COMBINING MULTIPLE TRANSLATION SYSTEMS FOR SPOKEN LANGUAGE UNDERSTANDING PORTABILITY

F. García, L.F. Hurtado, E. Segarra, E. Sanchis[*]

Dept. Sistemes Informàtics i Computació
Universitat Politècnica de València, Spain

Giuseppe Riccardi[†]

Dept. of Information Engineering and Computer Science. University of Trento, Italy

ABSTRACT

We are interested in the problem of learning Spoken Language Understanding (SLU) models for multiple target languages. Learning such models requires annotated corpora, and porting to different languages would require corpora with parallel text translation and semantic annotations. In this paper we investigate how to learn a SLU model in a target language starting from no target text and no semantic annotation. Our proposed algorithm is based on the idea of exploiting the diversity (with regard to performance and coverage) of multiple translation systems to transfer statistically stable word-to-concept mappings in the case of the romance language pair, French and Spanish. Each translation system performs differently at the lexical level (wrt BLEU). The best translation system performances for the semantic task are gained from their combination at different stages of the portability methodology. We have evaluated the portability algorithms on the French MEDIA corpus, using French as the source language and Spanish as the target language.

The experiments show the effectiveness of the proposed methods with respect to the source language SLU baseline.

Index Terms— Spoken Language Understanding, Statistical Models, Language Portability.

1. INTRODUCTION

In the last few years, different approaches have been developed for the problem of SLU. As in other speech areas, statistical modelization has been successfully used in SLU [1], [2], [3], and [4]. There are many types of applications for SLU techniques; one of the most interesting is their application in limited-domain spoken dialog systems. The main goal of dialog systems of this kind is to obtain some information from a database. Then, generally, the system must interact with the user in order to obtain the information needed to fill out a template to perform a query. The size of the vocabulary in dialog systems of these kinds is small or medium; the semantic labels involved in the understanding process are related to the database queries, and they are strongly related to some specific words or segments of words of the user turns in the dialog.

Given the aforementioned characteristics of the limited-domain spoken dialog systems, it is possible to obtain manually transcribed, segmented, and labeled corpora that can be used to learn statistical models for the semantic decoding process. However, even in this case, the transcription, segmentation and labeling of the corpora is very time-consuming, and it is only useful for a specific task and for a specific language. Apart from the time-consuming work, manual segmentation and labeling have the disadvantage that it is sometimes difficult to decide a priori which limits of the segments are more accurate to represent a specific semantic label and to better discriminate from other semantic labels. Many efforts to avoid or minimize this manual work have been made in the last few years. This is the case of boostrapping, active learning, and semisupervised learning techniques [5].

The work presented in this paper attempts to obtain a SLU system for a target language from an annotated target language corpus obtained using a source-to-target translation process. This approach is similar to those presented in [6], in [7], in [8], and others, but there are some differences. The availability of an small parallel corpus French-Italian in the MEDIA corpus made possible for these works the use of an Stochastic Machine Translation (SMT) system, the MOSES toolkit [9], to obtain a translation from source to target language of the corpus. In contrast, there are not any parallel corpus French-Spanish available in the MEDIA corpus, therefore, it is not possible to estimate an SMT system, to obtain the translations. Therefore, we decided to use a combination of several online general-purpose translators, available in the web, in order to obtain the translations of both, complete sentences and segments of sentences. All these translations has been done without any human supervision.

The SLU in the source language is designed for the French MEDIA Corpus [10], which is labeled in terms of concepts, and the target language is Spanish. We have studied two types

[*]This work is partially supported by the Spanish MICINN under contract TIN2011-28169-C05-01, and by the Vic. d'Investigació of the UPV under contracts PAID-00-09 and PAID-06-10

[†]The author work was partially funded by FP7 PORTDIAL project n. 296170

of SLU techniques: a classical Conditional Random Field (CRF) approach [11] and a Two-level stochastic model [1], which is based on a Stochastic Finite-State Transducers. A segmented and labeled training corpus is necessary in order to estimate the understanding models for the two techniques. In order to obtain this segmented and labeled corpus to learn the models in the target language we have proposed a process of translation on the source corpus (using multiple translation systems, that is, a combination of several online general-purpose translators). Logically, the understanding results of this proposal are very dependent on the quality of this translation as well as on the strong or weak correlation in the way the same concepts in the two languages (source and target) are expressed. Some problems may appear in the target SLU system due to errors or ambiguity in the translation process: lack of coverage in the vocabulary, lack of a good Language Model (LM), and difficulty in translating the segmentation of the source language to the target language.

