Using the World Wide Web for Learning New Words in Continuous Speech Recognition Tasks: Two Case Studies

Stanislas Oger, Vladimir Popescu, Georges Linarès

Laboratoire Informatique d'Avignon, University of Avignon, France
{stanislas.oger, vladimir.popescu, georges.linares}@univ-avignon.fr

Abstract

In this paper, Web-based lexicon augmentation is addressed: using various strategies for Out-Of-Vocabulary (OOV) word learning, we discuss their relevance in two types of applications: broadcast news, or topic-specific corpora transcription. The Web-based OOV word learning is first tested on the French news corpus ESTER; the same approach is applied to a very specific corpus concerning surgical interventions. These tests allow us to assess the value of the OOV word learning methods proposed, emphasizing their strengths and weaknesses, regarding the particular type of application considered.

1. Introduction

The Web constitutes a huge and continuously updated collection of data that could be useful in language processing tasks; in the past, it was considered a too imperfect corpus for knowledge learning, because the documents available are too noisy or too difficult to access (the search engines were scarce and not reliable) [1]. However, since recently this view changed, first in the Natural language processing community – where corpora collected from the Web began to be used for training part-of-speech (POS) taggers, morphological analyzers, etc. – and then in the Automatic Speech Recognition (ASR) field, where Web-collected corpora began to be used for language modeling purposes. However, large well-targeted corpora generally outperform Web-based language models in language processing tasks [2].

More specifically, for ASR purposes, the Web can be used to adapt the general-purpose language models of the speech decoders to topic-specific corpora. Most of the proposed approaches consist in collecting Web data to estimate augmented lexicons and relevant language models [3], [4].

Concerning lexicon augmentation, an important issue resides in the need to formulate queries which are fed into a search engine, in order to retrieve relevant documents. The way the words are chosen in the retrieved documents for lexical augmentation is essential, in order to ensure a reasonable word retrieval precision [5], [6], [7].

Focusing on these issues, a method for learning new words is presented in [5] and relies on the idea that the local word context of the OOV words can provide valuable clues concerning the characteristics of these OOV words. This information is used for retrieving the OOV words in the unlimited set of Web documents. This approach allows a better precision of the lexicon augmentation.

Another Web-based method for lexicon augmentation is proposed in [7], where the authors extract keywords from *entire* documents, in order to query the Web for data (accessed *via* Web pages) relevant to those documents. On the contrary, the approach proposed in [5] relies on the use of the *local* information in order to extract Web document excerpts. Therefore, the two approaches use the Web in similar ways for dealing with the same issue, augmenting vocabularies for reducing OOV word rates. However, the fact that in [7] entire documents are used yields, in principle (actual results in this respect are not reported in the paper), an important augmentation in the number of words in the lexicon. This leads to a lower precision in lexicon augmentation, which determines the generation of relatively noisier language models. Local methods should ensure a better precision, since only document excerpts that closely match the context of the OOV words are selected.

In this paper, we assess the method proposed in [5] for new words learning in several continuous speech recognition application types. More specifically, an analysis (emphasizing the strengths and weak points) of our lexical augmentation method will be performed on two different application scenarios: (a) broadcast news diarization, on the ESTER corpus, and (b) topic-specific corpora indexing, on a collection of audio recordings concerning surgical interventions (the AVISON corpus)[1].

A summary of the approach that we have proposed in [5] is presented in the first section. In the second section we describe the experimental configurations used for assessing the lexicon augmentation method; especially, the particular AVISON corpus is described in detail. Moreover, a particular method for extracting a topic-specific sub-corpus is proposed. In the third section, we present the performance figures concerning the OOV word coverage induced by our lexical augmentation method, as well as the effects on the recognition rates. A set of final remarks end up the paper.

2. Background

The goal of this paper is to assess the method presented in [5] for learning new words, in two different continuous speech recognition application types. Here is a summary of this method.

