, dates, numerical expressions, etc. Moreover, all these entities are spelled out. In a live dialog system these entities are usually handled by their associated grammars, either handcrafted by the developers or provided as built-in by the ASR system. In the context of automatic translation of transcriptions or ASR output, translating such entities adds additional noise; thus, handling these entities with a grammar in the target language is a better option. The step is beneficial for both off-the-shelf and Moses-based systems trained on out-of-domain corpora.

In the domain of LUNA corpus (IT Help Desk), one of the most frequent entities is numerical expressions: ticket numbers, phone numbers, etc.. Google Translate often converts word-numerical expressions into dates or reorders them. To reduce this translation noise the following procedure is implemented: (1) Each word-numerical expression in the target language (Spanish or Turkish) is converted to digits, i.e. “two thousand six” is converted to “2006”. (2) In the source language (Italian) these expressions are converted back to word-numerical form, i.e. “2006” is converted to “two thousand six”. The digit-form expressions are enclosed in XML tags to prevent their translation, a feature supported both by Google Translate and Moses.

3.4. Results and Discussion

Table 2 reports cumulative results of SMT output post-processing, SMT language style adaptation, and numerical entity pre- and post-processing steps for Google Translate on the LUNA development set. For output post-processing and language adaptation all the described steps are applied, since they improve BLEU score individually and in combination. Output post-processing improves the performance by almost 4 point for Spanish and 3 points for Turkish. The effect of style adaptation is greater for Spanish (2.88) than for Turkish (1.37), what is easily explained by the rich Turkish morphology; thus, greater data sparseness. Due to the high frequency of word-numerical entities in LUNA, numerical entity processing step improves the performance by additional 2.44 for Spanish and 4.31 points for Turkish. Considering all the pre- and post-processing steps, results indicate that off-the-shelf SMT systems like Google Translate can be adapted to the spoken utterance translation, and matching input and output language styles greatly improves performance irrespective of

Table 2. Cumulative effects of output post-processing, style adaptation and numerical entity processing for Google Translate on LUNA Development Set. Results are reported as 4-gram BLEU score.

<i>Pre- & Post- Processing</i>	<i>Language Pair</i>	
	ES-IT	TR-IT
Baseline	25.89	13.72
+ Post-Processing	29.78	16.47
+ Style Adaptation	32.66	17.84
+ Numerical Entities	35.10	22.15

Table 3. Performance of the style-adapted off-the-shelf SMT Google Translate, out-of-domain Europarl Moses, and in-domain LUNA Moses SMT systems on LUNA Development and Test Sets. Results are reported as 4-gram BLEU score.

<i>SMT System</i>	<i>Language Pair</i>			
	ES-IT		TR-IT	
	DEV	TEST	DEV	TEST
Google Translate	35.10	31.08	22.15	20.13
Europarl Moses	37.37	35.69	N/A	N/A
LUNA Moses	49.77	50.69	33.39	35.29

language distance (by 9.21 points for Spanish and 8.43 for Turkish).

Table 3 reports performance of the style-adapted SMT systems on LUNA Development and Test Sets. Affected systems are Google Translate and Europarl Moses. While the former includes the full pre- and post-processing, the latter includes only entity processing, since Europarl is adapted to conversational style prior to training. Comparing Tables 1 and 3, we reduce the performance difference between off-the-shelf and out-of-domain Spanish - Italian SMT from 9.19 to 2.27. However, both systems still perform more than 10 points below the in-domain SMT system.

4. DOMAIN ADAPTATION FOR SMT

SMT systems are drastically affected by differences in training and testing conditions. One of the drawbacks of using an off-the-shelf translation systems is not being able to access its translation and language models. Thus, any available in-domain data is not utilized. An alternative is to train the system using an open source tools such as Moses on out-of-domain parallel corpora like Europarl, and adapt it to the target domain. Thus, in this section we address the second limitation of cross-language SLU porting – scarce aligned data.