In this work, we obtained a Spanish corpus from the French MEDIA corpus in order to estimate the SLU systems and the LMs for Spanish. This has been done using different approaches: translating the sentences, translating the segments separately, using only one translator and combining translators. We have also developed two SLU systems, the Two-level system and the CRF system, for the French MEDIA corpus in order to compare the SLU results in both languages. A series of experiments with a development set has been performed to find the best translation combination as well as to determine the influence of the translation in the two SLU techniques. For evaluation purposes, we have generated and acquired a set of correct Spanish sentences. The results show that it is possible to obtain an accurate system using this approach.

2. TASK DESCRIPTION AND SEMANTIC LABELING

The MEDIA corpus [10] is a French dialogue corpus that simulates a telephone server for tourist information and hotel booking. It has been recorded using a Wizard of Oz technique. The corpus has 1,250 dialogs from 250 speakers; each speaker recorded five different hotel reservation scenarios. There is a total of 18,801 user turns with a vocabulary size of 2,715 words. These user turns were manually transcribed and conceptually rich with more than 80 manually annotated basic concepts.

For the Spanish SLU modelization, we used the training corpus defined in the MEDIA corpus and the techniques described in Section 3. From the test set of the MEDIA corpus, we selected two subsets: a subset of 323 sentences as the development set, and a subset of 1,012 sentences as the test set for the Spanish SLU system. We manually translated these French sentences to Spanish and we recorded them. These manually translated sentences were used as the reference in the evaluation of the Spanish SLU systems.

3. SEMANTIC MODELS

Two different SLU techniques have been studied, a generative technique (the Two-level) and a discriminant technique (a classical CRF).

To apply the Two-level technique [1], we assume that each user turn in the training set has a sequence of concepts (semantic units) associated to it, each of these concepts represents a piece of meaning of the user turn, and there is a segment (sequence of consecutive words) in the user sentence that is associated to each of these concepts. This approach consists of learning two types of finite-state models from the training set of pairs (u, c), where u is the sequence of segments and c is the corresponding sequence of concepts.

A model A_{s} for the *semantic language* is estimated from the sequences of concepts c that are associated to the input sentences. A set of models, *concept models* $A_{\mathrm{c_i}}$ (one for each concept c_{i}), is estimated from all the segments of words associated to this concept. The semantic model A_{s} represents the semantic information provided by the training data, and each concept model $A_{\mathrm{c_i}}$ represents the lexical and syntactic information for the corresponding concept c_{i}.

For the understanding process, all the models must be combined in order to take advantage of all the lexical, syntactic, and semantic constraints. To do this, the states of the stochastic automaton A_{s} are substituted by the corresponding stochastic automaton $A_{\mathrm{c_i}}$. Once this integrated automaton A_{t} is built, the understanding process consists of finding the best path in this automaton given the input sentence. In the experimentation, we used a model of 4-grams for the A_{s} automaton and 3-grams for the $A_{\mathrm{c_i}}$ automata.

CRFs have been successfully used for SLU tasks [4]. We applied CRFs to the MEDIA corpus using the CRF++ toolkit (http://crfpp.googlecode.com/svn/trunk/doc/index.html). We defined a set of basic features that includes only lexical information, setting a window such as incorporates the two previous and the two posterior words. A more complete set of features could be defined for applying the CRFs to SLU tasks [4], however, in this work we have not done a depth study of the best combination of features.

4. LEARNING MODELS

In order to obtain the semantic models in the target language from the initial semantic models, which are in the source language, we propose the following strategy. In a first step, we perform a translation of the training sentences. This translation can be done using an online general-purpose translator. One advantage of translators of this kind is that they are open domain; however, in counterpart they can introduce many errors. In order to avoid the problem of systematic errors introduced by a specific translator and to increase the coverage, we

