



# Investigating multiple approaches for SLU portability to a new language

Bassam Jabaian<sup>1,2</sup>, Laurent Besacier<sup>1</sup>, Fabrice Lefèvre<sup>2</sup>

<sup>1</sup> LIG, University Joseph Fourier, Grenoble - France

<sup>2</sup> LIA, University of Avignon, Avignon - France

{bassam.jabaian, laurent.besacier}@imag.fr , fabrice.lefevre@univ-avignon.fr

## Abstract

The challenge with language portability of a spoken language understanding module is to be able to reuse the knowledge and the data available in a source language to produce knowledge in the target language.

In this paper several approaches are proposed, motivated by the availability of the MEDIA French dialogue corpus and its manual translation into Italian.

The three portability methods investigated are based on statistical machine translation or automatic word alignment techniques and differ in the level of the system development at which the translation is performed.

The first experimental results show the efficiency of the proposed portability methods in general for a fast and low-cost SLU porting from French to Italian and the best performance are obtained by using translation only at the test level.

**Index Terms:** Dialogue systems, Spoken Language Understanding, Languages Portability, Statistical Machine Translation.

## 1. Introduction

The portability of a spoken language system from a language to another is the subject of many researches as presented by recent papers [1, 2, 3]. To build a Spoken Language Understanding (SLU) we need to have annotated data reflecting our knowledge of a specific domain. The language portability of this system consists on porting this data already available in a first language to produce annotated data in the second language. Let us define the first language as a source language and the second language as a target language.

Recently, it has been shown that the use of an automatic translation system at different level of the understanding process can help to achieve this goal. For example in [2] the authors automatically translate data from a language to another and then use a retrained stochastic grammar to perform recognition and interpretation in the target language.

Another possibility is to consider that the semantics of a domain is independent of the language. Then, one solution consists in translating the entire training corpus into the target language. As described in [3], sentences of the training corpus are composed of one or several chunks, each chunk having a semantic annotation. By translating the training corpus considering these chunks it is possible to match the semantic annotation of the source corpus with the translated corpus. In [3] it is shown that the portability of a SLU is possible with this approach using either manual or automatic translation.

The choice of an approach depends on theoretical considerations and also on the domain characteristics and the data available. Having data manually translated, human capacities and specific tools can make a difference on the choice of the approach.

In this paper we propose several approaches motivated by the availability of the French MEDIA corpus described in [4],

and the availability of a manual translation into Italian of a part of this corpus. It is important to note that in this paper (as opposed to [3]), since the manual translation is done at the sentence level, we don't assume that there is a simple match between the Italian and the French concept chunks.

The approaches studied in this paper are fully automatic, without human supervision during the portability process. The human contribution is limited to the translation of a part of the MEDIA corpus. Most of the approaches use a Statistical Machine Translation (SMT) process or, at least, an automatic bilingual word alignment process. The French set and its Italian translation are used as a parallel corpus to train a SMT system in both directions (French to Italian and Italian to French) using the Moses toolkit [6].

The main difference between the approaches presented in this study is not only the method used to port the system but also the level at which the SMT process takes place.

One first basic approach (that will be referred to as *Test Translation* in the rest of the paper) is to build a system on the data already annotated on the source language. Then, the input sentences in the target language are automatically translated back to the source language before being sent to the input of the source SLU system. In other words, the SLU system remains in the source language and the target language inputs are translated with a target-to-source SMT system.

A second approach (that will be referred to as *Tagged Translation*) is to build a new SLU system for the target language with an annotated target language corpus obtained using a source-to-target SMT system. In that case, each source annotated sentence, segmented into chunks is translated to the target language and the translated target chunks are used to train a new SLU system in the target language. This approach is similar to the one presented in [3] but without any human supervision.

The third approach (that will be referred to as *Alignment*) does not use any SMT system but only information extracted from word-to-word alignments of the source-target bilingual corpus. In that case, the target chunks and their associated semantic concept are directly inferred from these word alignments.