2.1. Overview

The problem we address here is how to automatically retrieve OOV words which have not been correctly transcribed by an automatic speech recognition engine, given that the automatic speech recognition engine is able to tag the erroneous OOV words in the automatic transcription.

Our approach makes the hypothesis that the local context of

[1]The AVISON corpus was provided by the IRCAD Institute (`http://www.ircad.fr`).

these OOV words in the transcription holds some information characteristic of the related OOV words. This information can be used to retrieve the words in a large text database. Moreover, a database large enough to be suitable for this application is hard to build and, above all, would have to be often updated in the case of transcribing news data, where topics are frequently unexpected and thus involve a high number of new words. This observation led us to consider the use of the World Wide Web as a large database, which is constantly updated. Given that base reflection, we developed a two-step approach:

1. We first use the context words of each OOV word to formulate queries, which are submitted to a search engine. For each OOV word, we query the search engine and take the top returned documents (usually the top 100). The candidate OOV words are then extracted from the documents depending on the way the queries are formulated. In [5] four query building strategies have been proposed.
2. The candidate OOV words are then inserted in the baseline language model and the speech segments that contained the OOV words (according to that baseline language model) are decoded again with the augmented language model. In [5] two strategies for inserting new words have been proposed; these methods are described in the sequel.

These two steps are described in more detail in Section 2.2.

2.2. Query Formulation

Several strategies have been proposed in [5] for formulating queries in order to extract relevant OOV words from Web documents.

The first strategy is called *n-gram strategy*. For each OOV word, it consists in extracting the n-gram containing the OOV word and substituting it by a wildcard character. For example, for the sentence "$w_1 w_2 w_3 w_4$" with w_4 the OOV word, the corresponding 3-gram query is "$w_2 w_3 (*)$". The candidate word lists are built with the words that replace the wildcard character $(*)$ in the retrieved documents.

The second strategy is called *pattern strategy*. Given a sentence containing an OOV word, first the function words (e.g. "what", "so", etc.) are suppressed, then wildcard characters are inserted between each word, and finally the OOV word is also replaced by a wildcard character. For example, for the sentence "$w_1 w_2 w_3 w_4$", with w_4 the OOV word and w_2 the function word, the corresponding pattern query of size 3 is "$w_1 * w_3 (*)$". The candidates are extracted by taking the words that replace the wildcard character between brackets in the retrieved documents.

The third strategy is called *semantic strategy*. First, the words contained in a small-sized window (between 5 and 20 words) around the OOV word are extracted. Then, they are sorted in decreasing order of frequency in the corpus where the search is being performed (here, the Web). Lastly, we take the first n words in this list, in order to build a query of size n. The query is processed as a bag of words, so that it allows for permutations of the keywords in the documents concerned. The candidates are all the unknown words found in the documents returned by the search engine.

The fourth strategy, called *semantics-driven strategy*, is a mixture between the first and the third. The principle is to build n-gram queries like in the first strategy and to add some *drive words*, which are relevant words extracted from the context. The drive words are added to the query without ordering constraints. For example, for the sentence "$w_1 w_2 w_3 w_4$" with w_4 the OOV word and w_2 the function word, the corresponding query of size 3 with 1 drive word is "$w_2 w_3 (*)'' + w_1$. The candidate word lists are built in the same way as with the n-gram strategy.

By using one of these techniques, we are able to retrieve for each erroneous OOV word a list of candidate words. In order to assess these strategies, we can measure the recall and the size of these lists of candidates; this is done in Section 4.2.

2.3. Dealing with New Words

Each of the previously-presented strategies delivers a set of candidate word lists, one list per OOV word. Then, one strategy is *a priori* chosen; we thus obtain *one* candidate word list for each OOV word. These lists (for all the OOV words) have to be inserted in the transcription. The insertion of the new words is accomplished by a second automatic transcription pass with the augmented lexicon. This approach has the advantage that it lets the decoding process automatically choose the correct word in the candidate list. This is done by automatically phonetizing the candidates (we used Festival [8] for the AVISON English corpus and LIA_PHON [9] for the ESTER french corpus) and inserting them in the language model recognition vocabulary. The probability given to these new words in the language model is a key point that led us to imagine two strategies.