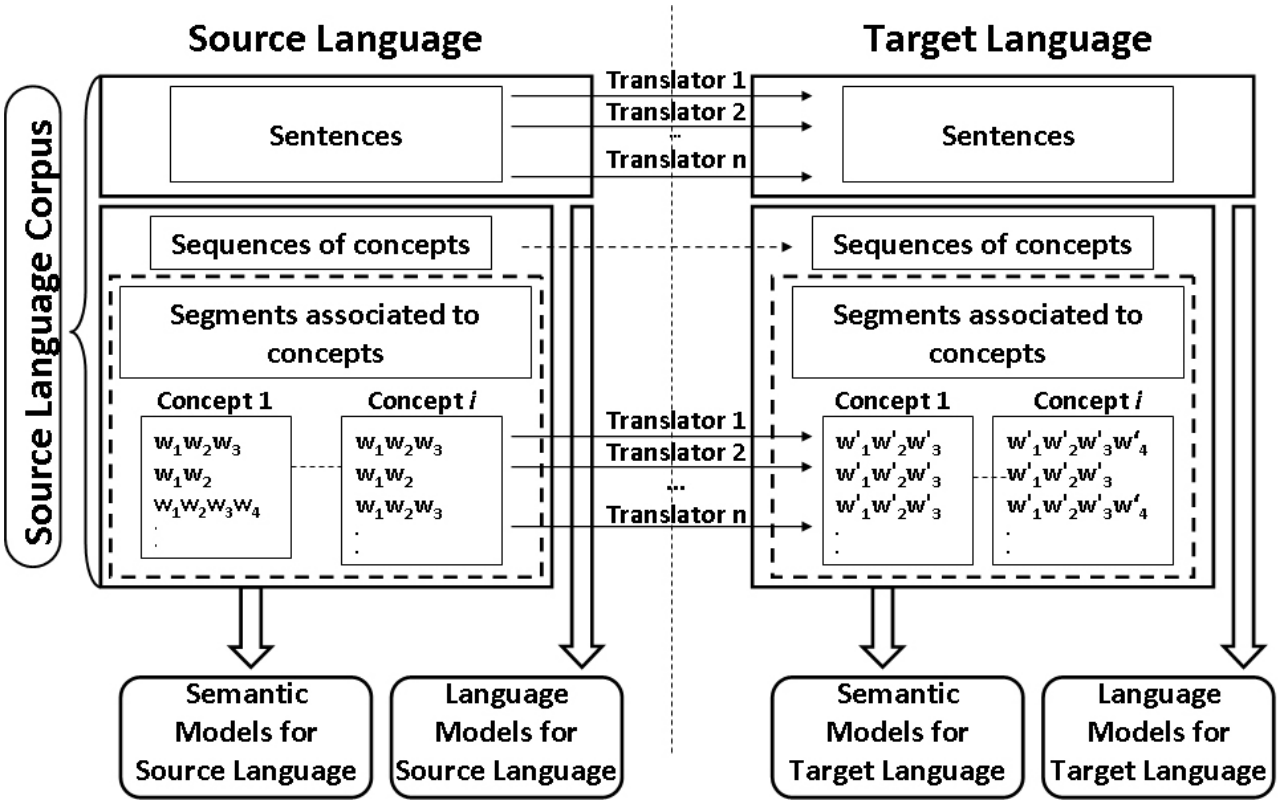

Fig. 1. Learning scheme

propose to use a combination of different translators. Table 1 shows the BLEU score of the five online general-purpose translation systems used in this work. These scores are measured on the manually translated media development set in Spanish.

Table 1. The BLEU score of the online general-purpose translators on the development set

	Translator	BLEU
t1	Apertium (http://www.apertium.org)	0.4724
t2	Lucy (http://www.lucysoftware.com)	0.6040
t3	Opentrad (http://www.opentrad.com)	0.5060
t4	Prompsit (http://www.prompsit.com)	0.5982
t5	Reverso (http://www.reverso.net)	0.7093

In the learning process, we assume that the same meaning is embedded in the same segment in both languages, that is, it is not distributed in many segments in the other language. This assumption is reasonable for closely related languages; however, it may not be accurate enough when the languages are very different. Figure 1 shows the scheme of the learning approach for the SLU and LM models in the target language:

- The translated sentences are used to learn the LM.
- The sequences of concepts (the same in the two languages) are used to estimate the semantic model, A_{s}, for the Two-level SLU system.
- The translations of the segments associated to each concept, c_{i}, are used to learn the $A_{\mathrm{c_i}}$ for the Two-level model, and to estimate the CRF model. If multiple translators are used, all the different translations are used to learn the these models.