Domain Adaptation is a rather well studied topic in machine translation research, and a variety of methods were proposed (see [10] for review). Phrased-based SMT tools, like Moses, generally require two models for translation: a translation model (phrase table) and a language model. With respect to the availability of bilingual in-domain data either of these

Table 4. Effects of domain adaptation with in-domain and close-to-domain language models for Europarl Moses Spanish-Italian SMT on LUNA Development and Test Sets. Results are reported as 4-gram BLEU score.

Transl. Model	Lang. Model	DEV	TEST
LUNA	LUNA	49.77	50.69
LUNA	LUNA+Web LUNA	51.34	50.77
Europarl	Europarl	37.37	35.69
Europarl	LUNA	48.11	44.65
	Web LUNA	46.58	40.82
	LUNA+Web LUNA	49.36	45.60
Europarl+LUNA	Europarl+LUNA	47.57	46.87
Europarl+LUNA	Europarl+LUNA +Web LUNA	49.66	48.95

models is adapted to the target domain. Simple SMT domain adaptation techniques are presented in [11]: (1) pooling large out-of-domain and small in-domain parallel corpora together to train the models, (2) using out-of-domain corpus for the translation model and in-domain data for the language model, (3) and their combinations. We follow the same approach, but additionally augment the data for training the language models with close-to-domain web crawled data, i.e. LUNA Web Corpus. Thus, for Europarl Moses system, we substitute Europarl trained out-of-domain language model with a language model trained on (1) Italian LUNA corpus – in-domain data, (2) LUNA Web corpus – close-to-domain data, or (3) both corpora. In all cases monolingual target language data is used. For the sake of completeness, we also present results on pooled data training and in-domain SMT with web data augmented language model. Since Europarl is not available for Turkish, all adaptation experiments are for Spanish - Italian.

Table 4 reports results on domain adaptation. The first observation is that augmenting the in-domain language model with close-to-domain web crawled improves the already high performance of the LUNA Moses by 1.57 for the development and 0.08 for the test set. Using in-domain language model to re-score translation hypotheses of the out-of-domain Europarl translation model improves performance by more than 10 points. Even though the gain of using close-to-domain language model is less, the performance is still more than 8 points higher than of the out-of-domain SMT. Training the language model on both in-domain and close-to-domain corpora outperforms both and falls only 0.41 points less than the in-domain SMT system for the development set; however, the difference increases to 5 points on the test set.

Pooling Europarl and LUNA corpora to train both translation and language models yields performance more than 10 points higher than the out-of-domain system. Augmenting the pooled data with close-to-domain data increases performance by additional 2 points, very close to the in-domain SMT.

The domain adaptation experiments show that adapting out-of-domain data trained SMT systems with monolingual

in-domain and close-to-domain data yields performance close to the in-domain SMT; thus, the translation of the source language corpora to build in-domain SMT for Test-on-Source SLU might not be necessary. Augmenting the limited in-domain data with larger web-crawled close-to-domain data is definitely beneficial: the Out-Of-Vocabulary rate (OOV) for LUNA corpus drop from 4.30% to 1.27% with the addition of close-to-domain data to the training set; consequently, better performance is expected.

5. TEST-ON-SOURCE SLU

In the Test-on-Source approach there is already an SLU model in the source language and SMT is deployed to translate the target language utterances to the source language. For the two target languages, Spanish and Turkish, utterances are translated to Italian, using the SMT systems described in the sections above. The translated utterances are the input to the SLU for semantic parsing (extraction of domain concepts).

A commonly accepted metric for SLU evaluation is Concept Error Rate (CER), which is based on the Levenshtein alignment of sentences and computed as the ratio between inserted, deleted and substituted concepts and the total number of concepts in the reference sentence. Since the SMT systems are optimized for BLEU during training and the target evaluation metric is CER, the behavior of systems might change.

The problem of optimizing the SMT directly for semantic parsing was addressed by tuning the mooses-based SMT (setting the model weights via Minimum Error Training) in [2]. The authors showed that such tuning reduces the CER. In this paper we follow a different approach exploiting the fact that Google Translate and Moses can output several translation hypotheses (n-best list). These hypotheses are parsed by the SLU and then re-ranked using in-domain RNN-based joint LM [5] trained on reference transcription word-concept pairs.

