Rapid Language Model Development for New Task Domains

Lucian Galescu, Eric Ringger, James Allen

Department of Computer Science
University of Rochester
Rochester, NY 14627 USA
{galescu, ringger, james}@cs.rochester.edu
http://www.cs.rochester.edu/research/trains

Abstract

Data sparseness has been regularly indicted as the primary problem in statistical language modelling. We go one step further to consider the situation when no text data is available for the target domain. We present two techniques for building efficient language models quickly for new domains. The first technique is based on using a context-free grammar to generate a corpus of word collocations. The second is an adaptation technique based on using out-of-domain corpora to estimate target domain language models. We report results of successfully using these two techniques individually and in combination to build efficient models for a spontaneous speech recognition task in a medium-sized vocabulary domain.

1. Introduction

Language models for continuous speech recognition are usually built from a large set of training sentences in a specific domain. However, many authors have pointed out the difficulty of getting a sufficiently large textual corpus. Consequently, there has been a lot of recent research on domain adaptation, that is, estimating domain-dependent models by using out-of-domain data.

Most approaches to adaptation (Rudnicky, 1995; Crespo *et al.*, 1997; Ito *et al.*, 1997) assume the availability of a good general model that can be made more specific by incorporating knowledge from a text corpus in a new domain. Other approaches (Iyer, Ostendorf & Gish, 1997; Iyer & Ostendorf, 1997; Witschel & Höge, 1997) perform adaptation from text data in other, different domains by trying to pick the "relevant" information, which is then combined with information from text data in the new domain. In contrast, we propose two techniques for domain adaptation when no corpus is available for the new domain. We also show their successful application on a language modelling task for a medium vocabulary domain.

The first technique proposed here is based on using a context-free grammar (CFG) to generate word n-grams. The second one is based on using out-of-domain corpora. These techniques can be used separately or in conjunction with any other technique, allowing for fast and inexpensive prototyping of low perplexity and good recognition accuracy language models.

2. Generating Artificial Corpora

When trying to build language models (LMs) for new domains for which no data is available, it is customary to use a "wizard of Oz" procedure to obtain a small text corpus (a few hundred to a few thousand utterances) for the new domain, generate a language model from this corpus, and (perhaps) interpolate it with a more general model. Rapid prototyping and incremental adaptation can be done by taking a bootstrapping approach: first a reduced set of sentences is used for adaptation, and then the LM is adapted incrementally as more in-domain sentences become available.

We propose an alternative method for obtaining good performance in very short time: build a CFG for the new domain and use it to generate an n-gram language model. The idea dates back to 1991, but it hasn't received much attention in the community. It was first used in the development of the VOYAGER system (Zue *et al.*, 1991), by employing the TINA parser in generation mode to obtain an artificial corpus. We took a different approach, closer to Jurafsky *et al.* (1994) and Popovici & Baggia (1997); we hand-coded a task-specific CFG from which we generated the artificial corpus, and obtained an n-gram LM from the text data in this corpus. Jurafsky *et al.* (1995) and other authors suggest using a (probabilistic) CFG as a language model, either stand-alone or in combination with a statistical LM. The grammar developed according to the methodology shown in this paper is meant only to be a source of realistic word collocations, and it would be too constraining and brittle if used as a LM. The advantage of our technique is that it requires less time and expertise, and doesn't need parameter tuning either by expensive hand-crafting or by corpus-based learning (we assume no target domain corpus is available).

We continue by describing the process of obtaining a CFG for a specific domain, with examples from the PACIFICA domain [1].

1. First we wrote down about one hundred sentences (about 800 words) of the type that would be likely to appear in conversations with the system, given its functionality. E.g.,

[1] We used the TRIPS dialog system (Ferguson & Allen, 1998) as a testbed for our experiments.