5. EXPERIMENTS AND RESULTS

To study the behavior of the proposed approach, some experiments were performed. In all the experiments of this section the modelization of the sequences of concepts and the sequences of words associated to the concepts for the Two-level model was done using 3-grams and 4-grams. In the case of CRF++, they have been learnt using a window of two

words before and two words after. We used the training corpus defined in the MEDIA corpus (which consists of 12,811 French sentences) for the estimation of the two SLU models for French. We used these models to understand the 3,468 transcribed test set sentences, which is the test set defined in the MEDIA corpus. Table 2 shows the results in terms of concept accuracy (CA), the rate of correctly understood concepts and the total number of concepts in the reference. These measures are a reference of the behavior of the SLU system in the original language and are comparable with the results from other authors [12].

Table 2. SLU results for the test French MEDIA corpus

	Two-level	CRFs
CA	84.4	87.4

From the test set of the MEDIA corpus, we selected a subset of 323 sentences as the development set for the Spanish SLU system. We manually translated these French sentences to Spanish. We used this development set to perform a series of experiments in order to select the best combination of 5 different online general-purpose translators that were used in the translation process from French to Spanish (see Figure 1).

We did all the possible combinations of 1 translator, 2 translators, and so on, having a set of 31 combinations. This experimentation was done applying the two SLU techniques. Figure 2 shows the results of the experimentation with the development set and all the combinations of translators from French to Spanish. According to the results, the best translator combination for each group (1 translator, 2 translators, and so on) was selected for the next understanding experimentation. As Figure 2 shows, the best CA was obtained taking into account a specific combination of 3 translators. When the other 2 translators were added to the best combination, the CA slightly decreased.

In order to obtain LMs for the speech recognition process in Spanish, we learned 31 trigram models from all the combinations of the 5 translations of the training corpus, in a similar way to the above experimentation. The LM selected to be used for the speech recognizer was the one with the lowest perplexity for the development set, 38.6. This LM corresponds with the one learned using all the 5 translations. The perplexity for the original development set in French was 24.0. We used the SRILM toolkit for the estimations of the LMs.

From the test set of the MEDIA corpus, we selected a subset of 1,012 sentences as the test set for the Spanish SLU system. These sentences were manually translated from French to Spanish. In order to perform experiments with speech these test sentences were recorded by 5 speakers.

The Loquendo speech recognizer (www.loquendo.com) was used for the speech experiments. The LM used by the speech

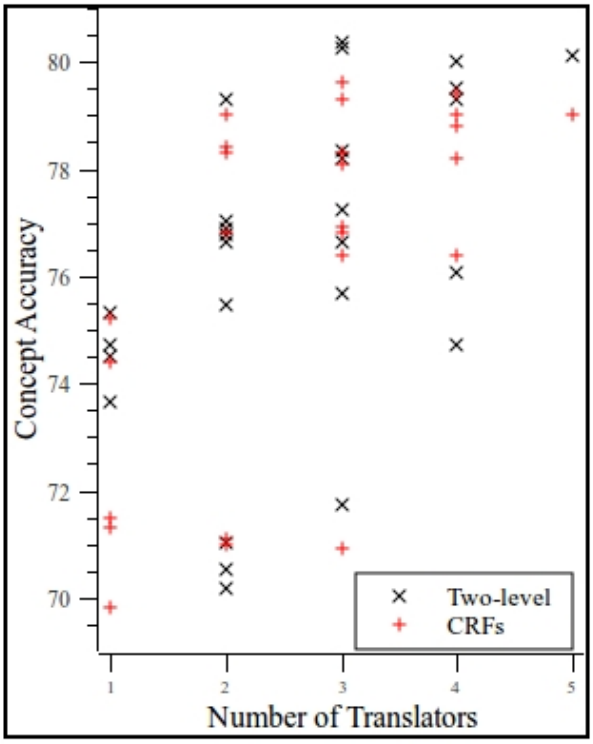

Fig. 2. Results of the SLU for the development set

recognizer were learned using the union of all the training sentences translated by the five translators. Using this LM, the Word Accuracy of the Loquendo speech recognizer was 83.7% for the test set.

In the experimentation, the two SLU systems were applied to the Spanish test set (both the written and the spoken sentences). Five experiments were performed with the best translator combination for each group selected from the development experiments (the best translator, the best combination of 2 translators, the best combination of 3 translators, and so on). The results of this series of experiments are shown in Table 3. In this table, the ASR is always the same for all speech experiments. As expected, the performance of the SLU systems from text were better than from speech. In comparison with the results for the SLU systems in French (Table 2), the loss in CA for the SLU systems in Spanish was rather small taking into account that a segmented and labeled corpus in Spanish was not available.

