MULTI-STYLE ADAPTIVE TRAINING FOR ROBUST CROSS-LINGUAL SPOKEN LANGUAGE UNDERSTANDING

Xiaodong He, Li Deng, Dilek Hakkani-Tur, Gokhan Tur

{xiaohe,deng,dilekha,gokhant}@microsoft.com

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

ABSTRACT

Given the increasingly available machine translation (MT) services nowadays, one efficient strategy for cross-lingual spoken language understanding (SLU) is to first translate the input utterance from the second language into the primary language, and then call the primary language SLU system to decode the semantic knowledge. However, errors introduced in the MT process create a condition similar to the "mismatch" condition encountered in robust speech recognition. Such mismatch makes the performance of cross-lingual SLU far from acceptable. Motivated by successful solutions developed in robust speech recognition, we in this paper propose a multi-style adaptive training method to improve the robustness of the SLU system for cross-lingual SLU tasks. For evaluation, we created an English-Chinese bilingual ATIS database, and then carried out a series of experiments on that database to experimentally assess the proposed methods. Experimental results show that, without relying on any data in the second language, the proposed method significantly improves the performance on a cross-lingual SLU task while producing no degradation for input in the primary language. This greatly facilitates porting SLU to as many languages as there are MT systems without any human effort. We further study the robustness of this approach to another type of mismatch condition, caused by speech recognition errors, and demonstrate its success also.

Index Terms— spoken language understanding, cross-lingual, adaptive training, multi-style training

1. INTRODUCTION

Porting a spoken language understanding (SLU) service from one language to another is of increasing interests to the community recently [11][12][2][15][28][20]. It aims at porting the SLU service that is built for a primary language to a second language, where we usually have no or very limited data to train a full SLU system reliably.

While there is a wide body of work on cross-lingual question answering (QA) and information retrieval (IR), mainly due to the shared task evaluations at the text retrieval conferences (TREC) [6] and the cross language evaluation forum (CLEF) [13], the work on cross-lingual SLU is rather limited to a few studies. In line with the previous work on QA, Jabaian et al. [12] proposed two strategies for SLU portability. The first one includes translating annotated corpora to the second language and training models for understanding examples in the second language. The second one instead translates the examples in the second language to the primary language and uses the primary language SLU models to analyze them. They also showed that training with additional noisy data increases robustness of models and a combination of the three modeling approaches results in the best performance [11].

In this paper, we focus on the second strategy, in which the input utterance is first translated into the primary language, and then the primary language SLU system is called to decode the semantic knowledge. Given the machine translation (MT) services, such as Microsoft and Google Online Translator [18][7], that are broadly available nowadays, the SLU service for the primary language can be efficiently extended to cover a variety of other languages with minimum cost using this strategy.

However, due to errors introduced in the MT process, the performance of cross-lingual SLU is far from acceptable. Similar to the "mismatch" condition encountered in robust speech recognition [3][16], the SLU system in the primary language is usually trained on clean data, while the input is noisy data with translation errors. This training/testing condition mismatch causes severe performance degradation. Therefore, building SLU systems that are robust to translation errors is crucial for cross-lingual SLU.

In [11], a smeared training data approach is proposed to address this issue. Assuming there are training data available in the second language, these data are first translated into the primary language, and then merged with other training data in the primary language to train the SLU model. Performance improvement on a cross-lingual SLU task is reported. However, since extra data are used compared to the baseline (i.e., the extra training data translated from the second language), it is not clear how much the improvement is exactly from introducing MT distortion in training. Moreover, that approach depends on the availability of the training data in the second language. Therefore, it is costly and may be not suitable for languages that have no or very limited resource.

In this work, without relying on any training data in the second language, we propose a method to adapt the clean training data in the primary language to the MT-distorted condition, so as to mitigate the training/testing condition mismatch problem for cross-lingual SLU. In our method, we first translate the clean training data in the primary language into the second language and then translate them back, both through MT. By doing this, we intentionally inject translation errors into the original clean training data. We then train the SLU system on these MT-distorted data and therefore adapt the SLU model to be more robust to inputs that contain MT errors. Compared to [11], one advantage of this training approach is that it doesn't require any data in the other languages, so we can port SLU to as many languages as there are MT systems without any human effort.

In another related work, Misu et al. use MT to translate the training data from the primary language to the second language (i.e., the first strategy) and run a back-translation for the purpose of

data filtering [20]. On the other hand, combining multiple MT systems could lead to superior translation output [9][17], which could further improve the performance of cross-lingual SLU significantly as demonstrated in [5].