USE ONE TRUCK TO MOVE ALL THE PEOPLE FROM ABYSS
HOW LONG DOES A HELICOPTER TAKE TO GET TO EXODUS
RESCUE THE PEOPLE AT EXODUS
HOW LONG DOES IT TAKE TO GET TO DELTA BY TRUCK
GO BACK TO ABYSS
HOW CAN I GET TO BARNACLE
CAN I FLY OVER THE FOREST

Starting from these, we hand-coded grammar rules for the TRIPS domain, following a methodology similar to the one developed in (Rayner & Carter, 1997) for semi-automatic adaptation of grammars. The overall procedure can be summarized in the following three steps:

2. Divide sentences into smaller parts which can be reasonably thought of as units that may be replaced by other words or phrases related to the task domain. E.g.,

USE / ONE TRUCK / TO / MOVE / ALL THE PEOPLE / FROM ABYSS
HOW / LONG / DOES / A HELICOPTER / TAKE / TO / GET / TO EXODUS

3. Tag the sentence parts. The tags denote semantic concepts, specific to the task domain. They are introduced in the grammar as non-terminals, and rules are added for their expansion into word sequences[2]. To differentiate these rules from the ones used to generate sentences, we'll call them tag rules and sentence rules, respectively. E.g.,

USE *a_vehicle* TO *transport who where_from*
HOW *long* DOES *a_vehicle* TAKE TO *move-i where_to*

a_vehicle ::= ONE TRUCK
who ::= ALL THE PEOPLE
a_vehicle ::= A HELICOPTER

4. If necessary, regroup rules produced by the previous step. This step may involve generalizing over several rules and/or splitting rules that may generate unnaturally long sentences. The rules affected here may be both sentence rules and tag rules. E.g.,

transport who what_with [*where_from*]
transport who what_with [*where_to*]
[*plan_intro*] [LET'S] USE *what_vehicle* TO *transport who*

who ::= ((*det* | ONE) GROUP [OF PEOPLE])
| ([ALL] THE (PEOPLE | GROUPS))
| (*number-pl* GROUPS [OF PEOPLE])

The final PACIFICA grammar contained slightly more than 200 sentence rules and about 80 tag rules. Note that, in order to reduce the generative power of the grammar, some of the very long patterns were split in step 4 into multiple shorter patterns that would generate more natural sentences. The apparent loss in coverage is not a real problem, since the main purpose of the grammar is to provide for realistic n-gram occurrences. As a side note, we also find this methodology of building the grammar very useful for fixing the vocabulary for the new domain.

Following some of the aforementioned authors, we could generate the artificial corpus by using a parser in generative mode. We see the fact that it doesn't depend on the availability of a task-specific parser (which might not be obtained as easily) as an advantage of our approach. We eventually used a general-purpose parser[3] to filter out ungrammatical sentences in the artificial corpus, but it turned out that the improvements brought by this additional step were not significant.

3. Reusing Corpora from Similar Domains

Another solution to the adaptation problem is to use out-of-domain (OOD) text data. Estimating a good language model for a particular domain would require huge quantities of data that are often unavailable, especially when building models for a new domain. Fortunately there are increasing amounts of data available for other domains. The problem is how to get the relevant information out of them, because using non-adapted out-of-domain LMs has been proven to be worse than not using any model at all.

3.1. Relevance of out-of-domain data

Out-of-domain data has been used before, mostly to improve LMs generated from insufficient in-domain data, and usually it has been pointed out that it is useful to the degree of its "similarity" or "relevance" to the target domain. However, few attempts have been made to quantify this "relevance". Two symmetrical measures of text similarity are given for syllable models by Matsunaga, Yamada & Shikano (1992). However, for word models the (directed) notion of "relevance" seems more appropriate. Iyer (1998) used three measures of relevance (although she seems to prefer using the notion of similarity), one for content relevance (taking into account word distributions in the compared corpora), one for style relevance (based on the posterior probability of POS n-grams), and another one, which is a modified version of the latter (POS n-grams are replaced by word n-grams), that attempts to account for both content and style.