The best results are obtained with a combination of three translators (t2+t3+t5) whose BLEU scores for the development set are the best for two of them (t2+t5). The combination of these three translators improves the use of each one of them separately. We believe that this combination is able to solve some of the errors of each translator, even when a lower BLEU translator (t3) was included. However, the inclusion of more translators does not provide better results.

Table 3. Results of the SLU for the test set and the best combinations of translators from French to Spanish

Translator combination	Text		Speech	
	Two-level	CRFs	Two-level	CRFs
	CA	CA	CA	CA
t5	76.0	74.7	70.8	69.4
t2+t5	77.6	79.7	73.3	75.3
t2+t3+t5	**79.7**	**80.6**	**74.8**	**76.6**
t1+t2+t4+t5	78.7	80.3	73.9	76.4
all	78.4	80.5	73.5	76.4

6. CONCLUSIONS

In this paper, we have presented an approach for the portability of semantic modelization between different languages. It has been shown that models obtained for the target language give good results in terms of concept accuracy and also that the behavior is not very dependent on the kind of semantic modelization. Some problems derived from the errors in the translation processes can be solved by combining different translators. This approach has been applied to two closely related languages (latin languages) that have similar sequentiality in the way of expressing the semantics. Future work could be done to extend this approach to other similar latin languages; additionally, the study of the extension of this approach to languages that have very different linguistic characteristics could be considered; however, this extension involves dealing with the lack of a similar sequentiality. Another interesting line of study would be to consider the models obtained for the target language as preliminary models that can be used in a real application and that can be improved by interacting with the users by means of active learning or other incremental learning techniques.

7. REFERENCES

[1] E. Segarra, E. Sanchis, M. Galiano, F. García, and L. Hurtado, "Extracting Semantic Information Through Automatic Learning Techniques," *IJPRAI*, vol. 16, no. 3, pp. 301–307, 2002.

[2] Yulan He and Steve Young, "Spoken language understanding using the hidden vector state model," *Speech Communication*, vol. 48, pp. 262–275, 2006.

[3] R. De Mori, F. Bechet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur, "Spoken language understanding: A survey," *IEEE Signal Processing magazine*, vol. 25, no. 3, pp. 50–58, 2008.

[4] Stefan Hahn, Patrick Lehnen, Georg Heigold, and Hermann Ney, "Optimizing CRFs for SLU Tasks in Various Languages Using Modified Training Criteria," in *Proc. of Interspeech'09*, 2009, pp. 2727–2730.

[5] G. Riccardi and D. Hakkani-Tur, "Active learning: theory and applications to automatic speech recognition," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 504 – 511, july 2005.

[6] Christophe Servan, Nathalie Camelin, Christian Raymond, Frdric Bchet, and Renato De Mori, "On the use of Machine Translation for Spoken Language Understanding portability," in *Proc. of ICASSP'10*, 2010, pp. 5330–5333.

[7] Bassamn Jabaia, Laurent Besacier, and Fabrice Lefvre, "Investigating multiple approaches for slu portability to a new language," in *Proc. of InterSpeech'10*, 2010, pp. 2502–2505.

[8] Bassamn Jabaia, Laurent Besacier, and Fabrice Lefvre, "Combination of stochastic understanding and machine translation systems for language portability of dialogue systems," in *Proc. of ICASSP'11*, 2011, pp. 5612–5615.

[9] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, Moran C., R. Zens, C. Dyer, Bojar O., A. Constantin, and E. Herbst, "Moses: Open source toolkit for statistical machine translation," in *Proc. of ACL'07*, 2007, pp. 177–180.

[10] H. Bonneau-Maynard, Sophie Rosset, C. Ayache, A. Kuhn, and Djamel Mostefa, "Semantic annotation of the French MEDIA dialog corpus," in *Proc. of InterSpeech'05*, Portugal, 2005, pp. 3457–3460.

[11] John Lafferty, Andrew McCallum, and Fernando Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. on Machine Learning*, 2001, pp. 282–289.

[12] Stefan Hahn, Marco Dinarelli, Christian Raymond, Fabrice Lefevre, Patrick Lehnen, Renato de Mori, Alessandro Moschitti, Hermann Ney, and Giuseppe Riccardi, "Comparing stochastic approaches to spoken language understanding in multiple languages.," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1569–1583, 2011.