РАЗРАБОТКА ФАКТОРНЫХ МОДЕЛЕЙ ЯЗЫКА ДЛЯ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РУССКОЙ РЕЧИ

Кипяткова И. С. (kipyatkova@iias.spb.su)[1,2]
Карпов А. А. (karpov@iias.spb.su)[1,3]

[1]Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, Россия

[2]Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП), Санкт-Петербург, Россия

[3]Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (ИТМО), Санкт-Петербург, Россия

В статье описывается процесс создания и исследования факторных моделей языка для системы автоматического распознавания русской речи. Различные факторные модели языка и базовая 3-граммная модель были обучены на текстовом корпусе, сформированном из интернет-сайтов ряда электронных газет, содержащем более 350 млн словоупотреблений. Были созданы факторные модели с фиксированными и с параллельными путями возврата, при этом использовалось 5 лингвистических факторов: словоформа, лемма, основа слова, часть речи и метка морфологических признаков. Оптимизация параметров моделей производилась с использованием генетического алгоритма. Созданные модели были внедрены в систему автоматического распознавания русской речи и используются на этапе переоценки списка лучших гипотез распознавания. В ходе экспериментов по распознаванию слитной русской речи со сверхбольшим словарем относительное уменьшение процента неправильно распознанных слов, полученное после выполнения переоценки списка гипотез распознавания с использованием факторных моделей языка, интерполированных с базовой 3-граммной моделью, составило 8%.

Ключевые слова: факторные модели языка, автоматическое распознавание речи, русская речь, корпусные исследования

DEVELOPMENT OF FACTORED LANGUAGE MODELS FOR AUTOMATIC RUSSIAN SPEECH RECOGNITION

Kipyatkova I. S. (kipyatkova@iias.spb.su)[1,2]
Karpov A. A. (karpov@iias.spb.su)[1,3]

[1]St. Petersburg Institute for Informatics and Automation of RAS (SPIIRAS), St. Petersburg, Russia

[2]St. Petersburg State University of Aerospace Instrumentation (SUAI), St. Petersburg, Russia

[3]ITMO University, St. Petersburg, Russia

In this paper, we present a study of factored language models (FLM) of Russian for rescoring N-best lists in automatic speech recognition (ASR) systems. We used 3-gram language models as baseline. Both 3-gram and factored language models were trained on a text corpus collected from recent Internet online newspapers; total size of the text corpus is about 350 million words (2.4 Gb data). For FLM creation, we used five linguistic factors: word-form, word lemma, stem, part-of-speech, and morphological tag. We studied several FLMs with two factors (word-form plus one of the other factors) using 2 fixed backoff paths: (1) the first drop was of the most distant word and factor, then—of the less distant ones; (2) the first drop was of the words in time-distance order, then drop of the factors in the same order. We investigated the influence of a factor set and backoff paths on language model perplexity and word error rate (WER). Also we created FLMs with some parallel generalized backoff paths. Optimization of the FLM parameters was carried out by means of the genetic algorithm. The FLMs were embedded in the automatic Russian speech recognition system with a very large vocabulary. Experimental results on continuous Russian speech recognition task showed a relative WER reduction of 8% when the FLM was interpolated with the baseline 3-gram model.

Key words: factored language models, automatic speech recognition, Russian speech, corpus studies

1. Introduction

The most widely used language models (LMs) are statistical *n*-gram models, which estimate the probability of appearance of a word sequence $X = (W_1, W_2, \ldots, W_m)$ in a text [16]. Rich morphology of the Russian language leads to increasing the perplexity of *n*-gram models. These models are efficient for many languages, but for Russian they do not work so well. Russian is a morphologically rich inflective language. This results in the increasing of vocabulary size as well the perplexity of *n*-gram language models. In [25], it was shown that changing the vocabulary size from 100K to 400K words increases the English model perplexity by 5.8% relatively, while the Russian model perplexity increases by as much as 39.5%.

A state-of-the-art alternative to n-gram language models is a factored language model (FLM) that for the first time was introduced in order to deal with the morphologically rich Arabic language [4]. Then it has been used for many other morphologically rich languages. This model incorporates various morphological features (factors) and it can be applied to inflective languages too. So, a word is represented as a vector of k factors: $w_i = (f_i^1, f_i^2, \ldots, f_i^k)$. Factors of a given word can be such as word-form, morphological class, stem, root, and other grammatical features. Probabilistic language model is constructed with sets of the factors.

