

# Автоматическое распознавание русской речи с применением факторных языковых моделей<sup>1</sup>

**Аннотация.** Статья посвящена исследованию факторных моделей языка для системы автоматического распознавания речи. 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. Модели использовались на этапе переоценки списка  $N$  лучших гипотез, генерируемых системой распознавания речи. Представлены результаты и анализ экспериментов по применению факторных моделей языка для задачи распознавания русской речи со сверхбольшим словарем.

**Ключевые слова:** факторные модели языка, автоматическое распознавание речи, русская речь, корпусные исследования, обработка естественного языка.

## Введение

Одним из этапов обучения системы распознавания речи является создание модели языка по обучающему текстовому корпусу. Подходы, применяемые для автоматической обработки текста на естественных языках, можно разделить на два класса: статистические и лингвистические [1]. Наиболее распространенными моделями языка в системах автоматического распознавания речи и машинного перевода являются статистические модели на основе  $n$ -грамм, цель которых состоит в оценке вероятности появления цепочки слов  $X = (W_1, W_2, \dots, W_m)$  в некотором тексте [2]. Данные модели показали свою эффективность для многих языков, но для русского они являются недостаточно действенными. Русский является синтетическим флективным языком с богатой морфологией, что приводит к существенному увеличению размера словаря системы распознавания, а также к увеличению коэффициента неопределенности (perplexity)  $n$ -граммных моделей языка (МЯ). В работе [3] было показано, что при увеличении размера словаря от 100 до

400 тыс. слов значение коэффициента неопределенности модели английского языка возрастает на 5,8%, а русского – на 39,5%.

Современной альтернативой  $n$ -граммной модели является факторная модель языка (ФМЯ), которая впервые была предложена в работе [4] для моделирования арабского языка и может быть использована для других языков с богатой морфологией. Эта модель объединяет различные признаки слова (факторы), при этом слово представляется как вектор  $k$  факторов  $Y_i = (F_i^1, F_i^2, \dots, F_i^k)$ . В качестве факторов могут использоваться: словоформа, часть речи, основа, корень слова и другие морфологические и грамматические признаки.

В работе [5] ФМЯ использовалась как на стадии распознавания речи, так и переоценки списка лучших гипотез. Результаты экспериментов показали, что при использовании модели для переоценки списка лучших гипотез количество неправильно распознанных слов уменьшилось на 0,8-1,3% в зависимости от тестового корпуса, а использование ФМЯ на этапе распознавания дало дополнительное уменьшение неправильно распознанных слов на 0,5%.

<sup>1</sup> Работа выполнена при финансовой поддержке фонда РФФИ (проекты № 15-07-04322 и 15-07-04415) и Совета по грантам Президента РФ (проекты № МК-5209.2015.8 и МД-3035.2015.8).

В [6] ФМЯ применялась для переоценки решетки гипотез распознавания. В качестве декодера использовался инструментарий НТК [7], который генерировал граф из 100 наилучших вариантов распознавания для каждого предложения с использованием биграммной МЯ и словарем объемом 5 тыс. слов. Затем была выполнена переоценка решетки с использованием морфемной и факторной МЯ. Количество неправильно распознанных слов, полученное с применением базовой МЯ, составило 8,4%. Использование ФМЯ позволило сократить количество неправильно распознанных слов до 7,08%.

В [8] ФМЯ для эстонского языка применялась для переоценки списка лучших гипотез распознавания, который был получен с использованием триграммной морфемной МЯ. Словарь состоял из 60 тыс. морфем, а ФМЯ содержала два фактора: словоформа и часть речи. Относительное уменьшение количества неправильно распознанных слов составило 7,3%.

ФМЯ используются и для распознавания мультязычной речи, в которой происходит смена языков (code-switching speech) [9]. В работе [10], помимо грамматических признаков слова, в ФМЯ были включены семантические признаки. ФМЯ была применена на стадии декодирования речи, для чего использовался декодер BioKIT [11]. Эксперименты по распознаванию речи на двух языках (английский и мандаринский китайский) показали относительное сокращение смешанной ошибки распознавания слов на 3,4%. В работе [12] для решения аналогичной задачи ФМЯ была объединена с МЯ на основе рекуррентных нейронных сетей. Полученная модель позволила сократить ошибку распознавания на 32,7% по сравнению с базовой триграммной МЯ.

