Investigating the Role of Machine Translated Text in ASR Domain Adaptation: Unsupervised and Semi-supervised Methods

Horia Cucu [#*1], Laurent Besacier [*2], Corneliu Burileanu [#3], Andi Buzo [#]

[#] *ETTI, University "Politehnica" of Bucharest*
Bucharest, Romania
[1] horia.cucu@upb.ro
[3] cburileanu@messnet.pub.ro

[*] *LIG, University Joseph Fourier*
Grenoble, France
[2] laurent.besacier@imag.fr

***Abstract*—This study investigates the use of machine translated text for ASR domain adaptation. The proposed methodology is applicable when domain-specific data is available in language X only, whereas the goal is to develop a domain-specific system in language Y. Two semi-supervised methods are introduced and compared with a fully unsupervised approach, which represents the baseline. While both unsupervised and semi-supervised approaches allow to quickly develop an accurate domain-specific ASR system, the semi-supervised approaches overpass the unsupervised one by 10% to 29% relative, depending on the amount of human post-processed data available. An in-depth analysis, to explain how the machine translated text improves the performance of the domain-specific ASR, is also given at the end of this paper.**

I. INTRODUCTION

Language and acoustic resources creation, for any given spoken language, is typically a costly task. For example, a large amount of time and money is required to properly create annotated speech corpora for automatic speech recognition (ASR), domain-specific text corpora for language modeling (LM), tagged text corpora for part-of-speech (POS) tagging, parallel text corpora for statistical machine translation (SMT), etc. The development of ASR systems for the already high-resourced languages (such as English, French or Mandarin, for example) is less constrained by this issue and, consequently, high-performance commercial systems are already on the market. On the other hand, for under-resourced languages, the above issue is typically the main obstacle.

Given this, the scientific community's concern with porting and adapting language and acoustic resources or even models from high-resourced languages to low-resourced languages makes perfect sense. Several algorithms and methods of adaptation have been proposed and experimented lately. The use of sub-word units for language modelling, applied to languages such as Somali, Amharic and Hungarian is described in [1]-[3], while [4] presents two techniques (cross-lingual and grapheme-based acoustic modeling) for bootstrapping acoustic models for a new language (Vietnamese). Several approaches for language portability (French to Italian) of dialogue systems are being investigated in [5]. In a previous work [6], we have analysed the use of machine translated text for language model portability (French to Romanian) in the context of ASR. The method proposed in [6] was evaluated in a preliminary fashion and was fully unsupervised. Similar efforts for porting written language resources using machine translation are being reported in [7] (for English to Icelandic) and [8] (for English to Japanese and vice-versa).

For Romanian (the target language of this work), the latest studies in ASR report only the usage of basic language models such as word loop grammars [9], [10], specific task grammars [11] or small vocabulary bigram models [11]. Moreover, a common planned future work for all these studies is to create better (general or domain-specific) language models for Romanian. In our previous work [6] we have used the Web as a language modeling resource and managed to create a satisfactory large-vocabulary LM, which has been evaluated in the context of ASR. Another obstacle for Romanian is the absence of diacritics in most of the corpora which are freely available on Internet. Our previous study [6] reports that diacritics restoration is mandatory for Romanian ASR and describes a basic diacritics restoration method. This method is used in this study as well because [6] also concludes that this diacritics restoration system is as good (from the ASR point of view) as the best diacritics restoration system available for Romanian [12].

This paper presents two SMT-based methodologies for porting a domain-specific (tourism) French corpus to Romanian, with the final goal of creating a domain-specific ASR system for Romanian. The unsupervised approach introduced in [6] is regarded as a baseline. The methodology used in [7] is fundamentally different than ours, because rule-based machine translation techniques are utilized for language portability (English to Icelandic), while the domain adaptation is being done with original (not machine translated) text. Moreover, our study also investigates *why and how* the machine translated text improves the domain-specific ASR system. This research is also different and more in-depth than the investigations conducted in [8], which focuses on SMT-based language portability as well, but only reports perplexity experiments, without investigating the implications on a full ASR system. In conclusion, the results given in these papers cannot be directly compared to ours.