We think that the relevance of an out-of-domain corpus should be judged with respect to the language modelling technique used, since ultimately only the information in the language model is what will affect the recognition, and statistical language models contain only part of the information present in the original corpus. Statistical language models are evaluated in terms of perplexity of an in-domain data set and, when used for speech recognition, in terms of their word error rate. Therefore we will use as an indicator of the relevance of an out-of-domain corpus O, the perplexity (PP) of an in-domain text corpus T wrt the back-off language model M_O derived from O, $PP_{M_O}(T)$, and the word error rate (WER) produced by the speech recognizer [4] using the M_O language model on an in-domain speech corpus S, $WER_{M_O}(S)$. We haven't attempted to develop

[2] () are used for grouping, | for separating alternatives and [] to denote optional expressions. The non-terminals are written in italics, actual words are capitalized.

[3] The parser (Allen, 1995) that we used here is based on a syntactic unification grammar and is not tied in any way to the PACIFICA domain.

[4] The TRIPS speech recognizer is based on SPHINX-II (Huang *et al.*, 1994).

a formal measure of relevance, but this is certainly an important aspect to consider in the near future.

3.2. Class-based LMs

As observed above, corpora differences are usually studied in terms of content and style (Iyer, 1998). Content is naturally characterized in terms of vocabulary. Style differences across domains were studied by Biber (1988) in terms of co-occurrence patterns between groups of words, where the grouping is done according to part-of-speech (POS) classes. For medium-sized vocabulary speech recognition systems a common way of classifying words is in terms of semantic concepts, specific to the task domain (Issar, 1996; Ward & Issar, 1996; Popovici & Baggia, 1997). We found that this kind of class can be used to account for style as well, and in addition provide a powerful means of adapting the content of out-of-domain corpora.

Usually class-based LMs are used to generalize observed word sequences to unseen sequences and thereby compensate for the insufficient data. We will show how this technique can be used for adaptation in the next section. The general procedure for generating class-based n-gram models (Issar, 1996) follows roughly three steps (Figure 1.b):

1. the text corpus is tagged according to some predefined class tag dictionary;
2. a back-off n-gram class model is computed from the tagged text corpus; and
3. the class model is converted to a word model using again the word-class mappings in the class tag dictionary[5].

3.3. The proposed methodology

In order to maximize the overlap between the word-class mappings for different domains, we put in a general dictionary all the most common words (eg, functional words, pronouns, many common-use words). The words that are domain-specific, or that have domain-specific senses, are grouped into separate dictionaries, one for each task domain. The class tags are assigned by hand. Thus, even if domain vocabularies may differ significantly, tagged corpora from different domains will look very similar. Also, words that are domain-specific can be easily spotted. Here are some examples of words from different domains that share the same tags:

Tag	Domain		
	ATIS	TDC	PACIFICA
city	L.A. SEATTLE	CORNING DANSVILLE	CALYPSO BARNACLE
transport	BUS CAB	TRAIN -	TRUCK HELI

The scheme we propose for using out-of-domain text corpora to build LMs for new domains can be summarized in the following steps:

[5] For LM building we made use of the CMU-Cambridge Statistical Language Model Toolkit (Clarkson & Rosenfeld, 1997).

