A Multilingual Spoken Language Understanding System

Sergio Laguna, Mayte Giménez, Marcos Calvo, Fernando García,
Encarna Segarra, Emilio Sanchis, Lluís-F. Hurtado

Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València
{slaguna,mgimenez,mcalvo,fgarcia,esegarra,esanchis,lhurtado}@dsic.upv.es

Abstract. We have developed a Multilingual Spoken Language Understanding System which is able to understand an utterance regardless of its language. In this web-based demo we have integrated different modules: language identification, automatic speech recognition, translation and speech understanding.

Keywords: Multilingual Spoken Language Understanding, Language Identification.

1 Introduction

This demo shows a Spoken Language Understanding (SLU) System which is able to semantically decode sentences from a restricted domain, which in this case is an information system about train timetables and fares, allowing us to use the DIHANA corpus [1] to train the models of the system. Furthermore, it has been endowed with mechanisms to work with languages other than the one the system was trained for. Therefore, we developed an automatic translation system to translate from new languages (French and English) to the system's language (Spanish). Also, we developed an automatic language identification module to detect which language user spoke and automatically set the translator for that language. All led to the development of a multilingual understanding prototype.

1.1 From a monolingual to a multilingual spoken language understanding system

We already had a language understanding system for Spanish spoken utterances [2]. But our aim is to expand it, so it would be able to understand several languages.

There are two approaches to achieve this goal:

- Train a new language understanding system for each language we want to understand. In this approach, we need to acquire and label new data and train a new system for each new language. This implies a significant effort.

– A semi-supervised approach for adapting a monolingual understanding system to sentences uttered in a new language. This, obviously simplifies porting SLU systems between languages. In a previous paper [3] we describe in deep this approach.

2 Description of the system

We are going to describe the Multilingual Spoken Language Understanding System we have developed.

Our process begins when the user provides an audio file to our system. This can be either and existing audio file, or can be recorded using the web browser.

Then, the system performs the following steps.

1. The system identifies the language of the audio choosing between English, French or Spanish.
2. Given a language, detected in the preceding step, the audio is recognized using a web-based recognizer.
3. If the detected language is not Spanish, the system will translate it into Spanish, as it is the language in which the models of the system are.
4. Then, a graph of words is created using the n-best hypothesis provided by the translator.
5. Finally, the system performs the semantic decoding of the graph of words, obtaining this way a set of detected concepts and a sequence of words associated to each concept.

Below, the system components are briefly described.

2.1 Language identification

Language identification can be stated as a classification problem.

We have developed a two phases approach to language identification:

– Acoustic-Phonetic Decoding (APD) of the spoken utterance using a set of Spanish phoneme models. This phase always uses the same phoneme models.
– Using the phonetic sequence from the transcription, the system assigns a language to it. It uses a language model of sequences of phonetic units learned for each language. The selection criterion is based on minimize the perplexity.

We used triphones as phonetic units, therefore we have context information. The acoustic model was learned from out-of-task corpus in Spanish. Moreover, the model of sequences of triphones used as language model was a trigram model of phonetic units.

Language models were learned from the APD output of 3446 spoken sentences uttered by several native English, French, and Spanish speakers. The English and French sentences are a translation of the DIHANA [1] corpus, which domain is restricted to information of long distance trains. However, Spanish sentences belong to the ALBAYZIN corpus which is a general-domain phonetically balanced corpus.

2.2 Automatic speech recognizer (ASR)

The input utterance is processed by an ASR for the language identified in the previous step. We use a general purpose, free-available web ASR, which means the ASR has no specific information about the task.

2.3 Automatic translation

These transcriptions are translated into the target language (Spanish) using a state-of-the-art Machine Translation system: MOSES [5]. The translation models were trained without using any manually translated data. Instead, a set of free-available web translators was used to translate the training sentences of the corpus from Spanish into the different languages, thereby building a parallel training corpus. MOSES provides as output a set of translation candidates (n-best list) of the transcription supplied by the ASR.

2.4 Graph of words generation

Our SLU system uses the n-best sentences obtained in the previous steps to generate a graph of words [4]. Figure 1 shows how this graph is obtained. First, a Multiple Sequence Alignment step is performed, in which we use a modification of the ClustalW software [6]. Then, we build the graph using the information contained in the alignment matrix, and computing the probabilities by the Maximum Likelihood criterion.

source utterance	*le prix du billet de train de six heures trente-cinq* *(the price of the train ticket for six thirty-five)* *(el precio del billete del tren de las seis treinta y cinco)*
multiple ASR outputs	le prix du billet train de sezer trente-cinq le prix du billet train de six vers trente-cinq le prix du billet train de six onze trente-cinq le prix du billet train des six heures trente cinq
multiple translations	el precio del billete de tren de sezer treinta y cinco el precio del billete de tren alrededor de las seis treinta y cinco el precio del billete del tren de las seis once y treinta y cinco el precio del billete de tren de las seis treinta de las cinco de la tarde

alignment matrix

el	precio	del	billete	de	tren	-	de	sezer	treinta	-	-	-	y	cinco	-	-	-
el	precio	del	billete	de	tren	alrededor	de	las	seis	treinta	-	-	y	cinco	-	-	-
el	precio	del	billete	del	tren	-	de	las	seis	once	y	treinta	y	cinco	-	-	-
el	precio	del	billete	de	tren	-	de	las	seis	treinta	-	de	las	cinco	de	la	tarde

Fig. 1. Steps for obtaining the graph of words from the original utterance *le prix du billet de train de six heures trente-cinq*, (*the price of the train ticket for six thirty-five*).