Применение ФМЯ для распознавания русской речи было предложено в работе [13]. ФМЯ была обучена на текстовом корпусе, содержащем около 10 млн словоупотреблений, размер словаря составил 100 тыс. словоформ. ФМЯ создавалась с использованием следующих факторов: словоформа, лемма, метка морфологических признаков, часть речи и фактор, содержащий информацию о роде, числе и лице словоформы. Для получения факторов использовался инструментарий TreeTagger [14]. ФМЯ применялась для переоценки списка лучших 500 гипотез. В результате экспериментов было получено относительное уменьшение количе-

ства неправильно распознанных слов на 4% по сравнению с результатами, полученными при распознавании речи с базовой триграммной МЯ. После интерполяции ФМЯ с триграммной относительное уменьшение количества неправильно распознанных слов составило 6,9%.

Данная статья посвящена разработке и исследованию факторных моделей русского языка и их применению в системах автоматического распознавания речи.

## 1. Создание факторных моделей русского языка

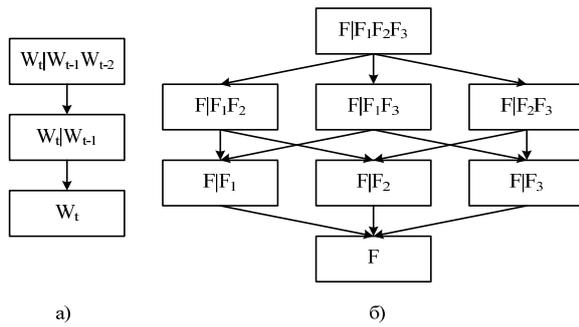
При создании ФМЯ необходимо выполнить следующие шаги [15]:

1. Выбрать подходящий набор лингвистических факторов при помощи применения методов, основанных на данных или лингвистических знаниях.

2. Построить наилучшую статистическую МЯ по этим факторам.

Одной из проблем создания статистических МЯ является недостаток обучающих данных (особенно для малоресурсных языков) [9]. Для решения этой проблемы используется методика возврата (back-off) [2], суть которой состоит в том, что, когда некоторая  $n$ -грамма отсутствует в обучающем текстовом корпусе или частота ее появления очень низкая, то вместо нее используется вероятность  $(n-1)$ -граммы, умноженная на коэффициент возврата. В  $n$ -граммных МЯ возврат осуществляется путем отбрасывания сначала наиболее дальнего слова, затем второго по дальности слова и т.д. Этот процесс показан на Рис. 1, а. Для ФМЯ нет очевидного пути возврата [4], любой фактор может быть опущен на каждом шаге выполнения процедуры возврата, и не является очевидным, какой фактор должен быть опущен первым. Таким образом, возможны несколько путей возврата, в результате получается граф возврата. Пример графа возврата ФМЯ представлен на Рис. 1, б. Граф показывает все возможные пути возврата, при которых один фактор опускается на каждом шаге процедуры возврата.

Для выбора оптимального набора факторов и пути возврата могут использоваться как лингвистические знания, так и методы, основанные на данных. В работе [5] показано, что автоматический метод, использующий генетические алгоритмы для оптимизации набора

Рис. 1. Пути возврата для  $n$ -граммной и факторной МЯ

- а) путь возврата для триграммной МЯ  
 б) граф возврата для трехфакторной МЯ

факторов, графа возврата и методик сглаживания, работает лучше, чем подбор параметров вручную, позволяя получить меньшее значение коэффициента неопределенности МЯ. Цель данного метода состоит в том, чтобы найти такую комбинацию параметров, которая создаст ФМЯ с низким коэффициентом неопределенности, вычисленным на текстовых данных, которые не использовались для обучения модели [15].

В ходе исследования для создания моделей языка был собран и автоматически обработан текстовый корпус, сформированный из Интернет-сайтов ряда электронных газет. Общий объем собранного корпуса после его обработки составил свыше 350 млн словоупотреблений, корпус содержит около 1 млн уникальных словоформ [16].

Для получения морфологических признаков словоформ русского языка в обучающих текстах использовалась программа VisualSynap проекта АОТ [17]. В работе были использованы 5 лингвистических факторов: словоформа, лемма, основа, часть речи, метка морфологических признаков. Таким образом, все слова в обучающем текстовом корпусе были заменены на факторы. Например, словоформа "схеме" заменяется на следующую последовательность факторов: "W-схеме L-схема S-схем P-сущ M-bc", где W – словоформа, L – лемма, S – основа, P – часть речи, M – метка морфологических признаков, содержащая всю грамматическую информацию о слове. В данном примере она означает, что словоформа является существительным женского рода в единственном числе и дательном падеже.