Compared to previous work (e.g. [11][12]), we have also studied three important issues regarding the robustness of the SLU systems. First, to minimize the cost and take the full advantage of cross-lingual SLU, there is a demand for building one SLU system in the primary language to serve all requests from both the primary language and other foreign languages (via MT). Therefore, it is crucial to build a system that is robust to cross-lingual users without sacrificing the performance for users in the primary language. In this work, we carried out quantitative study to investigate the impact of MT-distorted training data on the performances for both the cross-lingual test and clean test conditions. The second issue we studied is how, if any, the MT-distorted data helps the robustness of SLU models to other types of errors such as errors from automatic speech recognition (ASR). Moreover, for certain SLU tasks, there are extra knowledge resources available beside regular training data, such as named entity lookup tables, which provide valuable information for SLU. In this paper, we also studied the performance of the proposed approaches under this condition.

We evaluate the proposed methods on a Chinese-to-English cross-lingual slot filling task. For evaluation, we created an English-Chinese bilingual ATIS database which is constructed by manually translating the English ATIS database into Chinese, including both the sentences and the annotations. Then we carried out a series of experiments to evaluate the proposed methods on this database. As the experimental results show, our method significantly improves the cross-lingual SLU accuracy on the slot-filling task by up to 5% absolutely as measured in F1 measure, while the performance on the clean input in the primary language is kept without degradation. We further study the robustness of our approach to another type of mismatch condition caused by speech recognition errors, and observe significant improvements also.

2. MULTI-STYLE ADAPTIVE TRAINING FOR CROSS-LINGUAL SLU

2.1 Spoken language processing: A review

Semantic parsing of input utterances typically consists of 3 tasks, domain detection, intent determination, and slot filling. Originated from call routing systems, domain detection or intent determination tasks are typically treated as semantic utterance classification, and originated from natural language to semantic template filling systems such as the DARPA ATIS, the slot filling task is typically treated as sequence classification. An example sentence with slot annotations is provided in Fig. 2 (English).

In this work, we evaluate our methods on the slot filling task. The input is the sentence consisting of a sequence of words, and the output is a sequence of slot IDs, one for each word, tagged by a conditional random field (CRF) [14] based slot filling model, i.e., given the input word sequence $L_1^N = l_1, \dots, l_N$, the linear chain CRF models the probability of a slot sequence $S_1^N = s_1, \dots, s_N$ as follows:

$$p(S_1^N|L_1^N) = \frac{1}{Z}\prod_{t=1}^{N} H(s_{t-1}, s_t, l_{t-d}^{t+d})$$

where

$$H(s_{t-1}, s_t, l_{t-d}^{t+d}) = \sum_{m=1}^{M} \lambda_m h_m(s_{t-1}, s_t, l_{t-d}^{t+d})$$

and $h_m(s_{t-1}, s_t, l_{t-d}^{t+d})$ are features extracted from the current and the previous states s_t and s_{t-1} , plus a window of words around the current word l_t, with a window size of $2d + 1$.

2.2 Multi-style adaptive training

In cross-lingual SLU, we can first translate the input utterance from the second language to the primary language, and then call the primary language SLU system to decode the semantic knowledge. One benefit of this approach is that we only need to build and maintain one SLU system in the primary language. Given the machine translation (MT) services, such as Microsoft and Google Translators, widely available nowadays, this approach is particularly plausible when we need to port the SLU capability to a variety of new languages quickly with minimum cost.

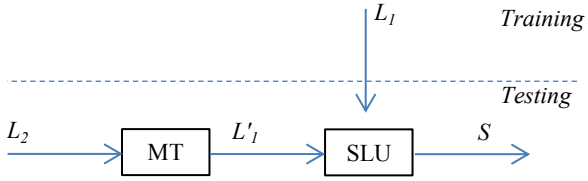
We illustrate this cross-lingual SLU strategy in Fig. 1. In the figure, we denote by L_1 and L_2 as the sentences in the primary and the second language, respectively, and denote by L'_1 and L'_2 as the sentences in the primary language and the second language that are translated from the other language, respectively. S is the semantic slot sequences produced by the SLU system.