2.5 Spoken Language Understanding

The goal of this module is to provide the best semantic interpretation according to the information encoded in the graph of words. This semantic interpretation is defined according to a restricted domain, which in this case is the scope of the DIHANA corpus, it is, an information system about train timetables and fares. This corpus provides a set of concepts that are relevant for the domain, as well as a set of utterances and their transcriptions, which are segmented and labeled in terms of these concepts. This information is very useful to train statistical semantic models. This way, we have trained statistical models to represent how the words are joined within a specific concept, as well as a model to represent how the concepts are chained. Table 1 shows an example of the output of the system, assuming that the input is a written sentence.

Input utterance	*hola buenos días quería saber los horarios de trenes para ir a Madrid* *(hello good morning I'd like to know the train timetables to go to Madrid)*
Semantic segments	*hola buenos días* : courtesy *quería saber* : query *los horarios de trenes para ir* : <time> *a Madrid* : destination_city

Table 1. Example of the outputs of the SLU module.

The semantic decoding algorithm uses the statistical semantic models to process the graph of words. This algorithm first finds the possible attachments of sequences of words represented by a path in the graph to any of the concepts of the task. Then, it finds the best sequence of concepts using these attachments [3]. In previous works we have evaluated the behavior of this approach to multilingual SLU, achieving a 82.28% of concept accuracy when the input language is Spanish and 77.60% for French.

3 Experimental results

In this section, we present different experiments carried out to evaluate the performance of our system that validate the hypothesis we set.

3.1 Language identification minimizing the perplexity

In order to verify our approach for language identification we have conducted several experiments. We used the SRILM Toolkit [8] to estimate the phonetic language models for the classiers and the HTK Speech Recognition Toolkit [9] to perform the phonetic transcriptions.

Previously, we have published the results of this experimentation [7]. Summing-up, as we expected, lower perplexity appears when the language of the sentence

and the language of the model are the same. Moreover, we evaluated the performance of the Language Identication system. The global accuracy of the system was 0.841.

Table 2 shows the perplexity of the test set for the different languages and the accuracy of the system. It shows that the use of trigrams of phonetic units learned using a corpus only in Spanish is not as critic as we a priori expected. Overall, this approach allows us to identify the language of spoken utterances with limited resources.

		Perplexity		
		Test language		
		French	English	Spanish
LM Trigrams APD	French	**8.24**	11.62	12.16
	English	10.79	**6.63**	11.29
	Spanish	11.27	10.86	**7.57**

	Accuracy		
	Test language		
	French	English	Spanish
LM	0.793	0.850	0.960

Table 2. Perplexity of the phonetic language models and accuracy of the system *minimizing the perplexity.*

3.2 Changing the acoustic model

In order to performe the APD of all user utterances, we used an acoustic model trained only with Spanish audios from the TC-STAR corpus. We thought that our APD could improve with an universal acoustic model that takes into account also the acoustic variations of English and French. However, we did not have the resources to train this universal acoustic model, so we trained a new acoustic model using only audios in Spanish and English.

		Detected language		
		FR	EN	SP
TC-STAR AM	French	**67%**	8%	25%
	English	8%	**67%**	25%
	Spanish	13%	8%	**79%**

		Detected language		
		FR	EN	SP
DIHANA AM	French	**82%**	13%	5%
	English	4.1%	**92,9%**	3%
	Spanish	0,3%	0,3%	**99,4%**

Table 3. Language identification using different acoustic models

From the results shown at Table 3, a universal acoustic model improves language detection. However, when we used it with real users the results were not as good as expected.

4 Conclusions

In this paper, we have presented a web-based Multilingual Spoken Language Understanding system. We integrated different modules in this demo: language identification, automatic speech recognition, automatic translation and an understanding system that we adapted to extract semantic interpretations from sentences in different languages.

References

1. Benedí, J.M., Lleida, E., Varona, A., Castro, M.J., Galiano, I., Justo, R., López de Letona, I., Miguel, A.: Design and acquisition of a telephone spontaneous speech dialogue corpus in Spanish: DIHANA. In: Proceedings of LREC 2006. pp. 1636–1639. Genoa (Italy) (May 2006)
2. Calvo, M., García, F., Hurtado, L.F., Jiménez, S., Sanchis, E.: Exploiting Multiple ASR Outputs for a Spoken Language Understanding Task. In: Speech and Computer, pp. 138–145. Springer International Publishing (2013)
3. Calvo, M., Garcıa, F., Hurtado, L.F., Jiménez, S., Sanchis, E.: Exploiting multiple hypotheses for Multilingual Spoken Language Understanding. CoNLL-2013 pp. 193–201 (2013)
4. Calvo, M., Hurtado, L.F., García, F., Sanchis, E.: A Multilingual SLU System Based on Semantic Decoding of Graphs of Words. In: Advances in Speech and Language Technologies for Iberian Languages, pp. 158–167. Springer (2012)
5. Koehn, P., et al.: Moses: Open Source Toolkit for Statistical Machine Translation. In: Proc. of ACL demonstration session. pp. 177–180 (2007)
6. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G.: ClustalW and ClustalX version 2.0. Bioinformatics 23(21), 2947–2948 (Nov 2007)
7. Sanchis, E., Giménez, M., Hurtado, L.F.: Language identification with limited resources. V Jornadas TIMM pp. 7–10 (2014)
8. Stolcke, A., et al.: SRILM-an extensible language modeling toolkit. In: INTERSPEECH (2002)
9. Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., et al.: The HTK book, vol. 2. Entropic Cambridge Research Laboratory Cambridge (1997)