The first strategy consists of giving to the new words the probability of the unknown word. This is the probability of occurrence of a word member of the unknown word class (actually, an OOV word) in the training corpus.

The second strategy consists in giving to the new words the probability of their part-of-speech class. This is the probability of occurrence of an unknown word of a given part-of-speech class in the training corpus. This is a finer probability than the previous one.

3. Experimental Configurations

The lexicon augmentation method described in the previous section is tested on two different application scenarios: (i) a broadcast news transcription task, using the ESTER corpus of spoken French news [10], and (ii) a topic-specific corpus transcription task, using the AVISON corpus of spoken documents related to surgical interventions. We thus hope to assess our lexicon augmentation method on two representative (and limit) cases of the transcription task: broadcast news and domain-specific speech material.

3.1. The Broadcast News Transcription Task

The experiments carried out on the broadcast news have already been presented in [5]. Thus, we used around 6 hours of French broadcast news in the ESTER 2005 corpus. The OOV word rate, with respect to a 65k baseline vocabulary, is around 1.03 %, out of the 62k words in the ESTER corpus. Most of them are named entities or technical words.

Hence, our first task comes to transcribing evenly-balanced multi-domain news corpora (containing spontaneous speech). In Section 4 we show how our approach behaves on this task.

3.2. Topic-Specific Transcription Task

The second test consists in transcribing a highly specialized domain-constrained speech corpus, AVISON.

3.2.1. The AVISON Corpus

The AVISON corpus contains around 20 hours of commented English surgical intervention films. The acoustic data are echoed in part by stenographic transcriptions, used for building a test corpus.

The spoken material in this corpus contains speech in several registers: read speech documenting surgical issues, spontaneous descriptions of surgical interventions, or spoken dialogues between surgeons and students.

For assessing the benefits of the lexicon augmentation approach on domain-specific corpora, there are at least two options: (i) using all the AVISON corpus, and (ii) extracting a topic-specific corpus, out of AVISON data. The first option would lead us to evaluate our method on domain-specific, but multi-topic data, which have a narrower purview than broadcast news data; nevertheless, we can go further and assess our OOV word retrieval method on even narrower scope data, by selecting a *topic*-specific sub-corpus, out of domain-specific AVISON data. We thus consider a sub-domain for simulating a case where one has a small corpus, and *new* data occur. These data are very specific, even in the medical domain. A short presentation of the method that we developed for extracting the topic-specific sub-corpus is presented in Section 3.2.2.

3.2.2. Choosing a Topic-Specific Sub-Corpus

In order to choose a topic-specific sub-corpus, we used an OOV word-based document selection. The principle of the method is to select a set of documents from the database, which maximize the number of OOV words with respect to the documents that are not in the set. This relies on the fact that the more OOV words are in the selected documents, the more they have a specific topic. This procedure is composed of four-step:

1. First, a term-frequency-inverse-document-frequency (tf-idf, [11]) based keyword extraction from the documents in the database is done. According to the interpretation of tf-idf, the obtained keywords are characteristic of certain documents with respect to the rest of the database. Thus, we expect to collect topic-specific keywords.

2. Then, bag-of-words are randomly built by associating up to eight of these keywords.

3. A sub-corpus is then built for each bag-of-words with the documents that contain all the words in the bag-of-words considered.

4. Finally, the sub-corpus that contains the most OOV words with respect to the rest of the database is considered as the targeted topic-specific sub-corpus.

We were thus able to choose a sub-corpus of around 2 hours of speech including prepared and disfluent speech. In this 13k words corpus, the OOV rate is of around 6 %, with respect to our 65k word baseline recognition vocabulary; most of the OOV words are domain-specific.