One commonly referred setup of this cross-lingual SLU strategy is illustrated in Fig. 1(a) [12]. In training, the SLU system is trained on the data in the primary language. In testing, the utterance spoken in the second language is first translated into the primary language via MT, and then the translated utterance is fed into the primary SLU system to produce the semantic slots. However, in this setup, the SLU system is trained on MT-distortion free data while the input is MT-distorted data. This training/testing condition mismatch causes severe performance degradation.

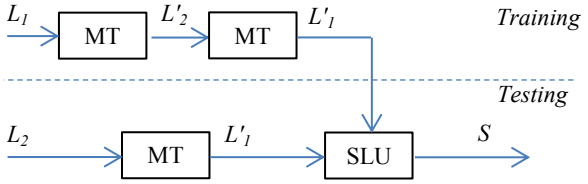
In order to mitigate the training/testing condition mismatch problem, we propose to adapt the distortion free training data in the primary language to the MT-distorted condition. Hereafter, we refer to the distortion free condition as *clean* condition and the MT-distorted condition as *noisy* condition for short. In our method, we first translate the clean training data in the primary language into the second language and then translate them back, both through MT. By doing this, we intentionally inject translation noise into the original clean training data. We then train the SLU system on these MT-distorted (noisy) data and therefore adapt the SLU model to be more robust to inputs that contain MT errors. In implementation, to recover annotations for the noisy training set, we performed a word alignment between the original clean data set and the newly-generated noisy data set, and then mapped the word level annotation from the former to the latter accordingly. Note that both of the clean and noisy sets are in the same primary language, therefore it is a monolingual word alignment is sufficient [9]. In this paper, we call this approach *adaptive training*, as in Fig. 1 (b).

However, training the SLU system solely on the noisy training set, though matching the cross-lingual condition well, may lead to performance degradation for clean input in the primary language due to a newly introduced mismatch between the *noisy* condition model and the *clean* input. To address this issue, we further proposed a multi-style adaptive training approach. As illustrated in Fig.1 (c), the SLU system will be trained on a training set consisting of multiple styles of data, including both clean and MT-distorted data. As will be demonstrated in the evaluation, this leads to a SLU system that is robust to various types of input. As shown in the experimental results, this method improves the performance

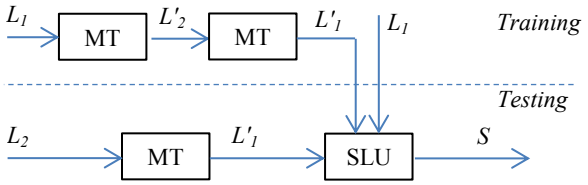
of cross-lingual SLU significantly, without performance degradation for clean input in the primary language.



(a) Clean training for cross-lingual SLU



(b) Adaptive training for cross-lingual SLU



(c) Multi-style adaptive training for cross-lingual SLU

Figure 1. Different training approaches for cross-lingual SLU.

3. ENGLISH-CHINESE BILINGUAL ATIS DATASET

Spoken language understanding (SLU) in goal oriented conversational systems aim to automatically identify the intent of the user as expressed in natural language and extract associated arguments (slots) [27]. This term has been mainly coined in the early 90s, by the DARPA (Defense Advanced Research Program Agency) Airline Travel Information System (ATIS) project. The ATIS task consisted of spoken queries on flight-related information. Understanding in this project was reduced to the problem of extracting task-specific arguments, such as *Destination* and *Departure Date*. An example utterance with annotations is shown in Fig. 2 (English). Participating systems employed either a data-driven statistical approach [19][22] or a knowledge-based approach [4][25][28]. An important by-product of the DARPA ATIS project was the ATIS corpus [23]. This corpus is the most commonly used data set for SLU research. In this paper, we use the ATIS corpus as used in He and Young [10], Raymond and Riccardi [24], and Tur et al. [26]. The training set contains 4,978 utterances selected from the Class A (context independent) training data in the ATIS-2 and ATIS-3 corpora, while the test set contains 893 utterances from the ATIS-3 Nov93 and Dec94 datasets. Each utterance has its named entities marked via table lookup, including domain specific entities such as city, airline, airport, and dates.

In order to facilitate the research of cross-lingual SLU, we manually translate the whole ATIS database from English to Chinese. In ATIS, usually the same intention is expressed in multiple different ways to cover variation of valid expressions. For the same purpose, the human translator is instructed to translate the sentence as literally as possible, without sacrificing fluency and adequacy in the translation. The translator is also instructed to map the slot annotations of English words to corresponding Chinese words based on his best knowledge. One example of this English-Chinese bilingual SLU database is illustrated in Fig. 2.