The performance of the different SLU systems built with these approaches depends not only on the performance of the chosen concept tagging method used to train them, but also on the quality of the SMT systems used during the portability process. In our study, Conditional Random Fields described in [5] are used to train our SLU systems and we evaluate their performance using the well-known concept error rate (CER), while the quality of the SMT systems is measured by means of the BLEU score.

The paper is structured as follows: Section 2 presents in more details the three proposed approaches for language portability. In Section 3 we will describe the corpus used in our experiments and the tools used to translate this corpus to Italian. Section 4 presents the experimental results while Section 5 concludes this work.

## 2. Spoken Language Understanding Portability

In this study, we used a SLU system developed on the MEDIA corpus (see Section 3). The Conditional Random Fields (CRF) as described in [5] are used to train the model. The corpus is tagged by concepts representing a semantic view of the information existing in the corpus. The SLU process extracts a list of concept tag hypotheses from an input sentence.

The generation of these concepts has been described in [7] and can be briefly explained as follows.

Let  $V$  be a vocabulary of words given as hypotheses of an ASR system.

Let  $C = c_1, \dots, c_n$  be a sequence of concept tags that can be hypothesized from a sequence of words  $W = w_1, \dots, w_n$  as an initial step.

To each concept  $c_i$  is associated a sequence of words in  $W$ ; a label is associated to each word; this label corresponds to the concept tag  $c_i$  and the position of  $w_i$ .

To denote concept boundaries the BIO formalism, as proposed by [7], is used. For example, the sentence:

« *Je voudrais réserver un hotel à Paris* »

will be represented as sequence of pairs  $(W, L)$ :

(Je, B\_command-tache) (voudrais, I\_command-tache) (réserver, I\_command-tache) (un, B\_objet) (hotel, I\_objet) (à, B\_localisation-ville) (Paris, I\_localisation-ville)

To train a CRF tagger we convert the training data to the BIO format then we use the CRF++ toolkit [11] to generate the model.

CRF is a log-linear model, and can be represented as follows:

$$p(c_1^N | w_1^N) = \frac{1}{Z} \prod_{n=1}^N H(c_{n-1}, c_n, w_{n-2}^{n+2})$$

with

$$H(c_{n-1}, c_n, w_{n-2}^{n+2}) = \sum_{m=1}^M \lambda_m \cdot h_m(c_{n-1}, c_n, w_{n-2}^{n+2})$$

The log-linear models are based on feature functions  $h_m(c_{n-1}, c_n, w_{n-2}^{n+2})$  representing the information extracted from the training corpus,  $Z$  is a normalization term and  $\lambda$  is estimated during the training process.

Talking about language portability we need now to define how machine translation intervenes in the process. In order to port a SLU system to a new language we propose several approaches and compare their performance. Our propositions can be divided into two main approaches based on the level at which the portability is done.

### 2.1. Test Translation

This first approach assumes the availability of a SLU system in the source language, and uses a SMT system to translate the test set from the target language to the source language, then use this translation as an input to the SLU system already available. In another way this approach consists in keeping the system as it is and translate the input sentence from the target to the source language.

The main problem of this approach is that the input of the system is a sentence generated automatically. It may be noisy and its quality highly depends on the performance of the SMT system.

### 2.2. Train Translation

Since CRF training needs only an annotated training corpus without any other information, the portability of the system can be made by translating the train corpus to the target language, and the challenge is to port the annotation from the annotated corpus to the translated one (in other words to create automatically a new training corpus). To perform this annotation portability we propose two different approaches.

#### 2.2.1. Translating with XML Tags (Tagged Translation)

In order to have a semantic representation of the new translated corpus we use the MOSES decoder with an option that allows to force particular sentence segmentation into phrases (or chunks). This is done by placing XML tags in the text to translate. This option projects the tags into the translated text and prevents any attempt to modify the boundaries coming from the source language segmentation.