4. Performance Assessment

4.1. System Configurations

In order to assess the lexicon augmentation methods described in Section 2 for the two transcription tasks presented in Section 3, we used the LIA continuous speech decoder *Speeral*, [12].

The system was fine-tuned for the English language by using the Festival phonetizer [8], for the English words. We used a combined language model, by interpolating general 3-grams learned on the HUB4 English corpus [13], with 3-grams learned on all the reference transcriptions available in the AVISON training corpus. This processing step proved to be essential for ensuring a baseline word error rate (WER) of around 48 % on the topic-specific AVISON sub-corpus. For the French ESTER corpus, in the broadcast news diarization task, such a step was not necessary, and a baseline word error rate of around 24 % was obtained.

The high WER obtained on the AVISON corpus can be explained by two facts. First, it is known that each OOV word produces about two errors on the transcription [14], thus the 6 % of OOV words are responsible of about 12 % of the transcription errors. Then, the English language model is learned with the HUB4 news corpus, and interpolated with a language model learned on the only 40k words of the AVISON textual corpus. It is clear that the language model is not suited to the task, but it is the best model that can be learned given the corpora. This scenario pertains to the situation where one wants to automatically transcribe a domain-specific corpus in order to index it. In this case, our approach is interesting because it allows us to retrieve OOV words that are relevant for indexing.

The OOV words are manually indicated in the transcription in order to simulate a perfect automatic OOV word detection process. A completely automatic procedure can be used, for example by using the work reported in [15] or [16].

The performance of the OOV word retrieval strategies are obtained with the Google search engine. For each OOV word, a query is generated, submitted to Google and the top 100 returned documents are used to build a candidate list.

4.2. Evaluations and Discussion

A first relevant performance measure for our vocabulary expansion method is the n-gram recall on the Web. In other words, the recall is the percentage of the n-grams containing OOV words in the test corpus, that are found on the Web. For example, for $n = 3$, queries such as $''w_1 w_2 w_o''$, with w_o the OOV word, are submitted to a search engine. If at least one document is returned, the n-gram is considered as existing on the Web. Thus, for the two scenarios considered in this article, the percentages of n-grams containing OOV words retrieved by the Google search engine are shown in Table 1.

n-gram order	1	2	3	4	5
Recall [%] ESTER	100	88.2	50.5	27.3	16.1
Recall [%] AVISON	100	98.8	87.1	56.6	32.0

Table 1: Recall of n-grams containing OOV words on the Web

From Table 1, we can see that more n-grams are found on the Web for the *domain-constrained* corpus (AVISON), than for the *domain-independent* broadcast news corpus (ESTER). In

n	ESTER						AVISON					
	n-gram strategy		pattern strategy		semantic strategy		n-gram strategy		pattern strategy		semantic strategy	
	REF	ASR	REF	ASR	REF	ASR	REF	ASR	REF	ASR	REF	ASR
2	14.0	4.7	20.0	7.3	32.6	18.5	18.1	15.1	21.1	15.3	68.5	55.4
3	18.1	5.1	20.3	5.0	39.7	27.8	20.5	11.6	18.7	9.9	61.2	44.4
4	16.4	2.3	17.5	2.0	45.9	35.2	21.4	9.2	6.5	3.8	55.6	35.1
5	13.8	1.9	12.3	1.2	50.2	40.9	14.1	4.1	1.5	1.0	23.5	24.1

Table 2: Recall [%] of OOV word retrieval on the best 100 Google ranked documents

n	ESTER						AVISON					
	n-gram strategy		pattern strategy		semantic strategy		n-gram strategy		pattern strategy		semantic strategy	
	REF	ASR	REF	ASR	REF	ASR	REF	ASR	REF	ASR	REF	ASR
2	145	322	411	475	16.0k	13.7k	526	439	356	309	15.9k	28.7k
3	49	207	139	166	19.0k	38.1k	147	108	88	70	30.2k	23.3k
4	13	34	34	21	37.9k	42.6k	62	38	18	16	33.6k	44.0k
5	4	9	15	8	44.9k	45.0k	25	13	8	9	20.3k	26.2k