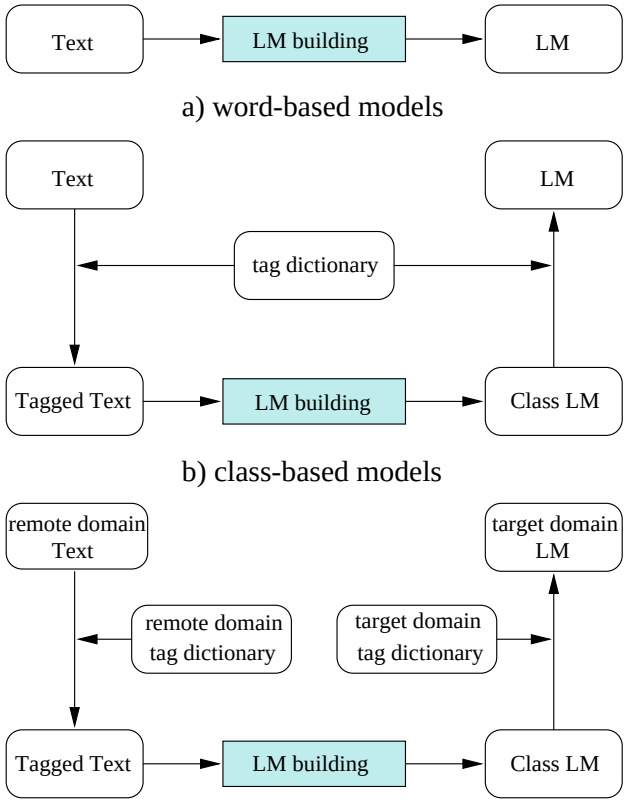

Figure 1: Block diagrams of the different LM generation procedures.

1. the corpora are tagged, each one according to the general tag dictionary and the appropriate domain-specific tag dictionary;
2. back-off class n-gram LMs are obtained from the tagged text corpora;
3. the class n-gram LMs are converted to word n-gram LM using the in-domain word-class dictionary.

Figure 1.c shows a diagram of this process. It should be clear that in step 3) tags will be expanded only with in-domain words, so this acts like a filter: all the out-of-domain n-grams will actually be discarded. Thus, only the relevant information from the OOD corpora will contribute to the LM being built.

It is possible that one OOD corpus that might look relevant to the target domain is very sparse. Several OOD corpora can be used by combining the information gotten from each of them. The combination can be done either by merging the counts, or by interpolating the language models (cf. Iyer, Ostendorf & Gish, 1997). We tried both methods, but we'll report only results with interpolated models, which performed better.

4. Experimental results

We ran perplexity and recognition experiments for the TRIPS transportation/logistics domain (Ferguson & Allen, 1998). The OOD corpora we selected were: the ATIS-2 and ATIS-3 air travel information corpora (MADCOW, 1992; Dahl *et al.*, 1992), consisting of 23k utterances and 200k word tokens, the TRAINS95 transportation scheduling domain corpus (Allen *et al.*, 1996), consisting of 5k utterances and 30k word tokens, and the TDC human-human spoken dialog corpus (Heeman & Allen, 1995), consisting of 7.5k utterances and 52k word tokens.

After setting up the TRIPS system, we also collected a small in-domain corpus (we will call it TRIPS as well) of 2389 utterances and 12656 word tokens. From these, we used 1500 randomly selected utterances as test data. We used the rest of 889 utterances to build class-based bigram models for the TRIPS domain. A model built from such a small training corpus is likely not to be very accurate, but its performance can provide an indication of how good a model one can hope to build. Therefore, we will be able to judge the performance of our new models by comparing them to the perplexity and recognition results for the TRIPS models. In addition, we will compare the new language models to a NULL model in which all the words have the same probability of occurrence.

The models are all open vocabulary bigram back-off models, with Witten-Bell discounting (Witten & Bell, 1991).

4.1. Testing the artificial corpus approach

To test the technique presented in section 2, we used the grammar to generate a corpus of sentences, by Monte Carlo sampling. This corpus was then filtered with the TRIPS parser. The final artificial corpus (call it PAC) contained about 20k sentences and 174k word tokens, and we used it to obtain class-based back-off bigram LMs, both with comprehensive and task-specific vocabularies. The results are shown in Table 1, and not only show a significant improvement over a zero-knowledge model (NULL), but also they are sufficiently close to the TRIPS model performance. Taking the TRIPS model's performance as optimal, the reduction of word error rate from the NULL model is more than 75% in the case of the class-based PAC model with task-specific vocabulary .