The example given before can be represented for this approach as:

```
<tag c=command_tache > Je voudrais  
reserver </tag> <tag c=objjet > un hotel  
</tag> <tag c=localisation_ville> a Paris  
</tag>
```

Using Moses that takes into account these XML tags as a phrase segmentation information, the following output is derived:

```
<tag c= command_tache > vorrei prenotare  
</tag> <tag c=objjet> un hotel </tag> <tag  
c= localisation_ville> a Parigi </tag>
```

All the train corpus is translated that way, then we introduce the BIO format and use it as a train corpus to build a SLU system in the target language.

A similar approach was proposed in [3] but, as we understand what is described in the paper, the translation outputs were manually post edited by Italian annotators whereas we do not perform any manual revision and use the data as it is (fully automatic process).

#### 2.2.2. Inferring Concepts using Word-to-Word Alignments (Alignment)

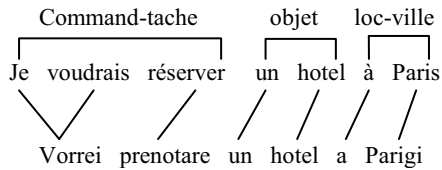
Word alignment is an important step of the statistical machine translation process. Several toolkits are available to do an unsupervised word alignment like GIZA++ [8] that uses IBM and HMM models, or Berkeley aligner [9] which uses the alignment by agreement model, and shows a good alignment quality.

Back to the main idea of porting the SLU to a new language with the least effort, we propose to use the word to word alignment approach to infer a training corpus in the target language.

In our case, the starting point is the sentence aligned bilingual corpus of 5.6k sentences (manual Italian translation of a part of the French training corpus). From this, it is not obvious to match directly the semantic concepts to the chunks in the target language, though many cases might be unambiguous as the example shown in Fig 1.

The idea is thus to use the word-to-word alignment information (in this experiment we use the Berkeley aligner) to match target language chunks and semantic concepts.

Fig 1: Example of inferring concepts using word to word alignment



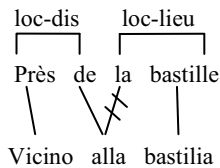
To do this, we developed an algorithm that uses alignment information and boundaries between chunks in the source corpus to infer concepts on the target corpus. For each chunk, the program maps the corresponding words in the target corpus.

In order to minimize the noise in the created corpus, words of the target language aligned to different words of the source corpus need to be treated specifically.

When a target language word is aligned to several words belonging to different concepts (see Fig 2), we have to take the decision on which concept must be associated to the target word. Our proposition is to consider it aligned with only the first word that occurs in the sentence of the source language. This simple strategy might not work for all cases, but it has the advantage to be consistent along the whole training data.

This strategy allows us to annotate the entire target corpus using the word to word alignment information and the concept tags of the source corpus.

Fig 2: Example of an ambiguous situations concept inference



### 3. Corpus and Tools Description

The French MEDIA corpus is the base corpus on which all the experiments of this paper have been made. As described in [4], this corpus covers a domain related to the reservation of hotel rooms and tourism information. The corpus is made of 1257 dialogs from 2500 speakers divided into three parts: a training set (approx. 13k sentences), a development set (approx. 1.3k sentences) and an evaluation set (approx. 3.5k sentences).

A part of the training set (approx. 5.6k sentences), as well as the development and test sets have been translated manually to Italian in the context of the LUNA project [3].

In this study, we use two statistical machine translation systems to obtain automatic translation from French to Italian and from Italian to French. To perform these translations we use the Moses toolkit [6]. Moses is a state-of-the-art phrase-based translation system using log-linear models.

We use the part of the training corpus of MEDIA, manually translated into Italian as a parallel corpus for Moses in both directions to learn translation models. We use the French part to obtain a French language model and the Italian part to obtain an Italian language model. Also, the development set of MEDIA with its translation is used as a parallel corpus to tune the log-linear weights of the SMT systems. Finally we obtain a French to Italian system with a BLEU score of 43.62, and an Italian to French system with a

BLEU score of 47.18. Those scores are measured on the manually translated test set of MEDIA. Since only one reference per utterance is used to evaluate BLEU, and since a small SMT training data is used (5.6k), these performances can be considered as acceptable.