Table 3: Average hypothesis-set size of OOV word retrieval on the best 100 Google ranked documents

n/m	ESTER				AVISON			
	REF		ASR		REF		ASR	
	recall	sets size	recall	sets size	recall	sets size	recall	sets size
2/1	24.0	268	8.7	292	41.8	1.3k	29.1	1.2k
2/2	26.1	789	8.1	306	57.8	3.9k	37.4	4.1k
2/3	27.0	1.3k	6.5	295	60.5	5.8k	38.1	6.2k
3/1	19.1	16	4.0	87	31.1	215	13.8	129
3/2	15.0	15	3.9	79	33.1	379	13.9	209
3/3	13.3	19	3.1	98	32.7	500	13.8	255

Table 4: Recall [%] and hypothesis-set size of OOV word retrieval, for the semantics-driven n-gram strategy with m drive words

other words, the relatively higher recall values that the Web exhibits for the surgical domain corpus AVISON can be explained by the fact that technical discourse is expected to be linguistically more stable than spontaneous or non-technical discourse encountered in news. We call here linguistic stability the fact that given a non-function word w and a corpus c, a few sentences constitute the context of the most of the occurrences of w in c. The more stable a database is, the less non-function words are found in different contexts and the more n-grams are easily found in other related databases, for example the Web.

However, the most interesting performance figures, when OOV word retrieval methods are to be evaluated, consist in the recall of a query building strategy, on Web-retrieved documents. In other words, the normalized number of OOV words that are found in the retrieved documents given a query strategy. The average size of the candidate sets is also presented, which is related to a precision measure. The larger the candidate sets are, the lower the precision is.

The retrieval strategies previously described in Section 2 are compared on the two corpora considered. The recall is evaluated with respect to both reference transcripts and ASR system outputs.

Thus, for the two tasks considered, we measure the performance of the candidate word sets built using queries formed according to three of the four strategies presented in the section 2.2: (i) building "hard" queries consisting of word n-grams, (ii) building "soft" queries, by extracting word patterns from the context, and (iii) building "semantic" queries by selecting relevant context words without ordering constraints. In Tables 2 and 3 we show recall (in percents) and "precision" (in number of words) for the OOV word retrieval methods proposed.

Obviously, the recall rate is worse with ASR outputs than with reference transcripts. This is explained by the transcription errors that are more frequent around the OOV words and thus provide erroneous contexts for query building.

Moreover, we see that recall rates are slightly higher for the domain-specific corpus (AVISON), than for the news, when n-gram-based and pattern-based strategies are used. This can be explained by the technical nature of the domain concerned by the AVISON corpus: the syntactic structures are highly constrained. In most of the sentences the functionally-related words (e.g. "forceps", "scissors", etc.) are highly substitutable; it is thus easier to retrieve the OOV words knowing their exact context. Hence, the pattern-based method, which exhibits better performance when compared to the n-gram-based retrieval on the ESTER news corpus, provides lower recall scores than the n-grams, on the AVISON corpus. This indicates that relaxing local syntactic constraints by using the pattern-based strategy is not essential on the AVISON corpus. This last remark confirms the observations we made about the figures in Table 1, that the AVISON corpus is more linguistically-stable than the ESTER corpus.