As we advocated earlier, we found that the use of the parser is not crucial. Experiments with a similar corpus of sentences generated by our grammar, but not filtered with the parser gave comparable perplexity and recognition results (the perplexity was worse by 6.78% relative but the accuracy was actually better by 0.95% relative), although the proportion of fully parsable sentences in this second corpus was just about 37%. This proves that our generative grammar provides a good coverage of the possible word collocations, which is what a bigram model encodes.

Although this approach provides good bigram coverage, the language model parameters may not be well adapted to the target domain. Since we assume no in-domain corpus is available, we can't do much about it. However, as we'll show in the next section, interpolating the artificial corpus-based language model with models derived from OOD spoken dialog corpora may help to remedy this deficiency quite well.

	comprehensive vocabulary		task-specific vocabulary	
	PP	WER	PP	WER
NULL	3600.0	66.3	1862.0	56.9
PAC	78.55	31.2	57.94	28.2
TRIPS	21.08	20.8	15.92	18.8

Table 1: Test set perplexities and word error rates of the PAC model compared to those for the NULL and TRIPS models. PAC and TRIPS are class-based models, while NULL is a word-base model.

	comprehensive vocabulary		task-specific vocabulary	
	PP	WER	PP	WER
ATIS	986.04	50.1	712.54	43.1
TDC	650.95	47.6	461.84	42.8
T95	722.93	44.3	430.31	40.3

Table 2: Test set perplexities and word error rates of the ATIS, TDC and T95 word-based models, with comprehensive and task-specific vocabularies.

4.2. Testing the usage of out-of-domain corpora

We obtained perplexity and recognition results (Tables 2 and 3) for several types of language models derived from the three OOD corpora: word-based models and class-based models with a comprehensive vocabulary (3600 words), and adapted word-based and class-based models, with a task-specific vocabulary (1862 words).

The OOD word-based LMs have word error rates better than if no model were used (an improvement of up to 24-29% relative), which shows that the respective domains are somewhat relevant to the TRIPS domain. However, their performance is extremely poor. The perplexity improvements over the baseline model are significant, too, but they still remain an order of magnitude higher than what we'd aim to.

Class-based models have better WER by 6.8-11.8% relative when compared to the word-based models. At the same time, the perplexity is reduced by 52-67% relative. Furthermore, when the vocabulary is restricted to the target domain, following the adaptation procedure described in section 3.3, the word error rates go even lower, by up to almost 16% relative in the case of class-based models. Simultaneously, the perplexity is reduced by 30-35% relative. Similar results are obtained for this technique in the case of word-based models, but these have significantly poorer performance than the corresponding class-based models.

Since the three OOD corpora and the artificial corpus

	comprehensive vocabulary		task-specific vocabulary	
	PP	WER	PP	WER
ATIS	468.71	46.7	305.45	39.3
TDC	260.88	42.0	184.03	37.5
T95	236.35	40.6	168.90	35.0

Table 3: Test set perplexities and word error rates of the ATIS, TDC and T95 class-based models, with comprehensive and task-specific vocabularies.

	PP (red.[%])	WER (red.[%])
TDC+T95	106.02 (37.22)	32.8 (6.3)
ATIS+TDC+T95	92.01 (45.52)	33.7 (3.7)
PAC+		
ATIS	41.85 (27.77)	26.4 (6.4)
T95	33.79 (41.68)	25.8 (8.5)
TDC+T95	33.08 (42.90)	26.1 (7.4)
ATIS+TDC+T95	32.86 (43.29)	26.3 (6.7)

Table 4: Test set perplexities and word error rates of the interpolated models and their relative reductions compared to the corresponding results of the best component.

have different characteristics, we would expect that by combining the corresponding language models we could obtain even better results. Indeed, the linear interpolation of various combinations of models provided significantly better performance (Table 4). The individual models are the adapted class-based models from above. In all cases, the interpolated models had lower perplexity and word error rates than the individual component models. The reduction in the word error rate was up to 6.3% relative compared to the best component model when only OOD models were interpolated, and up to 8.5% relative compared to the PAC model, when we interpolated this one with the OOD models. At the same time, the reductions in perplexity were up to 45.52% compared to the best component's perplexity for OOD models only, and up to 43.29% compared to the PAC model, when we interpolated this one with the OOD models.