In [23], a FLM was incorporated at different stages of speech recognition: N-best list rescoring and recognition stage. Recognition results showed an improvement of word error rate (WER) by 0.8–1.3% with the FLM used for N-best rescoring task depending on the test speech corpus; and the usage of FLM at speech recognition gave additional improving of WER by 0.5%.

A FLM was applied for lattice rescoring in [20]. The decoder generated a lattice of 100 best alternatives for each test sentence using a word-based bigram LM with 5K vocabulary. Then the lattice was rescored with various morpheme-based and factored language models. Word recognition accuracy obtained with the baseline model was 91.60%, and the usage of the FLM increased word recognition accuracy up to 92.92%.

In [3], a morpheme-based trigram LM for Estonian was used for N-best list generating. The vocabulary of the language model consisted of 60K word particles. Recognized morpheme sequences were reconstructed to word sequences. A FLM, which used words and their part-of-speech (POS) tags, was applied to rescore N-best hypotheses. A relative WER improvement of 7.3% was obtained on a large vocabulary.

FLMs are also used for code-switching speech [1, 9]. In [1], for code-switching speech the following factors were analyzed: words, POS tags, open class words, and open class word clusters. FLM was used at the speech decoding stage. For this purpose BioKIT speech decoder [21] was extended to support such models. Experiments on recognition of Mandarin-English code-switching speech showed a relative reduction of mixed error rate by 3.4%. In [2], a FLM was combined with recurrent neural networks (RNN) for Mandarin-English code-switching language modeling task. The combined LM gave a relative improvement of 32.7% comparing to the baseline 3-gram model.

An application of FLMs for Russian speech recognition is described in [22]. The FLM was trained on a text corpus containing 10M words with a vocabulary size of about 100K words. FLMs were created using the following factors: word, lemma, morphological tag, POS, and gender-number-person factor. TreeTagger tool [17] was used for obtaining linguistic factors. Influence of different factors and backoff paths on the perplexity and WER was tested. FLM was used for rescoring 500-best lists. Evaluation experiments showed that FLM allows achieving 4.0% WER relative reduction, and 6.9% relative reduction was obtained after interpolation of the FLM with the baseline 3-gram model.

2. Creation of Factored Language Models for Russian

There are two main issues at development of FLM [14]:

1. Choosing an appropriate set of factor definitions using data-driven techniques or linguistic knowledge.
2. Finding the best statistical model for these factors.

One of the problems at creating statistical LMs is the lack of training data (especially for under-resourced languages) [9]. To solve this problem backoff methods are used [16]. In word *n*-gram modeling, backing-off is performed by dropping first the most distant word, followed by the second most distant word, and so on until the unigram language model is used. This process is illustrated in Figure 2(a). In FLM, there is no obvious path of backoff [4]. In FLMs, any factor can be dropped at each step of the backoff process, and it is not obvious, which factor to drop first. In this case, several backoff paths are possible, that results in a backoff graph. An example of the backoff graph is presented in Figure 1(b). The graph shows all possible single step backoff paths, where exactly one variable is dropped per each step.

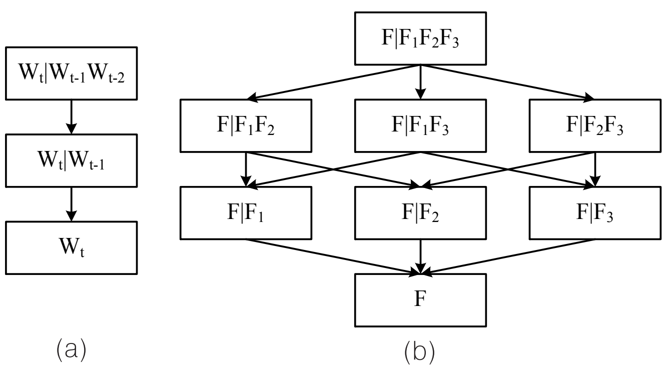


Fig. 1. Backoff graphs for *n*-gram and FLMs: (a) backoff path for a 3-gram language model over words; (b) backoff graph with three parent factors F1, F2, F3

In order to choose the best factor set and backoff path, linguistic knowledge or data-driven techniques can be applied. In [23], it was shown that an automatic method that uses Genetic Algorithm (GA) for optimization of the factor set, backoff path, and smoothing techniques, performs better than the manual search in terms of perplexity. The goal of this method is to find a combination of parameters that produces a FLM with a low perplexity on unseen test data [14].