In addition, the Romanian ASR system evaluated in this paper makes use of a SMT-extended phonetic dictionary. The method for extending the pronunciation dictionary is very similar to the one presented in [13] and [14] and will be briefly described and evaluated in Section III. Before this, Section II presents the unsupervised and semi-supervised adaptation methods. Section IV describes the experimental setup and Section V discusses the ASR evaluation results. A more in-depth analysis, explaining *why and how* this methodology improves ASR performance, is given in Section VI and, finally, Section VII draws some conclusions.

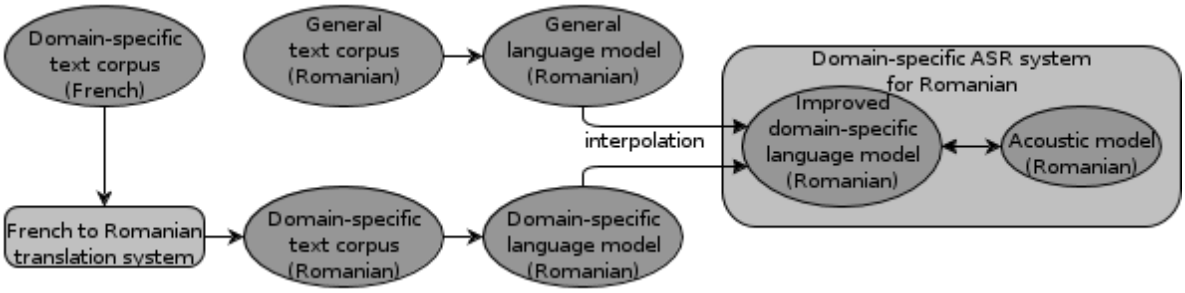

Fig. 1. The general translation-based ASR domain adaptation methodology

II. SMT-BASED LANGUAGE MODEL DOMAIN ADAPTATION

The method we propose aims to adapt a general language model for Romanian to a specific domain, given a French text corpus for that particular domain. The final goal is to use the adapted language model, among with an acoustic model, for a domain-specific ASR task for Romanian.

Fig. 1 depicts the general methodology we propose to create the domain-specific language model for Romanian. Basically, a French-to-Romanian translation system is required to translate the French domain-specific corpus to Romanian. The translation system can be a human expert performing manual translation, an unsupervised machine translation system or a combination of the two.

The language model trained using only the machine translated corpus can be utilized for speech recognition as well, but, as Fig. 1 illustrates, we expect to obtain better results if this domain-specific language model is interpolated with a general language model for the target language. Intuitively, the domain-specific language model will include many specific words and sequences of words, but it will probably have a poor coverage over the general language structures. On the other hand, the most frequent language structures are usually well modeled by a broader, out-of-domain language model, created using larger Romanian corpora. The interpolation of the two language models should lead to an improved domain-specific language model for Romanian.

A. Unsupervised SMT-based Adaptation Method

The domain adaptation methodology presented in [6], translates the in-domain French corpus using the online Google (French-to-Romanian) MT system. This translation is utilized to create the domain-specific language model without any human post-edition (unsupervised) and, obviously, the machine generated corpus contains several errors. Nevertheless, as we will show in the results section, the domain-specific language model created by this unsupervised method is much more suitable for the domain-specific ASR task than a general language model.

B. Semi-supervised SMT-based Adaptation Methods

For the unsupervised adaptation method described above we have used the imperfect Google MT system and obtained some speech recognition results. These results are better than the ones obtained using a general (out-of-domain) ASR system, which does not use any domain-specific information. Nevertheless, the ideal (performance-oriented) scenario would imply having the French corpus manually translated to Romanian by a human expert. The question that arises is: how much better would have been the results in this *ideal* scenario?

To progressively answer this question, we started to correct the Google translated corpus. Ideally, the whole corpus should have been corrected, but, due to time constrains, we have only corrected a part of it (*xx%*). This post-processed part of the corpus is further called *xx%GMTpp*. This part was afterwards concatenated with the rest of the Google translated corpus (denoted *rest%GMT*) to obtain a complete (100%) domain-specific corpus. This is the first semi-supervised method of obtaining a Romanian domain-specific corpus and it is graphically represented in the upper part of Fig. 2.