Concerning the "semantic" retrieval strategy, we first remark that the recall worsens for the AVISON corpus when query sizes increase, whereas for the ESTER corpus, an inverse effect is obtained. This is explained by the highly specialized nature of the domain concerned by the AVISON corpus and the fact that the queries are made of relevant context words: the longer a query is, the less likely one is to find a *relevant* document that contains all the keywords. On the other hand, "precision" (in the sense pointed out above) decreases when the query size increases on the AVISON corpus. This can be explained by the same specialized nature of this corpus: documents that match large queries are likely to be very long and not necessarily relevant. In fact, the recall values exhibit an interesting evolution, with respect to the specialization of the queries: from Table 2 we see that the recall depends non-monotonically on n, the size of the query. There is a value for n up to which the recall increases (4 for ESTER and 3 for AVISON) because the more relevant words are added to the queries, the more precisely they target the appropriate documents. After that, for greater query sizes, the recall decreases, because there are less documents that match the query and thus OOV words are lost.

We present in Table 4 the "precision" and recall for the driven n-gram OOV word retrieval strategy, the fourth presented in Section 2.2. This strategy consists in formulating n-gram queries that contain an ordered set of words, along with a set of drive words that semantically customize the query. As expected, this method improves the recall values for *both* corpora. Nevertheless, the performance gains are more important when the queries are shorter. This can be explained by the fact that over-specializing the queries results in a low amount of matching documents. Moreover, for the domain-specific AVISON corpus the recall improvements are much more important than for the broadcast news corpus. This improvement in the recall value is consistent with the relatively higher stability of the linguistic constructs in AVISON. However, "precision" worsens with the increase of the size of the queries in all cases. This is expected as well, since longer documents are more likely to match these specialized queries.

4.3. Decoding with augmented lexicon

This experiment consists in inserting the new words obtained with one of the OOV word retrieval strategies, in the recognition lexicon and perform a second ASR pass on the segments containing OOV words.

The baseline lexicon is augmented with the driven n-gram strategy (configuration $n/m = 2/1$ in Table 4) with an absolute recall of 29.1 %. We integrate the augmented lexicons in the ASR system by using the POS-based method described in Section 4.3. In Table 5 we show the effects of the lexicon augmentation methods on ASR performance, in terms of WER, precision and recall. Recall and precision bear different significances with respect to the previous discussion. Thus, the recall is the absolute ratio between the transcription OOV words that are successfully recovered in the final (lexicon augmented) transcription, and the total number of OOV words. The precision is the ratio between correct OOV words, and the total number of inserted new words in the final transcription.

Corpus	Recall	Precision	WER	base WER
ESTER	6.05	55.07	24.3	24.5
AVISON	28.7	49.9	47.7	48.7

Table 5: Recognizer precision [%], recall [%], WER [%] and baseline WER [%] with POS-based integration of augmented lexicon, on ESTER and AVISON corpora

The recall indicates that 29.1 % of the OOV words are recovered in the final transcription. This represents a potential gain of about 2 % WER (29.1 % of 6 % of OOV words). Moreover, a precision of about 50 % indicates that about as many wrong new words as correct new words are introduced by the ASR. This represents about 2 % of the words. However, the WER decreases 1 % in absolute value. This means that the lack of precision is strongly balanced by the correction of the erroneous zone around the retrieved OOV words. It is worth emphasizing that the new word insertion recall is about 98 %: 29 % of OOV words are in the augmented lexicon and 28.7 % are introduced in the final transcription.

5. Conclusions and Prospects

In this paper we have assessed several Web-based techniques that rely on word template search. First, we have described a framework for using the Web in order to augment lexicons with OOV words. Then, we have studied two prototypical application scenarios: broadcast news and specialized-domain audio data transcription. The results show the effectiveness of the proposed approaches, in both generic and topic-dependent tasks. Nevertheless, as expected, the proposed approaches are more efficient in domain-constrained tasks, especially in terms of recall gains. These gains are more significant for the domain-specific corpus, due to its linguistic regularity. The method that combines word sequences with additional drive words in queries provides the best recall for both corpora, especially for the domain-constrained data. These results are confirmed by the WER gains obtained on recognition experiments; nevertheless, rather than accuracy gains, the transcription intelligibility should be improved by the retrieval of meaningful OOV words, such as named entities, which are crucial in language understanding tasks.