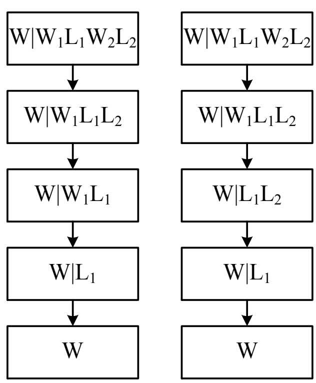
For the language model creation, we collected and automatically processed a Russian text corpus of some on-line newspapers. It contains news texts of different topics: politics, economy, culture, sport, etc. The procedure of preliminary text processing and normalization is described in [8]. The size of the corpus after text normalization and deletion of doubling and short (<5 words) sentences is over 350M words, as well as it contains above 1M unique word-forms.

The software "VisualSynan" of AOT project [18] was used for obtaining morphological features for words. This tool can make a morphological analysis of Russian, English, and German texts. Output of the morphological analysis is quite correct, although some errors are exist. We used 5 linguistic factors: word-form, its lemma, stem, part-of-speech (POS), and morphological tag. The training text corpus was processed to replace words with their factors. For example, the word-form '**схеме**' ("scheme") is replaced with the vector {W-**схеме**: L-**схема**: S-**схем**: P-**сущ**: M-bc}, where W means a word-form, L denotes a lemma, S is a stem, P is POS, M is a morphological tag, which indicates that the given word-form is a noun with feminine gender, singular, dative case.

2.1. FLMs with fixed backoff paths

We used the SRI Language Modeling Toolkit (SRILM) [19] for LM creation. At first, we created 2-factor LMs with the word-form plus one of the other factors. To create these models we used two fixed backoff paths:

1. The first drop was of the most distant word-form and factor, then—of the less distant ones (Figure 1a).
2. The first drop was of the word-forms in time-distance order, and then the drop of the factors in the same order (Figure 2a).



(a) (b)

Fig. 2. Backoff paths for WL model: (a) backoff path 1; (b) backoff path 2

For example, for the real trigram "**вагонов грузового состава**" ("wagons of freight train") the backoff path is the following:

вагонов грузового состава
грузового состава
состава

When creating 2-factor LM (word-form and lemma) this trigram is converted into a sequences of factors: "W-**вагонов** L-**вагон** W-**грузового** L-**грузовой** W-**состава** L-**состав**". Backoff paths can be the following:

Backoff path 1:

L-вагон W-вагонов L-грузовой W-грузового W-состава
L-вагон L-грузовой W-грузового W-состава
L-грузовой W-грузового W-состава
L-грузовой W-состава
W-состава

Backoff path 2:

L-вагон L-грузовой W-вагонов W-грузового W-состава
L-вагон L-грузовой W-грузового W-состава
L-вагон L-грузовой W-состава
L-грузовой W-состава
W-состава

When creating LMs, discounting techniques are used to assign nonzero probabilities to *n*-grams that were not observed in the training corpus by discounting probabilities of the observed *n*-grams [24]. Therefore, we investigated FLMs with different discounting techniques: (1) Good-Turing; (2) Unmodified Kneser-Ney; (3) Modified Kneser-Ney; (4) Witten-Bell; (5) Natural [24]. Perplexities of the created FLMs are shown in Table 1; they were calculated on text data consisting of phrases (33M word usage in total) from another online newspaper "Фонтанка.ru" (www.fontanka.ru), which was not used for LM training. The models built with backoff path 1 have smaller perplexities for all discounting techniques and factors. Some discounting techniques gave better results depending on factor combinations. The best perplexity was obtained using the LM with word-form and lemma factors created with the modified Kneser-Ney discounting technique. Also this discounting method gave better (smaller) perplexity for all other LMs excepting the model with word-form and part-of-speech factors. For this model the Good-Turing discounting method was the best. The largest (worst) LM perplexity was obtained using the model with word-form and stem factors with the Witten-bell discounting. The perplexity of the baseline 3-gram LM was 553 [10].