The second semi-supervised adaptation method regards the *xx%* of the French domain-specific corpus and the Romanian *xx%GMTpp* corpus as parallel corpora and uses them to train a domain-specific machine translation system. Undoubtedly, the resulting SMT system will be worse than Google's when *xx%* is small, but it may outperform Google's as *xx%* increases. The trained SMT system is afterwards used to translate the rest of the domain-specific corpus. This part of the corpus, which is obtained by translation with our own domain-specific SMT system, is further called *rest%dsMT*. In the end, as Fig. 2 depicts, *xx%GMTpp* is concatenated with *rest%dsMT* to obtain a complete (100%) domain-specific corpus.

Fig. 2 describes the whole methodology for obtaining the two Romanian partly-post-processed corpora, given the French domain-specific corpus. These two corpora have been further used to create domain-specific language models and, eventually, domain-specific ASR systems, using the general methodology presented in Fig. 1.

III. SMT-BASED PHONETIC DICTIONARY EXTENSION

Any ASR system uses a phonetic dictionary to translate between the graphical representation of words and the corresponding phonetic sequence (the actual pronunciation). For Romanian, we already had an extensive (600k words), general phonetic dictionary [9], which was manually updated with proper names (hotels, places, etc.) for the domain-specific task [6]. Still, the phonetic dictionaries used in the experiments presented in [6] lacked several thousand words within the corresponding LMs (all the LMs had 64k unigrams and among these only 45k – 50k words had phonetic representations). Consequently, an automatic grapheme to phoneme (G2P) algorithm had to be built and employed to overcome this problem (in order to have complete phonetic dictionaries regardless of the size of the LMs).

A SMT-based approach, similar to the ones presented in [13] and [14], has been adopted for this task. A SMT system generally translates text in a source language into text in a target language. Two components are required for training: a) a parallel corpus consisting of sentences in the source language and their corresponding sentences in the target language, and b) a language model for the target language.

For our specific task (G2P), we consider graphemes (letters) as "words" in the source language and sequences of graphemes (words) as "sentences" in the source language. As for the target language, its "words" are actually phonemes and its "sentences" are actually sequences of phonemes. Note that the natural language is Romanian (for both the graphemic and phonetic representations). Table I lists a few examples.

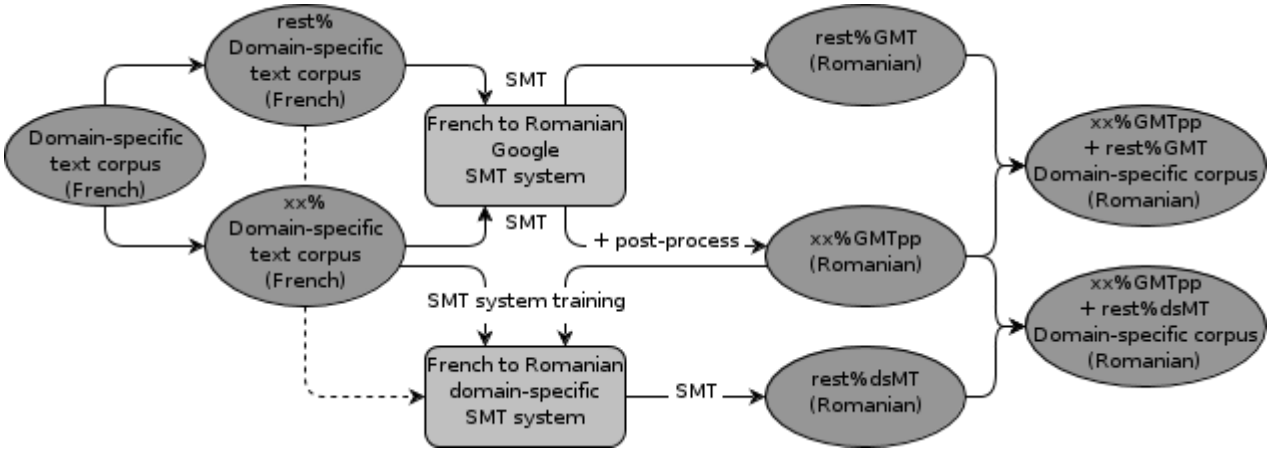

Fig. 2. The two semi-supervised SMT-based language portability methods

TABLE I
EXAMPLES WITHIN THE PHONETIC DICTIONARY (PARALLEL CORPUS)

Ex	Source language (graphemes)	Target language (phonemes)
1	d e z n o d ă m â n t u l	d e z n o d ə m i n t u l
2	a c h i t â n d	a c i t ɨ n d
3	t a p i ț e r i e	t a p i ʦ e r i e

The already available phonetic dictionary is exactly the parallel corpus needed for SMT training. It was randomly split into three parts: a) a training part (580k words), b) an optimization (tuning) part (10k words) and c) an evaluation part (10k words). The same phonetic dictionary, specifically the phonetic representations, serves as training corpus for creating the target language LM.