English:

Sentence	*show*	*flights*	*from*	*Boston*	*to*	*New*	*York*	*today*
Slots	O	O	O	B-dept	O	B-arr	I-arr	B-date

Chinese:

Sentence	显示	今天	从	波士顿	飞往	纽约	的	航班
Slots	O	B-date	O	B-dept	O	B-arr	O	O

Figure 2. Example of English-Chinese bilingual ATIS data set.

4. EVALUATION

4.1 Experimental condition

In this section we carried on Chinese-to-English cross-lingual SLU experiments to evaluate the proposed methods, i.e., the primary language is English and the second language is Chinese. The database used in the evaluation is described in section 3. We use the linear chain CRF model as described in section 2.1 for slot tagging. Following [24], the CRF++ toolkit is used. 5-fold cross validation on the training set is employed to pick the best regularization factor for CRF training. The machine translation service used in the experiments is the online Microsoft Translator. It is a large scale general purpose MT system that provides state-of-the-art translation service for about 80 language pairs. The performance of that MT system on the data set used in the experiments is presented in Table 1 in BLEU score [21]. It is interesting to see that the BLEU score of the MT-distorted training data, i.e., the training data after a round-trip MT, is higher than the cross-lingual test data which only goes through the MT process once. We find that because the same MT training data sets are usually used to train the translation models of two symmetric language pairs (e.g., English-to-Chinese and Chinese-to-English), it is possible that the translation error caused by mapping a source phrase to a wrong target phrase could lead to a correct source phrase when translating it back by phrase pairs learnt from the same translation samples. Nevertheless, substantial MT distortion is introduced as indicated by the lower-than-50% BLEU score (naturally, the clean data set has a BLEU score of 100% when using itself as reference).

The experimental results of the proposed methods are presented in the following sub-sections. Unless specified, the baseline model in these experiments is trained on the clean training data as illustrated in Fig. 1(a).

Table 1: BLEU scores of the MT-distorted training data and the Chinese-to-English cross-lingual test data, References are the clean training and test data in English, respectively.

Data set	BLEU
MT-distorted training data	42.6%
Chinese-to-English test data	33.4%

4.2 Experimental results

4.2.1 Experiments on adaptive training

We first evaluate the performance of adaptive training as illustrated in Fig. 1(b). In the experiment, the lexical/n-gram features extracted from a 5-word window are used for CRF. Experimental results are shown in Table 2. After trained on the MT-distorted data, the model is adapted to the cross-lingual condition well. This leads to a 5.1% improvement in F1 score for the Chinese-to-English cross-lingual test set. However, this approach also causes severe training/testing mismatch for the clean test condition, where the performance drops by 5.1%.

Table 2: Slot-filling results on the Chinese-to-English (chs-to-enu) cross-lingual test set and the clean test set using adaptive training, reported in F1 score.

Training set	chs-to-enu input	clean enu input
Clean data only (baseline)	68.37%	92.57%
MT-distorted data only	73.46% (+5.1%)	87.50% (-5.1%)

4.2.2 Experiments on multi-style adaptive training

We then evaluated the multi-style adaptive training method, with a focus on improving the performance on cross-lingual test while keeping the clean test performance from degradation. As illustrated in Fig. 2(c), starting from using solely the clean training data, we gradually add MT-distorted data for training. As shown in Table 3, the performance on the Chinese-to-English cross-lingual test set improves continuously when more and more MT-distorted data are added in, while the high accuracy on the clean test set maintains with almost no degradation. At the setting when all MT-distorted data are added, we observed a 5.3% improvement in F1 score, and only negligible degradation, -0.1%, on the clean test set.

Table 3: Slot-filling results on the Chinese-to-English (chs-to-enu) cross-lingual test and the clean test set using multi-style adaptive training under different conditions, reported in F1 score.

Training set	chs-to-enu input	clean enu input
Clean data only (baseline)	68.37%	92.57%
Clean+ 10% MT-distorted	70.76% (+2.4%)	92.54% (0.0%)
Clean+ 25% MT-distorted	71.46% (+3.1%)	92.45% (-0.1%)
Clean+ 50% MT-distorted	72.67% (+4.3%)	92.04% (-0.5%)
Clean+100% MT-distorted	73.69% (+5.3%)	92.50% (-0.1%)

4.2.3 Experiments on using extra Named Entity features

Table 4: Slot-filling results on the Chinese-to-English (chs-to-enu) cross-lingual SLU test and the clean English test set using extra NE features, reported in F1 score.