Table 1. Perplexity of FLMs with different discounting techniques and backoff paths

Factors	Discounting techniques									
	Good-Turing		Unmodified Kneser-Ney		Modified Kneser-Ney		Witten-Bell		Natural	
	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2
WM	573	696	593	724	566	691	749	898	761	916
WL	557	597	550	603	**529**	577	826	1007	747	779
WP	572	636	649	755	623	729	725	727	734	762
WS	617	685	617	701	595	672	879	1098	824	895

2.2. FLMs with parallel generalized backoff

We created also FLM using all factors and parallel generalized backoff. Since the creation of such a model requires a large amount of memory, we used only a part of the text corpus, which contains 100M words. We applied the genetic algorithm (GA) [5] to find the best backoff graph. As initial factors we used all mentioned above factors and discounting methods, as well as the time context of 2. GA was implemented using the population size of 10 and the maximum number of generation of 20 [11].

We chose two models, which are the best in the terms of perplexity for the experiments on Russian ASR. A backoff graph for the first model (FLM 1) is presented in Figure 3; the backoff graph for the second model (FLM 2) is presented in Figure 4. In these figures, a digit after a factor symbol denotes a time context.

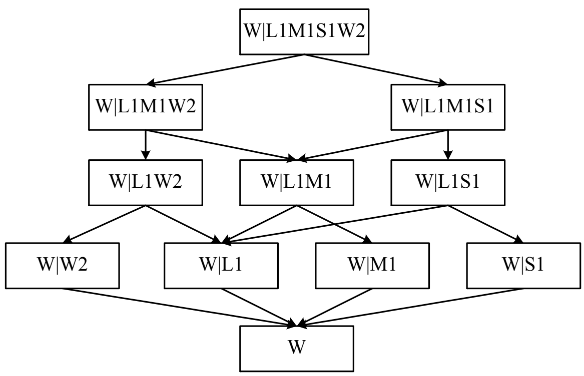


Fig. 3. Backoff graph for FLM 1

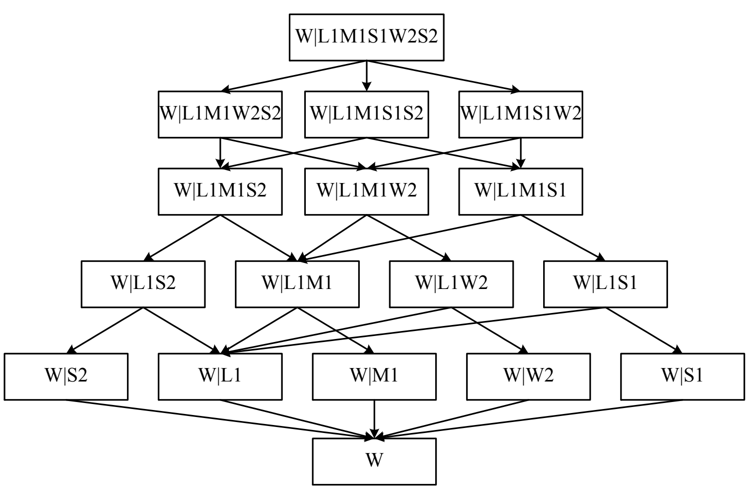


Fig. 4. Backoff graph for FLM 2

Both models use 4 factors: lemma, morphological tag, stem, word-form, and three discounting methods on different stages of backing-off: Unmodified Kneser-Ney, Modified Kneser-Ney, and Witten-Bell. The perplexity of FLM 1 is 589, and the perplexity of FLM 2 is 618.

3. Russian Speech Recognition System with FLM

Architecture of the Russian ASR system with developed FLMs is presented in Fig. 5.