We should however notice that adding drive words to the pattern-based queries does not seem to bring a robust recall improvement in both scenarios: whereas gains are indeed obtained in transcribing generic data (the ESTER corpus), those are apparently less compelling in domain-constrained contexts (the AVISON corpus), because the queries become too specialized in this latter case.

These results validate the initial idea that the short-term context holds some characteristic information about missing words. Moreover, most of the retrieved words are highly meaningful OOV terms, such as named entities. These terms are crucial in language understanding tasks, and the proposed methods should bring significant improvements to ASR-based speech mining systems [17].

In the near future we plan to extrapolate this method to language model adaptation by considering the use of Web data as an alternative strategy for n-gram estimation on low resourced domains.

6. Acknowledgements

This research has been partly funded by the National Research Authority (ANR), AVISON project (ANR-007-014). The authors wish to express their thanks for the support. The authors are also grateful to the IRCAD Institute for having provided the AVISON speech corpus.

7. References

[1] W. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[2] M. Lapata and F. Keller, "Web-based models for natural language processing," in *ACM Transaction Speech Language Processing*, 2005.

[3] A. Allauzen and J. Gauvain, "Open Vocabulary ASR for Audiovisual Document Indexation," in *Proceedings of the International Conference on Acoustics, Speech and Language Processing*, 2005.

[4] N. Bertoldi and M. Federico, "Lexicon adaptation for broadcast news transcription," in *Proceedings of ISCA ITRW Workshop on Adaptation Methods for Speech Recognition*, 2001.

[5] S. Oger, G. Linarès, F. Béchet, and P. Nocera, "On-demand new word learning using the world wide web," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 2008.

[6] T. Hain, J. Dines, G. Garau, M. Karafiat, D. Moore, V. Wan, R. Ordelman, and S. Renals, "Transcription of conference room meetings: an investigation," in *Proceedings of Interspeech*, 2005.

[7] H. Yu, T. Tomokiyo, Z. Wang, and A. Waibel, "New developments in automatic meeting transcription," in *Proceedings of the International Conference on Speech and Language Processing*, 2000.

[8] P. Taylor, A. Black, and R. Caley, "The architecture of the festival speech synthesis system," in *Proceedings of the third ESCA Workshop in Speech Synthesis*, 1998.

[9] F. Béchet, "Lia_phon: Un système complet de phonétisation de textes," *Traitement Automatique des Langues*, vol. 42, no. 1, pp. 47–67, 2001.

[10] G. Gravier, J. Bonastre, S. Galliano, and E. Geoffrois, "The ester evaluation campaign of rich transcription of french broadcast news," in *Proceedings of Language Resources and Evaluation Conference*, 2004.

[11] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.

[12] P. Nocera, C. Fredouille, G. Linares, D. Matrouf, S. Meignier, J. Bonastre, D. Massonié, and F. Béchet, "The LIA's French Broadcast News Transcription System," in *SWIM: Lectures by Masters in Speech Processing*, 2004.

[13] R. Stern, "Specifications of the 1996 hub-4 broadcast news evaluation," in *Proceedings of the DARPA Speech Recognition Workshop*, 1997.

[14] J. Gauvain, L. Lamel, and M. Adda-Decker, "Developments in continuous speech dictation using the ARPA WSJ task," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1995.

[15] L. Burget, P. Schwarz, P. Matejka, M. Hannemann, A. Rastrow, C. White, S. Khudanpur, H. Hermansky, and J. Cernocky, "Combination of strongly and weakly constrained recognizers for reliable detection of oovs," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 2008.

[16] H. Lin, J. Bilmes, D. Vergyri, and K. Kirchhoff, "OOV detection by joint word/phone lattice alignment," in *Proceedings of the IEEE Workshop Automatic Speech Recognition & Understanding*, 2007.

[17] R. De Mori, F. Bechet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur, "Spoken language understanding," *IEEE Signal Processing Magazine*, vol. 25, pp. 50–58, 2008.