The interpolation weights were obtained a posteriori, using an EM algorithm so as to minimize the perplexity of the TRIPS test data relative to the interpolated model [6]. While this shows that good results are possible, it doesn't provide a technique of "guessing" the right combination when no target domain data is available, as is the case for us.

Note that we used the same data for testing and for obtaining the interpolation weights; thus, it is reasonable to expect that the results in Table 4 are optimistic. At the time of the conference we will have more TRIPS data available so that we can re-run these experiments with disjoint data sets for training the interpolation weights and for testing

[6] It should be noted that a reduction in perplexity is not always accompanied by a reduction in the word error rate.

	PP	WER
TRIPS	15.92	18.8
TDC+T95+PAC+TRIPS	15.20	19.1
ATIS+TDC+T95+PAC+TRIPS	15.19	19.2

Table 5: The effect of interpolating an in-domain LM with the adapted LMs on the test set perplexities and word error rates.

the interpolated models. Consequently, we will also have a more reliable TRIPS model as a baseline reference.

We also started to investigate whether the adapted models might improve even on in-domain models obtained from very limited amounts of data. The results (Table 5) show a reduction in perplexity of almost 4.6%. The same caveats with respect to the interpolation and testing phases using the same data set apply here also. The interpolated model has been obtained by minimizing the perplexity of the TRIPS test data relative to the interpolated model, and this didn't provide a reduction in the word error rate. However, the increase in WER is too small to be significant. We intend to further investigate this issue.

5. Conclusions

The results obtained are very encouraging. We were able to build rapidly sufficiently good medium-sized vocabulary language models for a new task domain without having any domain-specific data. Each of the two language model adaptation techniques proposed, using artificial corpora, and using out-of-domain corpora, provided very good performance results, and their combination showed significant improvement over each one alone. We then used our system to collect real, unsimulated data (in contrast to the "wizard of Oz" technique). As soon as more target domain data is available, we think that better results can be obtained by further adaptation with models obtained from this data. We intend to devote more work to subsequent adaptation. An interesting area of further research is to find better ways of judging the relevance of the OOD corpora, better techniques of filtering the irrelevant parts, and better methods of combining OOD models so as to maximize the benefit provided by each model.

6. Acknowledgments

We would like to thank Nathaniel Martin for the initial collection of utterances in the PACIFICA domain and for initiating their abstraction into CFG rules, and the whole TRAINS/TRIPS group at the University of Rochester for helping us collect and transcribe the speech data and for system support. We are indebted to Sunil Issar and many others at CMU for SPHINX-II and other tools we used in this research.

This work has been funded in part by ARPA/Rome Laboratory contract no. F30602-95-I-1088, ONR grant no.

N00014-95-I-1088, and NSF grant no. IRI-9623665.

7. References

Allen, J.F (1995) The TRAINS-95 Parsing System: A User's Manual. TRAINS TN 95-1, Department of Computer Science, University of Rochester.

Allen, J.; Miller, B.W.; Ringger E.K.; Sikorski, T. (1996). A Robust System for Natural Spoken Dialogue. In *Proc. ACL'96* (pp. 62–70).

Biber, D. (1988). *Variations Across Speech and Writing*. Cambridge University Press, 1988.

Clarkson, P.R.; Rosenfeld, R. (1997) Statistical Language Modeling Using the CMU-Cambridge Toolkit. In *Proc. Eurospeech 1997* (pp. 2707–2710).