The implementation of the SMT system is based on the Moses Toolkit [15]. Moses is a widely known toolkit which is mostly used for SMT tasks, but can also solve generic transduction problems as the one presented above.

The training of a G2P translation model is similar to the one of a general translation model, as specified in the Moses documentation. The model's optimization should have been made by minimizing the phone error rate (PER), but this type of optimization module was not available. Therefore, for this process, we chose to use both of the two available methods: a) maximization of the BLEU score [16] (the default in Moses) and b) minimization of the position independent phone error rate (PIPER) [17]. The evaluation of the translation results has been made using the sclite tool in the NIST Scoring Toolkit [18]. Table II lists these results, in terms of BLEU score, phone error rate (PER) and word error rate (WER). Please note that BLEU score is the default evaluation metric for MT systems, but is not suitable for our specific task (G2P). The third system listed in Table II has been further used for all the phonetisations needed in our further experiments.

TABLE II
SMT-BASED G2P RESULTS

Exp	Optimisation	BLEU	PER	WER
1	none	98.89	0.53%	4.79%
2	BLEU	99.49	0.33%	3.24%
3	PIPER	99.39	0.31%	2.76%

Once this G2P system was available, a single phonetic dictionary (97k words) has been created (for all ASR experiments) in the following manner:

a) all the vocabularies for all the LMs have been concatenated creating one single vocabulary,
b) all the words within the vocabulary which were found in the 600k words phonetic dictionary were phonetised using the 600k words phonetic dictionary,
c) all the other words were phonetised using the G2P system.

IV. EXPERIMENTAL SETUP

A. Speech Database and the Acoustic Models

All ASR experiments presented in the next section use the same HMM-based acoustic model. The 36 phonemes in Romanian are contextually modelled with 4000 HMM senones and 16 Gaussian mixtures per senone state [6]. The acoustic model was previously created and optimized (using the CMU Sphinx Toolkit [19]) with a training speech database of about 54 hours of Romanian read speech. This speech database was progressively developed and now comprises isolated words, general newspaper articles and domain-specific (library) dialogues [9]. The texts were recorded by 17 speakers (7 males and 10 females). The phonetic dictionary used in all the ASR experiments is the one described in the previous section.

For the tourism-specific ASR task, the test speech database was obtained as follows: 300 phrases were randomly selected out of the French tourism-specific corpus, manually translated to Romanian and recorded by three speakers. The size of the test database is about 55 minutes. Obviously, the 300 phrases were removed from the French tourism-specific corpus before it was further used for training the language models.

B. Text Corpora and the Language Models

Two text corpora are needed for the ASR experiments in this study: a general Romanian corpus and a domain-specific French corpus (just as presented in Fig. 1).

The domain-specific French corpus comprises tourism specific transcriptions of spontaneous speech. The Google-translated version consists of about 10k phrases summing up to a total of 64k words.

The general Romanian corpus has been acquired using the Web as a resource [6]. This corpus has been subject to various pre-processing operations (among which diacritics restoration). It comprises different types of news and discussions in the European Parliament. It consists of 9.8M phrases summing up to a total of about 169M words.

All the language models used in the ASR experiments are tri-gram, closed-vocabulary language models and have been created using the SRI-LM Toolkit [20]. This toolkit was also used for the interpolation of the general language model with the various domain-specific

language models. The interpolation was systematically done with the weights 0.1 for the general LM and 0.9 for the domain-specific LM because our goal was to create a domain-specific ASR (the domain-specific LM should prevail). The interpolation weights tuning has not been considered for the moment. For the general LM, the number of unigrams had to be limited to the most frequent 64k due to the ASR decoder (Sphinx3) limitation.