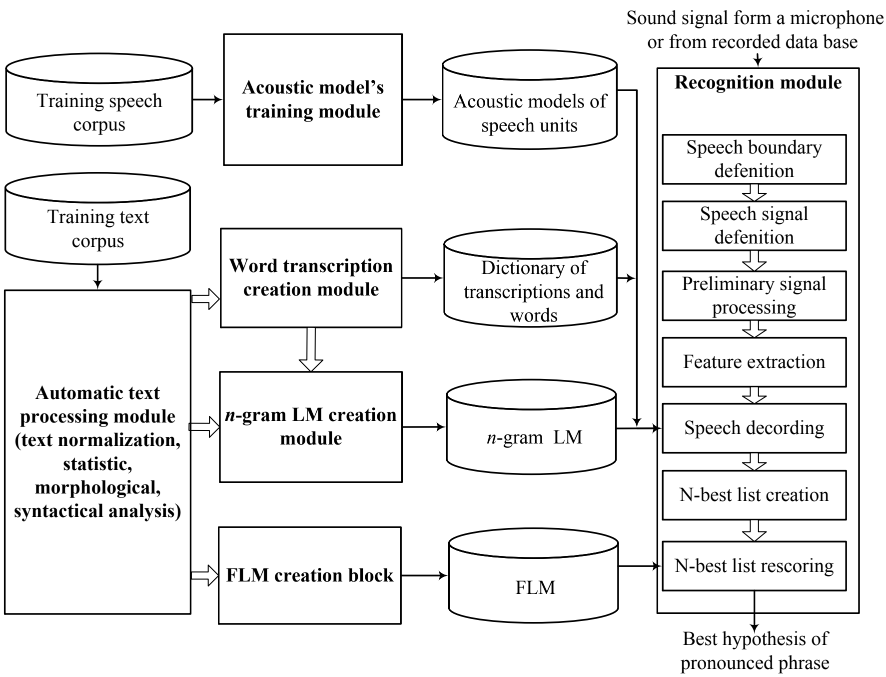


Fig. 5. Architecture of Russian ASR system with FLM

The system works in 2 modes [12]: training and recognition. In the training mode, acoustic models of speech units, a phonemic vocabulary of word-forms, as well as *n*-gram and factored LMs are created. In the speech recognition mode, an input speech signal is transformed into the sequence of feature vectors (Mel-Frequency Cepstral Coefficients with the 1st and 2nd order derivatives are used), and then the search of most probable hypotheses is performed with the help of preliminary trained acoustic and language models. FLM is used at the stage of post-processing for N-best list rescoring. Thereby, on the speech recognition stage, 3-gram LM is used for creating N-best list and then FLM is applied for rescoring obtained N-best list of hypotheses and for selection of the best recognition hypothesis for pronounced phrase.

4. Experiments on continuous Russian speech recognition

4.1. Training and testing speech corpora

For training the speech recognition system we used our own corpus of spoken Russian speech, created by SPIIRAS in 2008–2009 in the framework of Euronounce project [6, 7]. The speech data were collected in clean acoustic conditions, with 44.1 kHz sampling rate, 16-bit audio quality. The signal-to-noise ratio (SNR) of 35–40 dB at least was provided. The database consists of 16,350 utterances pronounced by 50 native Russian speakers (25 male and 25 female). Each speaker pronounced more than 300 phonetically-balanced and meaningful phrases. Total duration of the speech data is about 21 hours.

Acoustic models were created with the help of HTK toolkit [26]. As for acoustic features, we used 13-dimentional MFCCs with the 1st and 2nd order derivatives calculated from the 26-channel filter bank analysis of 20 ms long frames with 10 ms overlap. Cepstral mean subtraction (CMS) is applied to audio feature vectors. For acoustic modeling, continuous density Hidden Markov Models (HMM) were used, and each phoneme was modeled by one HMM.

To test the system we used a speech corpus that contains 500 phrases pronounced by 5 speakers (each speaker pronounced the same 100 phrases). The phrases were taken from the materials of the on-line newspaper "Фонтанка.ru" that was not used in the training data.

4.2. Study of FLMs with fixed backoff paths

Russian ASR system was built on the base of Julius ver. 4.2 decoder [15]. System's performance was estimated by the word error rate (WER) measure. At the speech decoding stage, 3-gram LM was used. WER obtained with this model was 26.54% [10]. The vocabulary size was 150K words. The out-of-vocabulary rate for the test set was 1.1%. The baseline ASR system produced 20-best lists of hypotheses for each pronounced phrase. The rescoring of the 20-best lists was carried out using created FLMs. The recognition results are summarized in Table 2.