Crespo, C.; Tapias, D.; Escalada, G.; Alvarez, J. (1997). Language Model Adaptation for Conversational Speech Recognition Using Automatically Tagged Pseudo-Morphological Classes. In *Proc. ICASSP'97* (pp. 823–826).

Dahl, D.A. *et al.* (1992) Expanding the Scope of the ATIS Task: The ATIS-3 Corpus. In *Proc. ARPA Human Language Technology Workshop '92* (pp. 45–50).

Ferguson, G.; Allen, J. (1998). TRIPS: An Intelligent Integrated Problem-Solving Assistant. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, Madison, WI, 26-30 July, 1998. To appear.

Heeman, P.A.; Allen J. (1995). The TRAINS 93 Dialogues. TRAINS TN 94-2 (corpus available from LDC).

Huang, X.; Alleva, F.; Hwang, M.-Y.; Rosenfeld, R. (1994) An Overview of the SPHINX-II Speech Recognition System. In *Proc. ARPA Human Language Technology Workshop '93* (pp. 81–86).

Issar, S. (1996). Estimation of Language Models for New Spoken Language Applications. In *Proc. ICSLP'96* (pp. 869–872).

Ito, A.; Saitoh, H.; Katoh M.; Kohda, M. (1997) N-Gram Language Model Adaptation Using Small Corpus for Spoken Dialog Recognition. In *Proc. Eurospeech'97* (pp. 2735–2738).

Iyer, R. (1998) Improving and Predicting Performance of Statistical Language Models in Sparse Domains. PhD Dissertation, Boston University.

Iyer R.; Ostendorf, M. (1997) Transforming Out-of-Domain Estimates to Improve In-Domain Language Models. In *Proc. Eurospeech'97* (pp. 1975–1978).

Iyer, R.; Ostendorf, M.; Gish, H. (1997) Using Out-of-Domain Data to Improve In-Domain Language Models. TR ECE-97-001, Boston University.

Jurafsky, D.; Wooters, C.; Tajchman, G.; Segal, J. (1994) The Berkeley Restaurant Project. In *Proc. ICSLP'94* (pp. 2139–2142).

Jurafsky, D.; Wooters, C.; Segal, J.; Stolcke, A.; Fosler, E.; Tajchman, G.; Morgan, N.(1995) Using a Stochastic Context-Free Grammar as a Language Model for Speech Recognition. In *Proc. ICASSP'95* (pp. 189–192).

MADCOW (1992) Multi-Site Data Collection for a Spoken Language Corpus. In *Proc. DARPA Speech and Natural Language Workshop '92* (pp. 7–14).

Matsunaga, S.; Yamada, T.; Shikano, K. (1992) Task Adaptation in Stochastic Language Models for Continuous Speech Recognition. In *Proc. ICASSP '92* (pp. 165–168).

Popovici, C.; Baggia, P. (1997) Language Modelling for Task-Oriented Domains. In *Proc. Eurospeech '97* (pp. 1459–1462).

Rayner, M.; Carter, D. (1997) Hybrid Language Processing in the Spoken Language Translator. In *Proc. ICASSP'97* (pp. 107–110).

Rudnicky, A.I. (1995) Language modeling with limited domain data. In *Proc. ARPA Spoken Language Technology Workshop*, 1995.

Ward, W.; Issar, S. (1996) A Class Based Language Model For Speech Recognition. In *Proc. ICASSP '96* (pp. 416–419).

Witschel, P.; Höge, H. (1997) Experiments in Adaptation of Language Models for Commercial Applications. In *Proc. Eurospeech'97*, (pp. 1967–1970).

Witten, I.T.; Bell, T.C. (1991) The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression. *IEEE Transactions on Information Theory*, 37(4):1085–1094.

Zue, V.; Glass, J.; Goodine, D.; Leung, H.; Phillips, M.; Polifroni, J.; Seneff, S. (1991) Integration of Speech Recognition and Natural Language Processing in the MIT VOYAGER System. In *Proc. ICASSP'91* (pp. 713–716).