C. Domain-Specific SMT System

The second semi-supervised domain adaptation method (discussed in Section II) assumed the existence of a domain-specific SMT system. This SMT system has also been developed with the Moses Toolkit. No optimization has been made due to the small amount of available data. The *xx%* of the French domain-specific corpus and the *xx%* Google translated and then post-processed Romanian corpus were regarded as parallel corpora and were used for training. The same post-processed corpus was also used to create a domain-specific language model (also needed to train the SMT system).

The size of the training corpus had varied from 500 phrases (5% of the domain-specific corpus) to 4000 phrases (40% of the domain-specific corpus). There was no optimization corpus and no test corpus because neither optimization, nor evaluation was performed. The method and consequently the translation system, were evaluated only in the framework of ASR adaptation (see experiments in Section V).

Obviously, this domain-specific SMT system could have been further ameliorated by improving the LM for the target language and/or by optimizing BLEU or some other metrics. These optimizations haven't been performed yet, because our main interest in this study was to validate the methodology.

V. Experimental Results

The evaluation of all the language models was done in terms of perplexity (PPL), out-of-vocabulary (OOV) rate, trigram hits and speech recognition word error rate (WER), all calculated on the test set. Among these four performance figures, the most important is the WER because, in fact, it is the only ASR performance figure. Nevertheless, the other three metrics also lead to important conclusions.

The baseline (the results for the unsupervised method) is presented in Table III. We see that the unsupervised adaptation method produces a domain-specific language model (Exp 0) which is significantly better than the general language model. The interpolation of these two language models issues an even better language model (Exp 100).

TABLE III
Baseline - Unsupervised Scenario Results

Exp	LM	PPL	OOV [%]	3gram hits [%]	WER [%]
-	out-of-domain LM	164.7	4.27	51.0	29.7
0	domain-specific LM	40.8	3.15	31.1	18.7
100	domain-specific LM interpolated with out-of-domain LM	42.5	0.80	55.4	16.2

The results obtained for the semi-supervised adaptation methods are presented in the next tables. Table IV shows the results for the domain-specific language models before interpolation with the general language model, and Table V shows the results after interpolation with the general language model.

The left part of the tables (experiments 1 – 5 and 101 – 105) evaluates the language models created using the first semi-supervised method (as described in Section II.B). This means that the Google SMT system has been used to translate the whole French corpus and *xx%* of the translated corpus has been post-processed. The resulted Romanian corpus has been used to create the language models evaluated in these experiments. We will further refer to these systems as "first-method systems".

The right part of the tables (experiments 6 – 10 and 106 – 110) evaluates the language models created using the second semi-supervised method (as described in Section II.B). The Google SMT system has been used to translate only *xx%* of the French corpus. This part was afterwards post-processed and used to train the domain-specific SMT system. The latter was needed to translate the rest of the French corpus. The resulted Romanian corpus was used to create the language models evaluated in these experiments. We will further refer to these systems as "second-method systems".

TABLE IV
Domain-Specific Language Models Results (before Interpolation with General LM)

		+ rest% GMT						+ rest% dsMT			
Exp	xx%GMTpp	PPL	OOV [%]	3gram hits [%]	WER [%]	Exp	xx%GMTpp	PPL	OOV [%]	3gram hits [%]	WER [%]
0	00%	40.8	3.15	31.1	18.7	0	00%	40.8	3.15	31.1	18.7
1	05%	34.8	2.08	34.0	15.1	6	05%	31.8	6.68	35.3	22.0
2	10%	32.5	1.76	35.2	14.6	7	10%	28.4	3.95	38.4	17.4
3	20%	28.7	1.50	37.9	13.0	8	20%	25.3	2.88	41.2	15.4
4	30%	26.3	1.39	39.4	12.7	9	30%	23.6	2.30	42.1	14.2
5	40%	24.8	1.39	41.3	12.5	10	40%	23.5	1.98	42.7	13.6

TABLE V
Improved Domain-Specific Language Models Results (after Interpolation with General LM)

		+ rest% GMT						+ rest% dsMT			
Exp	xx%GMTpp	PPL	OOV [%]	3gram hits [%]	WER [%]	Exp	xx%GMTpp	PPL	OOV [%]	3gram hits [%]	WER [%]
100	00%	42.5	0.80	55.4	16.2	100	00%	42.5	0.80	55.4	16.2
101	05%	34.4	0.80	56.0	14.6	106	05%	36.3	0.80	58.8	14.2
102	10%	32.4	0.53	56.8	13.9	107	10%	30.1	0.53	58.6	12.7
103	20%	29.0	0.48	57.7	13.1	108	20%	26.7	0.48	59.5	12.6
104	30%	26.6	0.48	58.2	12.4	109	30%	24.3	0.48	59.9	11.6
105	40%	25.2	0.48	59.1	12.2	110	40%	23.8	0.48	60.2	11.5

Please note that experiments 0 and 100 are repeated on both the left and the right side of the tables because they represent the baseline for all the other experiments.