Table 2. WER obtained after 20-best list rescoring (%)

FLMs	Discounting techniques									
	Good-Turing		Unmodified Kneser-Ney		Modified Kneser-Ney		Witten-Bell		Natural	
	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2
WM	27.87	28.00	27.79	28.09	28.15	28.16	**27.30**	27.55	27.58	27.40
WL	28.45	28.78	28.37	28.82	28.28	28.99	27.83	28.39	27.88	28.67
WP	28.61	28.61	28.58	28.71	28.63	28.88	27.72	28.48	28.33	28.91
WS	29.93	30.24	29.78	30.19	30.02	30.28	29.01	29.46	28.90	29.91

FLMs	Discounting techniques									
	Good-Turing		Unmodified Kneser-Ney		Modified Kneser-Ney		Witten-Bell		Natural	
	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2
Interpolated models										
WM+3-gram	25.00	24.89	24.57	24.93	**24.44**	24.78	24.94	25.22	25.41	25.36
WL+3-gram	25.51	25.54	25.54	25.67	25.58	25.43	25.51	25.47	24.98	25.47
WP+3-gram	25.21	25.28	25.30	25.32	25.07	25.24	25.47	25.64	25.60	25.43
WS+3-gram	25.97	25.92	26.03	25.86	25.88	25.90	26.05	25.86	25.49	25.90

We produced lists of 20-best hypotheses and rescored them using created FLMs. The best results were obtained using the LM with word-form and morphological tag factors created with the Witten-Bell discounting. Optimal WER value was 27.30%. So, the WER was worse than one obtained before N-best list rescoring. For models with other combination of factors the Witten-Bell discounting also gave better results, although in terms of perplexity this discounting method was not the best. Then we carried out linear interpolation of FLMs with the baseline 3-gram LM. The lowest WER=24.44% was obtained after interpolation of the baseline model with the FLM, in which word-form and morphological factors were used. This model was created using modified Kneser-Ney discounting technique with the backoff path 1.

Then, we produced N-best lists with the number of hypotheses from 10 to 50 and performed their rescoring using FLM with the Modified Kneser-Ney discounting technique interpolated with 3-gram model. Recognition results are presented in Table 3. Also in the table oracle WER, which is minimal value of WER that can be obtained choosing the most accurate hypothesis from N-best list, is shown. From the table we can see that rescoring of 20-best list gives better results.

Table 3. WER obtained after rescoring of N-best lists (%)

Language models	N=10		N=20		N=50	
	Path 1	Path 2	Path 1	Path 2	Path 1	Path 2
3-gram (oracle WER)	18.52		16.63		15.34	
3-gram + WM	24.83	24.94	**24.44**	24.78	24.55	24.66
3-gram + WL	25.79	25.71	25.58	25.43	25.60	25.37
3-gram + WP	25.43	25.54	25.07	25.24	25.15	25.26
3-gram + WS	25.82	26.01	25.88	25.90	25.90	26.10

4.3. Study of LMs with parallel generalized backoff

Then experiments on rescoring 20-best lists using FLMs with the parallel generalized backoff method were conducted. Obtained results are presented in Table 4. The use of FLMs for N-best list rescoring did not improve the ASR results. Therefore, we have performed a linear interpolation of FLMs with the baseline model. The best WER was obtained with the FLM 1 interpolated with the baseline 3-gram LM (WER=24.53%).

Table 4. WER obtained after rescoring 20-best lists with parallel generalized backoff (%)

Language models	WER, %
3-gram	26.54
FLM 1	27.94
FLM 2	28.56
FLM 1 + 3-gram	**24.53**
FLM 2 + 3-gram	24.74

Figure 5 shows the 20-best list of ASR for the Russian phrase: “Основой нашего эфира станет мировая музыкальная классика во всем многообразии жанров, стилей и направлений» (“The base of our broadcast will become world classical music in variety of genres, styles, and trends”). The hypotheses are ranked according to descending probability. After rescoring of this 20-best list using FLM 1 interpolated with the baseline 3-gram LM, the hypothesis #4 was selected as the best one. So, after N-best list rescoring we obtained the correct hypothesis for this utterance.