First, we briefly describe our SLU and RNN-based joint LM re-ranking, and then present the results on re-ranking of the style adapted and domain adapted SMT systems.

5.1. Spoken Language Understanding Module

One of the recent approaches to SLU is based on conditional random fields (CRF) [12]. CRFs are discriminative undirected graphical models which have been successfully used for segmenting and labeling sequential data. CRFs model the conditional probability of the concept sequence given the word sequence. In the SLU module we have used different types of features. The first type of features is orthographic, aimed at decreasing the data sparseness problem. These features consider the first or last i characters of the word, where i changes from 1 to 5. Another type of features is word bigrams that consist of pairs of the previous and the current word, the current and the next word, and also, the previous and the next word. In addition to these features we have bi-

Table 5. Test-On-Source SLU performance of SMT systems on the LUNA Test Set. 1-Best SLU CER for the baseline and style-adapted systems, 100-Best RNN-LM re-ranked CER, and 100-Best oracle CER (in parentheses) are reported.

<i>SMT System</i>	BL	SA	RNN-LM
Spanish - Italian			
Google Translate	43.00	36.10	34.60 (31.10)
Europarl Moses	39.20	35.40	31.30 (22.80)
LUNA Moses	25.80	N/A	25.30 (20.70)
Turkish - Italian			
Google Translate	56.90	50.40	49.20 (44.70)
LUNA Moses	39.20	N/A	37.90 (27.70)

nary features to label numeric words. CRFs also consider the previous token’s output label as a feature for the model. All these features are independent of each other in the window of ± 1 tokens. The source language SLU performance on the LUNA Test Set has CER of 21.50.

5.2. RNN-based Joint Language Model Re-Ranking

Considering word-to-concept alignment constrains to optimize language models (LMs) improves SLU performance [13]. A Neural Network (NN) LM to optimize the SLU performance, which is a joint model that is built over word-concept pairs, was proposed in [5]. The given LM is based on a recurrent NN (RNN) that uses a modified version of the class-based RNN structure given in [14]. This RNN-based joint LM is used to re-rank the n-best list of semantic parses of the translation hypotheses. Translation scores of the SMT systems are combined with the scores of the RNN-based joint LM. Specifically, translation and LM scores provided by the SMT (mooses) are extracted and the LM score is substituted with the RNN-based joint LM probability. In case of Google Translate there is no separate language model score; thus, the re-ranking is solely RNN-LM score based.

5.3. Results and Discussion

Table 5 reports the SLU performance of the baseline and style-adapted systems (including all pre- and post-processing) and the in-domain LUNA Moses SMT in terms of CER; as well as the performance of 100 best RNN-LM re-ranking (oracles of 100-best are given in parentheses). The first observation is that the performance of the systems in terms of CER is in line with their performance in terms of BLEU, i.e. in-domain SMT perform the best and the off-the-shelf SMT the worst. This holds for the baseline, style-adapted and RNN-LM re-ranked systems. Style adaptation significantly improves performance of both Google Translate and Europarl Moses. The benefits of re-ranking is greater for out-of-domain SMT than for Google Translate. This is explained by the fact that Google Translate outputs only a few translation hypotheses (on average 4.5 hypotheses per sentence),

Table 6. Test-On-Source SLU performance of the domain-adapted Spanish - Italian Moses SMT systems on the LUNA Test Set. 1-Best SLU CER, 100-Best RNN-LM re-ranked CER, and 100-Best oracle CER (in parentheses) are reported.

Transl. Model	Lang. Model	SLU	RNN-LM
LUNA	LUNA	25.80	25.30 (20.70)
LUNA	LUNA+Web LUNA	26.00	26.00 (22.80)
Europarl	Europarl	35.40	31.30 (22.80)
Europarl	LUNA	31.20	29.80 (23.60)
	Web LUNA	32.70	31.30 (25.20)
	LUNA+Web LUNA	31.20	30.00 (24.50)
Europarl+LUNA	Europarl+LUNA	28.40	27.20 (23.10)
Europarl+LUNA	Europarl+LUNA +Web LUNA	27.90	26.30 (22.10)

while for the Moses-based systems we use 100 hypotheses. Performance improvements hold across language pairs.