Training set	chs-to-enu input	clean enu input
Clean data only (baseline)	76.88%	94.44%
MT-distorted data only	79.26% (+2.4%)	91.23% (-3.2%)
Clean+100% MT-distorted	79.78% (+2.9%)	94.00% (-0.4%)

ATIS provides labels marked for named entities (NE) via NE table lookup (see section 3), which provide valuable information for SLU. In this section, we conducted experiments using additional features derived from these NE labels. The results are shown in Table 4. Compared the first row of Table 4 (baseline) to that of Table 3, adding the named entity features improves the slot filling accuracy substantially for both cross-lingual and clean tests. On top of this strong baseline, still, the proposed multi-style adaptive training further improves the slot filling accuracy on the Chinese-to-English cross-lingual test by 2.9% absolute, with a small -0.4% degradation on the clean-input test. This demonstrates that the gain of the proposed method is complementary to gains from using NE features.

4.2.4 Experiments on input with ASR errors

We also evaluated the robustness of the MT distortion-adapted model on non-MT type errors. In this experiment, the input for SLU is the recognition hypothesis from a generic dictation ASR system and has a word error rate (WER) of 13.8%, while this is significantly higher than the best reported performances of about 5% WER [29], this provides a more challenging and realistic framework. The experimental results are in Table 5. Interestingly, using solely the MT-distorted data for training does not help or hurt the performance on ASR-noise input. This is because the MT errors are different from the ASR errors, so the MT error patterns learned by the model may help some cases but hurt some others and the overall gain is neutral. However, after combined with the clean data in multi-style training, a 2.0% absolute gain in F1 score can be observed. The gain is smaller than that on cross-lingual tests, but still significant. These results indicate that information learned from MT-type of noise could improve the robustness of the model to other types of noise, too, but in a less effective degree.

Table 5: Slot-filling results reported in F1 score. The test input is ASR hypothesis in English.

Training set	Without Named Entity feature	With Named Entity feature
Clean data only (baseline)	81.15%	84.66%
MT-distorted data only	81.09% (-0.1%)	84.24% (-0.4%)
Clean+100% MT-distorted	83.11% (+2.0%)	86.67% (+2.0%)

5. DISCUSSION AND FUTURE WORK

In this paper, we study the effects of a training/testing "mismatch" condition, due to MT errors, on cross-lingual SLU. In order to mitigate this mismatch, we develop a multi-style adaptive training method. Without relying on data in the second language, the proposed method adapts the clean training data in the primary language to the MT-distorted condition, so as to address the training/testing condition mismatch problem for cross-lingual SLU. The method proves highly effective as is demonstrated through a series of cross-lingual SLU experiments.

The condition mismatch problem in cross-lingual SLU shares connection with similar problems in ASR and MT, and the methods in this paper are also motivated by similar solutions developed in speech recognition (e.g., unstructured and model-domain adaptation [3]) and machine translation (e.g., domain adaptation by data selection [1]). In the future, one of our focuses will be on the cross-fertilization of robust modeling and adaptation approaches between cross-lingual SLU and ASR/MT. On the other hand, a full cross-lingual SLU system consists of multiple components including ASR, MT, and SLU. Inspired by work of end-to-end discriminative learning for the ASR and MT components in speech translation [8][30], end-to-end optimization of all three components jointly for cross-lingual SLU as an extension of the straightforward noise-adaptation method presented in this paper is another key direction of our future work.

6. REFERENCES

[1] A. Axelrod, X. He, and J. Gao, "Domain adaptation via pseudo in-domain data selection," in Proceedings of EMNLP, 2011.

[2] N. Camelin, C. Raymond, F. Bechet, R. De Mori, "On the use of machine translation for spoken language understanding portability," in Proceedings of the ICASSP 2010.

[3] L. Deng, "Front-End, Back-End, and Hybrid Techniques to Noise-Robust Speech Recognition," in D. Kolossa and R. Hab-Umbach (eds.) Robust Speech Recognition of Uncertain Data, pp. 67-99, Springer Verlag, 2011.

[4] J. Dowding, J. M. Gawron, D. Appelt, J. Bear, L. Cherny, R. Moore, and D. Moran, "Gemini: A natural language system for spoken language understanding," in Proceedings of the ARPA Workshop on Human Language Technology, Princeton, NJ, March 1993.