#1 <s> мы заранее договорились что разговор нужны для публикации проста надо познакомиться поближе </s>
#2 <s> мы заранее договорились что разговор наш не для публикации проста надо познакомиться поближе </s>
#3 <s> мы заранее договорились что разговор нужны для публикаций портала познакомиться поближе </s>
#4 <s> мы заранее договорились что разговор наш не для публикации просто надо познакомиться поближе </s>
#5 <s> мы заранее договорились что разговор наш не для публикации портала познакомиться поближе </s>
#6 <s> мы заранее договорились что разговор нужны для публикаций просто надо познакомиться поближе </s>
#7 <s> мы заранее договорились что разговор нужны для публикаций проста надо познакомиться поближе </s>
#8 <s> мы заранее договорились что разговор наш мир для публикации проста надо познакомиться поближе </s>
#9 <s> мы заранее договорились что разговор нож не для публикации просто надо познакомиться поближе </s>
#10 <s> мы заранее договорились что разговор нашли для публикации проста надо познакомиться поближе </s>
#11 <s> мы заранее договорились что разговор нож не для публикации портала познакомиться поближе </s>
#12 <s> мы заранее договорились что разговор наш ни для публикации проста надо познакомиться поближе </s>
#13 <s> мы заранее договорились что разговор наш мир для публикации просто надо познакомиться поближе </s>
#14 <s> мы заранее договорились что разговор нужные для публикации проста надо познакомиться поближе </s>
#15 <s> мы заранее договорились что разговор наш не для публикаций портала познакомиться поближе </s>
#16 <s> мы заранее договорились что разговор наш мир до публикации проста надо познакомиться поближе </s>
#17 <s> мы заранее договорились что разговор нужный для публикации проста надо познакомиться поближе </s>
#18 <s> мы заранее договорились что разговор на шнидер публикации проста надо познакомиться поближе </s>
#19 <s> мы заранее договорились что разговор нашли для публикации просто надо познакомиться поближе </s>
#20 <s> мы заранее договорились что разговор наш ни для публикаций портала познакомиться поближе </s>

Fig. 6. An example of N-best list of recognition hypotheses

Table 4 shows that the WER obtained after applying the LMs with parallel backoff paths slightly increased comparing to results obtained after applying the models with fixed backoff paths. The reason for this is that models with parallel backoff paths were trained on a portion of the corpus (100M word usage). The disadvantage of FLMs with many factors and parallel backoff paths is that these models require a large amount of memory; in our case training these models required 64 Gb RAM memory. However, it is possible to obtain decreasing WER even by training these models using a small train corpus that is an obvious advantage of FLMs.

Our experimental results are consistent with those obtained in [22], but we used another morphological parser—AOT [18] instead of TreeTagger [17]. For our experiments we used the training text corpus of 350 million words that is in 35 times larger than the set in [22]. Moreover, our WER results are better than reported in [22], and they confirm the hypothesis that the use of FLM for N-best list rescoring improves recognition accuracy. Also we can conclude that we obtained a larger relative reduction of WER in comparison with some other researches for other languages (for example, reported in [3, 20, 23]).

5. Conclusion

The study of FLMs showed that the inclusion of addition linguistic information in language models can improve the performance of ASR systems. In this paper, we compared different factor sets in terms of the word error rate. We obtained relative WER reduction of 8% comparing to the baseline ASR system. In further research, we plan to investigate FLMs with other factors as well as other types of statistical language models.

This research is partially supported by the Council for Grants of the President of Russia (Projects No. MK-5209.2015.8 and MD-3035.2015.8), by the Russian Foundation for Basic Research (Projects No. 15-07-04415 and 15-07-04322), and by the Government of the Russian Federation (Grant No. 074-U01).

References

1. *Adel H., Kirchhof K., Telaar D., Vu N. T., Schlippe T., Schultz T.* (2014), Features for factored language models for code-switching speech, Proceedings of 4th International Workshop on Spoken Language Technologies for Under-resourced languages (SLTU-2014), St. Petersburg, Russia, pp. 32–38.
2. *Adel H., Vu N. T., Schultz T.* (2013), Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria.
3. *Alumae T.* (2006), Sentence-adapted factored language model for transcribing Estonian speech, Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP- 2006). Toulouse, France, pp. 429–432.
4. *Bilmes J. A., Kirchhoff K.* (2003), Factored language models and generalized parallel backoff, Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Stroudsburg, USA, Vol. 2, pp. 4–6.
5. *Gladkov L. A., Kureychik V. V., Kureychik V. M.* (2006), Genetic algorithms [Geneticheskie algoritmy]. 2nd ed., Fizmatlit, Moscow.