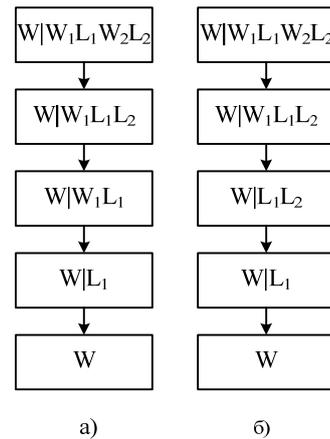


Рис. 2. Фиксированные пути возврата ФМЯ с факторами слово и лемма

- а) первый путь возврата  
 б) второй путь возврата

### 1.1. ФМЯ с фиксированным путем возврата

Для автоматизированного создания МЯ использовалась программа SRILM [18]. Вначале по обучающему текстовому корпусу были созданы двухфакторные модели, включающие словоформу и один из других факторов. Для создания этих моделей использовались 2 варианта путей возврата [19, 20]:

1. Сначала опускаются дальняя словоформа и фактор, затем – ближняя словоформа и фактор (Рис. 2, а).

2. Сначала опускаются словоформы в порядке их удаленности от рассматриваемой, а затем – факторы в том же порядке (Рис. 2, б).

При создании МЯ обычно используются различные методики сглаживания, которые применяются для того, чтобы приписать ненулевые вероятности  $n$ -граммам, не встретившимся в обучающем корпусе, за счет уменьшения вероятностей встретившихся  $n$ -грамм. Поэтому нами были исследованы ФМЯ с использованием различных методик сглаживания [21]. Применялись следующие методики: 1 – Good-Turing, 2 – Unmodified Kneser-Ney, 3 – Modified Kneser-Ney, 4 – Witten-Bell, 5 – Natural. В Табл. 1 показаны коэффициенты неопределенности (perplexity) для созданных ФМЯ. Коэффициент неопределенности был вычислен для корпуса, содержащего фразы из материалов интернет-газеты «Фонтанка.ru», объемом 33 млн словоупотреблений. У моделей, имеющих путь возврата 1, значение коэффициента неопределенности оказалось меньше, незави-

Табл. 1. Коэффициенты неопределенности для ФМЯ с различными методиками сглаживания и путями возврата

Факторы	Методики сглаживания ФМЯ									
	1		2		3		4		5	
	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2
WM	573	696	593	724	566	691	749	898	761	916
WL	557	597	550	603	<b>529</b>	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

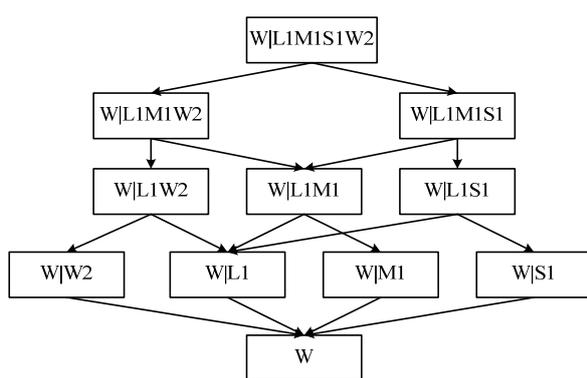


Рис. 3. Граф возврата для ФМЯ 1 с параллельными путями возврата

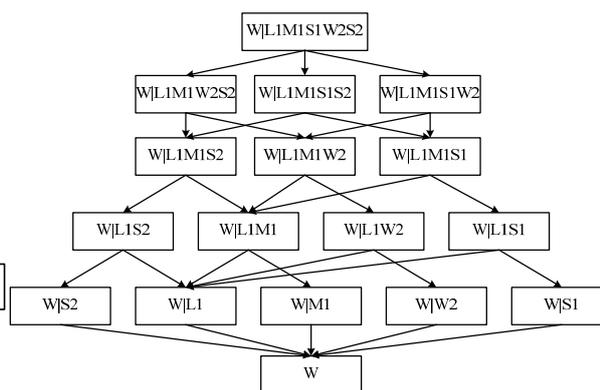


Рис. 4. Граф возврата для ФМЯ 2 с параллельными путями возврата

симо от используемых факторов и методик сглаживания. Наименьший коэффициент неопределенности имеет двухфакторная модель (словоформа и лемма), имеющая путь возврата 1 и использующая методику сглаживания 3. Применение этой методики сглаживания дало лучшие результаты для всех моделей, кроме содержащей факторы словоформа и ее части речи. Для этой модели лучший результат был получен с применением методики сглаживания 1. Для сравнения, коэффициент неопределенности базовой три-граммной модели равен 553.