These systems allow to derive an automatic translation of the remaining (not manually translated) part of the train set; so a full translation of the MEDIA train (manual + automatic) is available. Table 1 gives a brief overview of the available corpora at this point.

Table 1: Overview of the MEDIA corpus and its Italian translation (total number of sentences).

MEDIA data	Train	dev	Test
<b>French MEDIA</b>	13K	1.3K	3.5K
<b>Italian manual</b>	5.6K	1.3K	3.5K
<b>Italian automatic</b>	7.4K	-	-

### 4. Experiments and Results

The MEDIA test corpus and its manual translation to Italian are used as tests to compare the performance of the various approaches. As evaluation criteria, we use the CER which is defined as the ratio of the sum of deleted, inserted and substituted concepts by the total number of concepts in the reference.

Our experiments can be divided into two main parts:

1. **Partial corpus:** only the subset of the training data of the MEDIA corpus that is manually translated (5.6k) is used. A baseline CRF tagger using unigram and bigram is trained on this part of the MEDIA train corpus as a reference to compare with, then we apply the three methods: Test Translation, Tagged Translation, and Alignment as described in Section 2. Results are reported in Table 2. Comparing the three portability approaches shows that the *Test Translation* method is slightly better than the *Tagged Translation* method. In both cases, the results are encouraging and the drop in performance compared to the same SLU system is less important than we would have expected, considering even more that the SMT system can be improved in the future.

Table 2: Evaluation (CER %) of Italian models trained on 5.6k sentences manually translated (Performance of French model on the same amount of data is given as a reference)

Model	Test	Sub	Del	Ins	CER
<b>FR</b>	FR	2.9	13.9	1.7	18.5
<b>Test Translation</b>	IT	5.1	15.2	2.8	23.1
<b>Tagged Translation</b>	IT	5.4	17.8	1.1	24.2
<b>Alignment</b>	IT	4.7	16.7	1.5	22.8

2. **Full corpus:** the entire training corpus of MEDIA (13K) is now used. In that case, the experiments presented in Table 2 are repeated but on the full MEDIA corpus. The manual translation of the training set is completed automatically (see last line of Table 1) to be used for the alignment method. Results are given in Table 3. The difference between the scores obtained by the French model trained on the entire MEDIA corpus

and on a subset on this corpus (from 12.9% to 18.5%) shows that the size of the training data has an important influence on the performance of a CRF model.

The results also show clearly that the alignment method is improved from 22.8% to 20.5%. This CER remains higher than the one obtained on the full corpus with the *Test Translation* approach (19.9%).

Even the score obtained on the full corpus with the *Tagged Translation* method decreases (from 24.2% to 22.7%) but it remains the method giving the worst performance.

It is important to note that these results were obtained on the test data of the MEDIA corpus manually transcribed and not on the output of a speech recognizer and the methods rating may be different on noisy data.

Table 3: Baseline French SLU vs. SMT/SLU approaches on the full MEDIA corpus (CER %)

Model	Test	Sub	Del	Ins	CER
FR	FR	3.1	8.1	1.8	12.9
Test Translation	IT	5.2	12.1	2.6	19.9
Tagged Translation	IT	3.7	16.9	2.1	22.7
Alignment	IT	3.1	15.0	2.3	20.5

In this last experiment, we intend to evaluate the gain obtained using additional information in the SLU system. A useful information is the Part Of Speech (POS) tagging.

Since we do not have a POS tagger for Italian we propose to use the alignment information obtained on the French-Italian parallel corpus to enrich each Italian word by its corresponding French word (or words) in the parallel corpus. That way, we think that the source language words are useful to disambiguate the target language word, just as POS tags are doing in a conventional approach.

We already have the automatic alignments of the parallel corpus, and we use this information to “tag” Italian words. Each word from the Italian train corpus is “tagged” by French word or words which are aligned with it. Then, a SLU CRF model is trained using this enriched corpus.