6. *Jokisch O., Wagner A., Sabo R., Jaeckel R., Cylwik N., Rusko M., Ronzhin A., Hoffmann R.* (2009), Multilingual speech data collection for the assessment of pronunciation and prosody in a language learning system, Proceedings of SPECOM'2009, St. Petersburg, Russia, pp. 515–520.
7. *Karpov A., Kipyatkova I., Ronzhin A.* (2011), Very Large Vocabulary ASR for Spoken Russian with Syntactic and Morphemic Analysis, Proceedings of International Conference INTERSPEECH-2011, Florence, Italy, pp. 3161–3164.
8. *Karpov A., Markov K., Kipyatkova I., Vazhenina D., Ronzhin A.* (2014), Large vocabulary Russian speech recognition using syntactico-statistical language modeling, Speech Communication, Vol. 56. pp, 213–228.
9. *Karpov A., Verkhodanova V.* (2015), Speech Technologies for Under-Resourced Languages of the World [Rechevye tehnologii dlja maloresursnyh jazykov mira] // Problems of Linguistics [Voprosy Jazykoznanija], Vol. 2, pp. 117–135.
10. *Kipyatkova I., Karpov A.* (2013), Lexicon Size and Language Model Order Optimization for Russian LVCSR // Springer International Publishing Switzerland. M. Zelezny et al. (Eds.): SPECOM 2013, LNAI 8113, pp. 219–226.
11. *Kipyatkova I., Karpov A.* (2014), Study of Morphological Factors of Factored Language Models for Russian ASR // Springer International Publishing Switzerland. A. Ronzhin et al. (Eds.): SPECOM 2014, LNAI 8773, pp. 451–458.
12. *Kipyatkova I., Karpov A., Verkhodanova V., Zelezny M.* (2014), Modeling of Pronunciation, Language and Nonverbal Units at Conversational Russian Speech Recognition // International Journal of Computer Science and Applications, Vol. 10, N 1, pp. 11–30.
13. *Kipyatkova I., Verkhodanova V., Karpov A.* (2014), Rescoring N-best lists for Russian speech recognition using factored language models, Proceedings of 4th International Workshop on Spoken Language Technologies for Under-resourced languages (SLTU-2014), St. Petersburg, Russia, pp. 81–86.
14. *Kirchhoff K., Bilmes J., Duh K.* (2007), Factored Language Models Tutorial, Tech. Report UWEETR-2007-0003, Department of Electrical Engineering, University of Washington.
15. *Lee A., Kawahara T.* (2009), Recent Development of Open-Source Speech Recognition Engine Julius, Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2009), Sapporo, Japan, pp. 131–137.
16. *Moore G. L.* (2001), Adaptive Statistical Class-based Language Modelling, PhD thesis, Cambridge University.
17. *Schmid H.* (1994), Probabilistic part-of-speech tagging using decision trees, Proc. International Conference on New Methods of Language Processing, Manchester, UK, pp. 44–49.
18. *Sokirko A. V.* (2004), Morphological modules on www.aot.ru [Morfologicheskie moduli na sajte www.aot.ru], Proceedings of International Conference "Dialogue-2004" [Trudy Mezhdunarodnoj konferencii "Dialog-2004"], pp. 559–564.
19. *Stolcke A., Zheng J., Wang W., Abrash V.* (2011), SRILM at Sixteen: Update and Outlook, Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop ASRU'2011. Waikoloa, Hawaii, USA.

20. *Tachbelie M. Y., Teferra Abate S., Menzel W.* (2009), Morpheme-based language modeling for Amharic speech recognition, Proceedings of the 4th Language and Technology Conference, LTC-2009, Posnan, Poland, pp. 114–118.
21. *Telaar D., Wand M., Gehrig D., Putze F., Amma C., Heger D., Vu N. T., Erhardt M., Schlippe T., Janke M., Herff C., Schultz T.* (2014), BioKIT — Real-time decoder for biosignal processing, Proceedings of Interspeech-2014, Singapore, pp. 2650–2654.
22. *Vazhenina D., Markov K.* (2013), Factored Language Modeling for Russian LVCSR, Proceedings of International Joint Conference on Awareness Science and Technology & Ubi-Media Computing, Aizu-Wakamatsu city, Japan, pp. 205–210.
23. *Vergyri D., Kirchhoff K., Duh K., Stolcke A.* (2004), Morphology-Based Language Modeling for Arabic Speech Recognition, Proceedings of ICSLP 2004, pp. 2245–2248.
24. *Whittaker E. W. D.* (2000), Statistical Language Modelling for Automatic Speech Recognition of Russian and English, PhD thesis, Cambridge University.
25. *Whittaker E. W. D., Woodland P. C.* (2003), Language modelling for Russian and English using words and classes, Computer Speech and Language, Vol. 17, pp. 87–104.
26. *Young S.* et al. (2009), The HTK book, Cambridge Univ. Press.