## 1.2. ФМЯ с параллельными путями возврата

Для выбора наилучших лингвистических факторов и оптимального пути возврата МЯ нами использовался генетический алгоритм [22]. В качестве начальных факторов использовались перечисленные выше пять факторов (словоформа, лемма, основа, часть речи, метка морфологических признаков), временной контекст, равный двум (т.е. для вычисления вероятности появления слова использовались факторы для цепочки из двух предшествующих ему слов) и те же методики сглаживания. При этом в генетическом алго-

ритме размер популяции был равен 10, а максимальное количество генераций – 20 [23, 20].

Для экспериментов по распознаванию речи были выбраны две модели с наименьшим коэффициентом неопределенности. Обе модели содержали одни и те же 4 фактора: словоформа, лемма, морфологическая метка, основа слова и использовали три методики сглаживания (2, 3 и 4). Граф возврата для ФМЯ 1 представлен на Рис. 3, для ФМЯ 2 – на Рис. 4. Коэффициент неопределенности для ФМЯ 1 – 589, для ФМЯ 2 – 618.

## 2. Система распознавания слитной русской речи с ФМЯ

Архитектура разработанной автоматической системы распознавания слитной русской речи с внедрением ФМЯ представлена на Рис. 5. Система работает в режимах обучения моделей и распознавания (декодирования) речи. В режиме обучения создаются модели акустических единиц речи, фонематический словарь словоформ,  $n$ -граммная и факторная МЯ. В режиме распознавания входной речевой сигнал преобразуется в последовательность векторов признаков