Several conclusions can be drawn given the results in Table IV. First, we observe that even when a small amount such as 5% of the Google translated text was post-processed, all the performance figures are significantly better for the first-method system (Exp 1 compared to Exp 0). On the other hand, the second-method system that uses these 5% displays a significantly higher WER (Exp 6 compared to Exp 0). Even if the trigram hits and perplexity are better, the out-of-vocabulary rate is much worse and it causes the higher WER. This happens because these 5% (500 phrases) are not enough to train a robust SMT system; many words cannot be translated by this system, resulting in a pseudo-Romanian domain-specific corpus, which is clearly not suited for language modelling.

A second important conclusion is that both semi-supervised methodologies issue better and better ASR systems as more machine translated phrases are being post-processed (the only exception is the one presented and explained above). The growth in performance saturates as more and more data is being post-processed.

Comparing the left and the right parts of Table IV, we see that, when the same amount of data is post-processed, the second-method systems systematically display better trigram hits. This means that the newly developed SMT systems produced translations which include some new and useful trigrams. Nevertheless, the WERs for the second-method systems are higher due to the higher OOV rates. The OOV rate problem was already explained: the domain-specific SMT systems can only translate the words found in the small training corpus, leaving the other words in their "French version". On the other hand, the Google MT system is able to adequately translate all the phrases to Romanian.

In conclusion, Table IV states that the first-methodology systems are more robust than the second-methodology systems (when the domain-specific language models are not interpolated with the general language).

Let's take a look now at the results in Table V. First of all, the same trend of lower WERs as more and more phrases are being post-processed can be observed for both semi-supervised methodologies. We observe that after the interpolation (Table V), the OOV rates are equal for the two methodologies (when the same amount of data is post-processed). This means that the lack of coverage which characterized the second-method systems before interpolation (Table IV) has been overcome. Consequently, the second-methodology systems continue to be better in terms of perplexity and trigrams hits, but now outperform the first-method system in terms of WER (thanks to the fewer OOV words).

Comparing the corresponding lines in the two tables, we conclude that, after interpolation, the OOV rates and the trigram hits are much better. Consequently, the WERs are also lower for these ASR systems (the ones which benefit from the large coverage of the general language model).

To conclude this section: when the domain-specific language model, created using the second semi-supervised methodology, is interpolated with a general language model, the relative improvement in WER (for the corresponding ASR system) varies between 12% and 29% depending on the amount of machine translated text that was manually corrected (post-processed). Instead, if the first semi-supervised methodology is used, the relative improvement in WER is generally smaller (10% to 25%).

VI. In-Depth N-Gram Hits Analysis

As shown in the previous section, the improved domain-specific language models have a good ability to predict (55% to 60% trigram hits) both domain-specific words sequences and out-of-domain words sequences (thanks to the interpolation with a general LM). In this work, the general language model was the same for all experiments, so, if we want to answer the question "why and how the proposed methodologies bring improvements in ASR?", we have to analyse the various domain-specific language models *before* interpolation. Table IV showed the results for all these language models. Some of them (Exp 0, 5 and 10) were selected and analysed in Table VI from the point of view of their ability to predict specific words (trigrams example). The selected language models have been created with corpus obtained using the unsupervised methodology (Exp 0), the first semi-supervised methodology (Exp 5) and the second semi-supervised methodology (Exp 10).

Table VI shows seven trigram examples and analyses the way the language models manage to predict the bolded word in the given context. "3-gram" means the LM was able to predict the bolded word in the given trigram context and "2-gram" means the LM needed to back-off to bigrams to predict the bolded word. "1-gram" means the LM needed to back-off to unigrams to predict the bolded word and "OOV" asserts the LM cannot predict the bolded word (it is out-of-vocabulary).