To test this model we need to tag the Italian test set by French words before generating the hypothesis by the SLU. To do that we automatically generate a word to word translation of the Italian test corpus to French. In order to have a reference to compare the quality of this approach we also tag the French MEDIA corpus using the LIA tag [10] and we train a model on this corpus. Two Italian alignment models with French tags are trained, one based only on the manual parallel corpus to be comparable with results shown in Table 2, and the other based on the full parallel corpus (manually+automatically translated) to be comparable with results shown in Table 3. We use exactly the same evaluation tools and corpus. Results are given in Table 4.

Comparing the score obtained with the alignment method, with and without additional information shows us that, in the context of our experiments, adding information (even noisy) to the training process increase the performance of the SLU model. The addition of source language words to disambiguate the SLU input in the target language improves the Italian SLU system from 22.8 to 21.2 (5.6k corpus) and from 20.5 to 20.1 (full corpus), and seems efficient when few data is available to train the SLU system.

Table 4: Evaluation (CER %) of French and Italian models trained with additional knowledge

Model	Test	Sub	Del	Ins	CER
FR+POS tags	FR	3.6	6.4	2.2	12.1
5.6k Alignment+FRtags	IT	4.4	15.7	1.2	21.2
13K Alignment+FRtags	IT	3.8	14.4	1.9	20.1

## 5. Conclusion

Several approaches to port a SLU system from a language to another were presented and studied in this paper.

We suggest to port it either at the test level by translating automatically the test set, or at the train level by translating the training data and using different method to port the semantic annotation.

The results show the interest of the proposed approaches as the performances in the target language all are only few percents below the source language.

The best performance is obtained by directly translating the user’s request and making use of the original SLU system. We are aware that the results shown in this paper are evaluated on a clean test set and we plan to reproduce them on speech recognizer’s outputs as soon as they will be available.

## 6. Support

This work is supported by the ANR funded PORT-MEDIA project (ANR 08 CORD 026 01). For more information about the PORT-MEDIA project, please visit the project website, [www.port-media.org](http://www.port-media.org)

## 7. References

- [1] D. Suendermann, K. Evanini, J. Liscombe, P. Hunter, K. Dayanidhi, R. Pieraccini, “From rule-based to statistical grammars: Continuous improvement of large-scale spoken dialog system,” in *ICASSP 2009*.
- [2] D. Suendermann, J. Liscombe, K. Dayanidhi, R. Pieraccini, “Localization of speech recognition in spoken dialog systems: How machine translation can make our lives,” in *INTERSPEECH 2009*.
- [3] C. Servan, N. Camelin, C. Raymond, F. Bechet, R. De Mori, “On the use of machine translation for spoken language understanding portability,” in *ICASSP 2010*.
- [4] H. Bonneau-Maynard, S. Rosset, C. Ayache, A. Kuhn, and D. Mostefa, “Semantic annotation of the French media dialog corpus,” in *Ninth European Conference on Speech Communication and Technology. ISCA, 2005*.
- [5] J. Lafferty, A. McCallum, F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *ICAL 2010*.
- [6] P. Koehn, H. Hoang, A. Birch, C. Callisonburch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and E. Herbst, “Moses: Open source toolkit for statistical machine translation,” in *ACL 2007*.
- [7] C. Raymond and G. Riccardi, “Generative and discriminative algorithms for spoken language understanding,” in *INTERSPEECH 2007*.
- [8] Franz Josef Och, Hermann Ney. "Improved Statistical Alignment Models", *In ACL 2000*.
- [9] P. Liang, B. Taskar, D. Klein, “Alignment by agreement,” in *HLT 2006*.
- [10] LIA-TAGG is available online: [http://lia.univ-avignon.fr/fileadmin/documents/Users/Intranet/chercheurs/bechet/download\\_fred.html](http://lia.univ-avignon.fr/fileadmin/documents/Users/Intranet/chercheurs/bechet/download_fred.html)
- [11] CRF++ is available online: <https://crfpp.sourceforge.net>