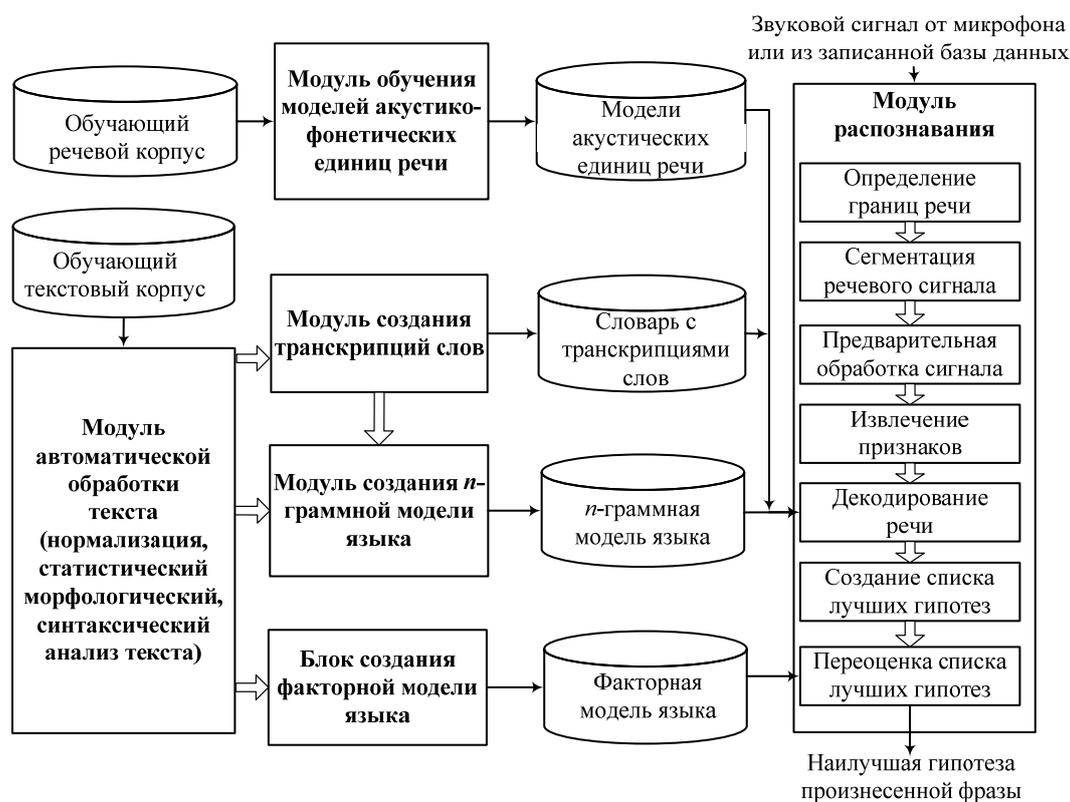


Рис. 5. Архитектура системы распознавания слитной русской речи с внедрением ФМЯ

(используются мел-частотные кепстральные коэффициенты с их двумя производными). Затем производится поиск наиболее вероятной гипотезы с использованием предварительно обученных акустических и языковых моделей. При этом ФМЯ используется на этапе постобработки для переоценки списков лучших гипотез распознавания (N-best list). Таким образом, на этапе распознавания речи используется триграммная МЯ, с помощью которой создается список лучших гипотез распознавания, а ФМЯ применяется для переоценки полученного списка гипотез и выбора наилучшей гипотезы распознавания фразы.

### 3. Экспериментальные исследования по распознаванию слитной русской речи

#### 3.1. Обучающий и тестовый корпус речи

Для обучения акустических моделей и тестирования системы распознавания речи использовались два различных речевых корпуса. Обуча-

ющий речевой корпус содержит записи произнесенных дикторами 327 фонетически сбалансированных осмысленных фраз на русском языке, каждая фраза была произнесена 50 дикторами-носителями русского языка (по 25 женщин и мужчин). Общий объем корпуса составил 13,5 Гб, длительность записей – более 21 ч. Запись проводилась в специальной звукоизолированной комнате, отношение сигнал/шум составляло 35 дБ [24]. Акустические модели создавались с помощью инструментария НТК (Hidden Markov Model Toolkit) [7]. Для извлечения информативных признаков речевого сигнала использовались энергия сигнала и 12 мелчастотных кепстральных коэффициентов с их двумя производными. В качестве акустических моделей использовались лево-правые скрытые марковские модели с тремя состояниями.

Для тестирования системы использовался корпус, содержащий 100 слитно произнесенных фраз, которые взяты из материалов интернет-газеты «Фонтанка.ru». Каждая фраза была произнесена пятью дикторами. Длина каждого предложения составила от 6 до 20 слов. Сессия записи для каждого диктора длилась от 20 до 40 мин,

при этом чистая речь составила 15-30 мин. Общий объем корпуса – 200 Мб аудиоданных.

### 3.2. Исследование ФМЯ с фиксированным путем возврата

Система автоматического распознавания слитной русской речи была построена на основе декодера речи Julius [25]. Оценка качества распознавания проводилась по критерию качества распознавания речи с использованием показателя процента неправильно распознанных слов в речи (WER — Word Error Rate).

На этапе декодирования речи использовалась триграммная МЯ с применением модели WER=26,54%. Размер словаря распознавания (лексикон) составил 150 тыс. слов, количество отсутствующих в словаре слов - 1,1%. Системой распознавания речи был создан список лучших 20 гипотез распознавания для каждой произнесенной фразы, и проведена переоценка данного списка с использованием ФМЯ. Результаты представлены в Табл. 2.

Наилучшие результаты достигнуты при применении ФМЯ, содержащей словоформы и метки морфологических признаков, созданной с помощью методики сглаживания 4 (Witten-Bell).

Количество неправильно распознанных слов - 27,3%. Таким образом, после переоценки списка лучших 20 гипотез процент ошибок увеличился. Поэтому была выполнена интерполяция факторных моделей с базовой триграммной моделью. Наименьший процент неправильно распознанных слов (WER=24,44%) был получен для базовой модели, линейно интерполированной с факторной моделью, в которой использовались словоформы и морфологические таги (метки). Эта модель была создана с применением методики сглаживания 3 (Modified Kneser-Ney) и пути возврата 1.

Затем была выполнена переоценка списков лучших 10 и 50 гипотез распознавания с использованием ФМЯ с методикой сглаживания 3, интерполированная с триграммной моделью. Результаты представлены в Табл. 3, где показано минимально достижимое количество неправильно распознанных слов для каждого списка гипотез (Oracle WER) - минимальное значение процента неправильно распознанных слов, которое можно получить, выбрав максимально точную гипотезу из данного списка. Из Табл. 3 видно, что наилучшие результаты были получены после переоценки списка из 20 гипотез.