Note that there are trigrams which can be very well predicted by all the analysed language models (type a), but also trigrams that can only be predicted by the domain-specific language models (type b). The importance of the interpolation with the broader, general language model is motivated by its higher trigram hits (51%) and by trigrams which can only be well predicted by it (type c). The plus brought by the semi-supervised methods is revealed by examples of type d.

TABLE VI
N-gram Hits for the General and Domain-Specific Language Models (Examples)

			Trigram examples							
			a	**b**	**b, d**	**c, d**	**c**	**c, d, e**	**f**	**Type**
			o cameră **single**	care acceptă **animale**	într-o locație **liniștită**	puteți să-mi **dați**	şi acum **pentru**	prea scumpă **într-un**	nopți pentru **Belfort**	**Ro text**
Exp	**Language model**	**3-gram hits [%]**	a **single** room	which accepts **animals**	in a **quiet** place	can you **give** me	and now **for**	too expensive **in a**	nights at **Belfort**	**En text**
-	general LM	51.0	3-gram	1-gram	1-gram	3-gram	3-gram	2-gram	OOV	
0	0%GMTpp + 100%GMT	31.1	3-gram	3-gram	2-gram	1-gram	1-gram	1-gram	1-gram	
5	40%GMTpp + 60%GMT	41.3	3-gram	3-gram	3-gram	3-gram	1-gram	2-gram	1-gram	
10	40%GMTpp + 60%dsMT	42.7	3-gram	3-gram	3-gram	3-gram	1-gram	3-gram	1-gram	

Only a few trigrams can be better predicted by the second-method systems, when compared to the first-method systems (see the small difference in trigram hits and examples of type e). And, of course, there are examples of trigrams which cannot be well predicted by any of the analyzed language models (type f). The frequency of occurrence for these six types is difficult to estimate, but the big picture is illustrated by the trigram hits column.

VII. CONCLUSIONS

This study proposed several language portability methodologies to address the absence of domain-specific text resources for a particular language, given domain-specific data in a different language. We have particularly investigated the possibility of porting a tourism-specific French corpus to Romanian with the final goal of creating a tourism-specific ASR system for Romanian. Several SMT-based methods were proposed to create domain-specific language models, which were eventually evaluated in the context of ASR. The baseline methodology used French-Romanian SMT in an unsupervised fashion. Two other semi-supervised methodologies, which benefit from human post-processed data, were introduced and compared with the baseline. The relative improvement in WER brought by the semi-supervised methods varies from 10% to 29%, depending on the amount of machine translated text that was manually corrected (post-processed).

The two semi-supervised SMT-based domain-adaptation methods represent the novelty of this paper. Several other studies ([6], [7], [8]) have used machine translation output for domain-adaptation, but only in an unsupervised fashion. In this work we have showed that human intervention (error correction) is very effective even if only a small amount of data is corrected. Moreover, if just a small part of the machine translated output is subject to correction then the human intervention is practically inexpensive, while the boost in performance is significant.

In the same context of ASR, SMT principles and methodologies were also used to extend the pronunciation dictionary. Although the use of SMT for G2P is not a new idea [13], the good results obtained for Romanian (0.31% PER) are worth mentioning. This G2P conversion method is presented in this paper with the sole purpose of outlining another application of SMT in the context of ASR. Consequently, we do not make any comparison with other, more elaborated G2P conversion methods such as the one presented in [21].

A more in-depth analysis, explaining the reasons why the proposed methodologies bring improvements in ASR, was also made and several examples of trigrams were given to illustrate the various language prediction scenarios.

Pragmatically, this study summarized the needed resources and proposed an SMT-based methodology, which could be used to develop a domain-specific ASR system for any under-resourced language, given that specific resources are available for a high-resourced language.

On the short term, the results presented in this study could be improved by tuning the domain-specific SMT system and the LM interpolation weights. A possible improvement could also be obtained by combining the semi-supervised methods.

On the long term, we plan to further validate the adaptation methodology by applying it for other specific domains and also for other pairs of source-target languages. Another interesting perspective would be the usage of the proposed methodology when domain-specific data is available in more than one high-resourced (source) languages.