SLU performance and the results of RNN-based joint LM re-ranking for domain-adapted Spanish - Italian Moses-based SMT are reported in Table 6. Even though, in-domain SMT augmented with web crawled data has higher BLEU score (see Table 4), it produces worse SLU results. Similarly, for the out-of-domain SMT trained on Europarl, augmenting the in-domain LM with web crawled data does not improve SLU performance. However, for the SMT with pooled-data training, adding web crawled data to the in-domain corpus, improves performance by 0.5.

The results of 100-best RNN-LM re-ranking are in-line with 1-best SLU results for the domain adapted systems: the only benefit of adding web crawled data is observed in pooled-data training condition. The benefit of re-ranking is proportional to the amount of out-of-domain data in the language models of SMT. Thus, Europarl Moses benefits the most, CER drops by 4.1%, and reaches the performance of the SMT adapted by web-crawled data only.

6. CONCLUSION

In this paper we proposed methods for dealing with the limitations of cross-language SLU porting such as scarceness of aligned data and unmatched style of conversation transcriptions and written text: style adaptation, domain adaptation, and semantic parse re-ranking with in-domain RNN-based LM. We evaluate end-to-end SLU system porting on both close and distant language pairs: Spanish - Italian and Turkish - Italian; and achieve significant improvements both in translation quality and SLU performance.

7. REFERENCES

[1] B. Jabaian, L. Besacier, and F. Lefèvre, “Investigating multiple approaches for SLU portability to a new language,” in *Proc. of INTERSPEECH*, 2010.

[2] B. Jabaian, L. Besacier, and F. Lefèvre, “Combination of stochastic understanding and machine translation systems for language portability of dialogue systems,” in *Proc. of ICASSP*, 2011.

[3] B. Jabaian, L. Besacier, and F. Lefèvre, “Comparison and combination of lightly supervised approaches for language portability of a spoken language understanding system,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 3, 2013.

[4] F. Lefèvre, F. Mairesse, and S. Young, “Cross-lingual spoken language understanding from unaligned data using discriminative classification models and machine translation,” in *Proc. of INTERSPEECH*, 2010.

[5] A.O. Bayer and G. Riccardi, “Joint language models for automatic speech recognition and understanding,” in *Proc. of SLT*, 2012.

[6] M. Dinarelli, S. Quarteroni, S. Tonelli, A. Moschitti, and G. Riccardi, “Annotating spoken dialogs: from speech segments to dialog acts and frame semantics,” in *Proc. of EACL Workshop on Semantic Representation of Spoken Language*, 2009.

[7] P. Koehn, “Europarl: A parallel corpus for statistical machine translation,” in *Machine Translation Summit X*, 2005.

[8] K. Papineni, S. Roukos, T. Ward, and W. Zhu, “BLEU: a method for automatic evaluation of machine translation,” in *Proc. of ACL*, 2002.

[9] A. Stolcke, “SRILM - an extensible language modeling toolkit,” in *Proc. of ICSLP*, 2002.

[10] N. Bertoldi and M. Federico, “Domain adaptation for statistical machine translation with monolingual resources,” in *Proc. of WMT*, 2009.

[11] P. Koehn and J. Schroeder, “Experiments in domain adaptation for statistical machine translation,” in *Proc. of WMT*, 2007.

[12] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *Proc. of ICML*, 2001.

[13] G. Riccardi and A.L. Gorin, “Stochastic language models for speech recognition and understanding,” in *Proc. of ICSLP*, 1998.

[14] T. Mikolov, S. Kombrink, L. Burget, J. Cernock, and S. Khudanpur, “Extensions of recurrent neural network language model,” in *Proc. of ICASSP*, 2011.