[5] F. García, L-F. Hurtado, E. Segarra, E. Sanchis, G. Riccardi, "Combining multiple translation systems for spoken language understanding portability." In Proceedings of the IEEE-SLT 2012.

[6] F. Gey and D. Oard, "The TREC-2001 cross-language information retrieval track: Searching Arabic using English, French or Arabic queries," in Proceedings of TREC, 2001.

[7] Google Translation online service: http://translate.google.com/

[8] X. He, L. Deng, and W. Chou, "Discriminative learning in sequential pattern recognition," IEEE Signal Processing Magazine, 2008.

[9] X. He, M. Yang, J. Gao, P. Nguyen, and R. Moore. "Indirect HMM based Hypothesis Alignment for Combining Outputs from Machine Translation Systems," in Proceedings of EMNLP, 2008.

[10] Y. He and S. Young, "A data-driven spoken language understanding system," in Proceedings of the IEEE ASRU Workshop, U.S. Virgin Islands, December 2003.

[11] B. Jabaian, L Besacier, and F. Lefevre. "Combination of stochastic understanding and machine translation systems for language portability of dialogue systems," in Proceedings of the ICASSP 2011.

[12] B. Jabaian, L Besacier, and F. Lefevre. "Investigating multiple approaches for SLU portability to a new language," in Proceedings of Interspeech 2010.

[13] M. Kluck and F. Gey, "The domain-specific task of CLEF - specific evaluation strategies in cross-language information retrieval," in Workshop of the Cross-Language Information Evaluation Forum, CLEF 2000, Lisbon, Portugal, 2000.

[14] J. Lafferty, A. McCallum, and F. Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proceedings of the ICML, 2001.

[15] F. Lefevre1, F. Mairesse and S. Young "Cross-Lingual Spoken Language Understanding from Unaligned Data using Discriminative Classification Models and Machine Translation," in Proceedings of Interspeech 2010.

[16] X. Lei, J. Hamaker, and X. He, "Robust feature space adaptation for telephony speech recognition," in Proceedings of InterSpeech, 2006

[17] C-H. Li, X. He, Y. Liu, and N. Xi, "Incremental HMM alignment for MT system combination," in Proceedings of the ACL, 2009.

[18] Microsoft Translation online service: http://www.bing.com/translator/

[19] S. Miller, R. Bobrow, R. Ingria, and R. Schwartz, "Hidden understanding models of natural language," in Proceedings of the ACL, Las Cruces, NM, June 1994.

[20] T. Misu, E. Mizukami, H. Kashioka, S. Nakamura and H. Li, "A Bootstrapping Approach for SLU Portability to a New Language by Inducting Unannotated User Queries." in Proceedings of the ICASSP 2012.

[21] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. "BLEU: a method for automatic evaluation of machine translation," in Proceedings of ACL 2002.

[22] R. Pieraccini, E. Tzoukermann, Z. Gorelov, J.-L. Gauvain, E. Levin, C.-H. Lee, and J. G. Wilpon, "A speech understanding system based on statistical representation of semantics," in Proceedings of the ICASSP, San Francisco, CA, March 1992.

[23] P. J. Price, "Evaluation of spoken language systems: The ATIS domain," in Proceedings of the DARPA Workshop on Speech and Natural Language, Hidden Valley, PA, June 1990.

[24] C. Raymond and G. Riccardi, "Generative and discriminative algorithms for spoken language understanding," in Proceedings of the Interspeech, Antwerp, Belgium, 2007.

[25] S. Seneff, "TINA: A natural language system for spoken language applications," Computational Linguistics, vol. 18, no. 1, 1992.

[26] G. Tur, D. Hakkani-Tur, and L. Heck, "What is left to be understood in ATIS?" in Proceedings of the IEEE SLT Workshop, Berkeley, CA, 2010.

[27] G. Tur and R. D. Mori, Eds., Spoken Language Understanding: Systems for Extracting Semantic Information from Speech. New York, NY: John Wiley and Sons, 2011.

[28] W. Ward and S.Issar, "Recent improvements in the CMU spoken language understanding system," in Proceedings of the ARPA HLT Workshop, March 1994.

[29] S. Yaman, L. Deng, D. Yu, Y.-Y. Wang, and A. Acero, "An integrative and discriminative technique for spoken utterance classification," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 6, 2008.

[30] Y. Zhang, L. Deng, X. He, and A. Acero, "A novel decision function and the associated decision-feedback learning for speech translation," in Proceedings of the ICASSP 2011.