ACKNOWLEDGMENT

The research reported in this paper was funded by the Sectoral Operational Program Human Resources Development 2007-2013 of the Romanian Ministry of Labor, Family and Social Protection through the Financial Agreement POSDRU/6/1.5/S/16.

REFERENCES

[1] N. Abdillahi, P. Nocera, and J.F. Bonastre, "Automatic transcription of Somali language," in *Proc. of ICSLP2006*, 2006, pp. 289-292.

[2] T. Pellegrini, and L. Lamel, "Investigating Automatic Decomposition for ASR in Less Represented Languages," in *Proc. of ICSLP2006*, 2006.

[3] P. Mihajlik, T. Fegyó, Z. Tüske, P. Ircing, "A Morpho-graphemic Approach for the Recognition of Spontaneous Speech in Agglutinative Languages – like Hungarian," in *Proc. of Interspeech2007*, 2007.

[4] V.B. Le, L. Besacier, "Automatic Speech Recognition for Under-Resourced Languages: App. to Vietnamese Language," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 8, pp. 1471-1482, 2009.

[5] B. Jabaian, L. Besacier and F. Lefevre, "Combination of Stochastic Understanding and Machine Translation Systems for Language Portability of Dialogue Systems," in *Proc. of ICASSP 2011*, 2011, p. 5612.

[6] H. Cucu, L. Besacier, C. Burileanu, A. Buzo, "Enhancing Automatic Speech Recognition for Romanian by Using Machine Translated and Web-based Text Corpora," in *Proc. of SPECOM2011* (to appear), 2011.

[7] A. Jensson, K. Iwano, S. Furui, "Development of a speech recognition system for Icelandic using machine translated text," in *Proc. of SLTU2008*, 2008.

[8] H. Nakajima, H. Yamamoto, T. Watanabe, "Language Model Adaptation with Additional Text Generated by Machine Translation," in *Proc. of COLING*, 2002, vol. 2, pp. 716-722.

[9] H. Cucu, A. Buzo, C. Burileanu, "Optimization Methods for Large Vocabulary, Isolated Words Recognition in Romanian Language," *UPB Scientific Bulletin*, series C, vol. 73, no. 2, pp. 179-192, 2011.

[10] E. Oancea, I. Gavăt, O. Dumitru, D. Munteanu, "Continuous Speech Recognition for Romanian Language Based on Context Dependent Modeling," in *Proc. of International IEEE Conference „COMMUNICATIONS 2004"*, 2004, pp. 221-224.

[11] D. Militaru, I. Gavăt, O. Dumitru, T. Zaharia, S. Segărceanu, "Protologos, System for Romanian Language Automatic Speech Recognition and Understanding," in *Proc. of SPED09*, 2009, pp. 21-32.

[12] C. Ungurean, D. Burileanu, V. Popescu, C. Negrescu, A. Dervis, "Automatic Diacritics Restoration for a TTS-based E-mail Reader Application," *UPB Scientific Bulletin*, series C, vol. 70, no. 4, 2008.

[13] A. Laurent, P. Deléglise, S. Meignier, "Grapheme to phoneme conversion using an SMT system," in *Proc. of Interspeech09*, 2009, p. 708.

[14] P. Karanasou and L. Lamel, "Comparing SMT Methods for Automatic Generation of Pronunciation Variants", in *Proc. of IceTAL2010*, 2010, p. 167.

[15] (2011) The Moses Toolkit website. [Online]. Available: http://www.statmt.org/moses/

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

[17] N. Bertoldi, B. Haddow, and J.-B. Fouet, "Improved Minimum Error Rate Training in Moses," *The Prague Bulletin of Mathematical Linguistics*, pp. 1-11, Feb. 2009.

[18] (2011) The NIST Scoring Toolkit (SCTK) website. [Online]. Available: ftp://jaguar.ncsl.nist.gov/current_docs/sctk/doc/sctk.htm

[19] (2011) The CMU-Sphinx Speech Recognition Toolkit website. [Online]. Available: http://cmusphinx.sourceforge.net

[20] (2011) The SRI-LM Language Modeling Toolkit website. [Online]. Available: http://www-speech.sri.com/projects/srilm

[21] M. Bisani and H. Ney, "Joint-sequence models for grapheme-to-phoneme conversion," Speech Communications, vol. 50, no. 5, pp. 434-451, May 2008.