Табл.2. Количество неправильно распознанных слов (WER, %), полученное после переоценки списка 20 лучших гипотез ФМЯ

ФМЯ	Методики сглаживания МЯ									
	1		2		3		4		5	
	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2
WM	27,87	28,00	27,79	28,09	28,15	28,16	<b>27,30</b>	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
Интерполированные модели										
WM+3-гр.	25,00	24,89	24,57	24,93	<b>24,44</b>	24,78	24,94	25,22	25,41	25,36
WL+3-гр.	25,51	25,54	25,54	25,67	25,58	25,43	25,51	25,47	24,98	25,47
WP+3-гр.	25,21	25,28	25,30	25,32	25,07	25,24	25,47	25,64	25,60	25,43
WS+3-гр.	25,97	25,92	26,03	25,86	25,88	25,90	26,05	25,86	25,49	25,90

Табл.3. Количество неправильно распознанных слов (WER, %), полученное после переоценки списков N лучших гипотез

Модели языка	N=10		N=20		N=50	
	Путь 1	Путь 2	Путь 1	Путь 2	Путь 1	Путь 2
3-граммная (oracle WER)	18,52		16,63		15,34	
3-граммная+WM	24,83	24,94	<b>24,44</b>	24,78	24,55	24,66
3-граммная+WL	25,79	25,71	25,58	25,43	25,60	25,37
3-граммная+WP	25,43	25,54	25,07	25,24	25,15	25,26
3-граммная+WS	25,82	26,01	25,88	25,90	25,90	26,10

### 3.3. Исследование ФМЯ с параллельными путями возврата

Авторами были выполнены эксперименты по переоценке списков лучших 20 гипотез с использованием ФМЯ с параллельными путями возврата (Табл. 4). Применение этих моделей не привело к снижению ошибок распознавания, поэтому они были интерполированы с триграммной моделью. Наилучшие результаты получены при интерполяции базовой модели с ФМЯ 1, количество неправильно распознанных слов составило 24,53%.

Табл. 4. Количество неправильно распознанных слов (WER, %) после переоценки списка из 20 лучших гипотез ФМЯ с параллельными путями возврата

Модели языка	WER, %
Триграммная	26,54
ФМЯ 1	27,94
ФМЯ 2	28,56
ФМЯ 1 + триграммная	<b>24,53</b>
ФМЯ 2 + триграммная	24,74

Из Табл.4 видно, что количество неправильно распознанных слов, полученное после применения моделей с параллельными путями возврата, незначительно увеличилось по сравнению с результатами, полученными после применения моделей с фиксированным путем возврата. Это связано с тем, что модели с параллельными путями возврата были обучены только на части корпуса (100 млн словоупотреблений). Для обучения такой модели на сервере потребовалось 64 Гб оперативной памяти. Таким образом, недостатком ФМЯ с большим числом факторов и параллельными путями возврата является то, что для их обучения требуется большой объем оперативной памяти. Однако даже при обучении моделей на небольшом текстовом корпусе можно снизить ошибки распознавания, что является несомненным достоинством применения статистических моделей языка данного типа.

### Заключение

Проведенное исследование ФМЯ показало, что включение дополнительной лингвистической информации в модель языка позволяет улучшить качество автоматического распознавания речи. Выполнено сравнение влияния различного набора

факторов на точность распознавания речи. Относительное уменьшение количества неправильно распознанных слов по сравнению с результатами, полученными с использованием триграммной модели, составило 8%. В дальнейшей работе планируется исследовать ФМЯ с дополнительными факторами, а также другие типы статистических моделей русского языка.

### Литература

1. Смирнов И.В., Шелманов А.О. Семантико-синтаксический анализ естественных языков. Часть I. Обзор методов синтаксического и семантического анализа текстов // Искусственный интеллект и принятие решений, № 1, 2013, С. 41–54.
2. Moore G.L. Adaptive Statistical Class-based Language Modelling. PhD thesis, Cambridge University, 2001. 193 p.
3. Whittaker E.W.D., Woodland P.C. Language modelling for Russian and English using words and classes // Computer Speech and Language, Vol. 17, 2003, pp. 87-104.
4. Bilmes J. A., Kirchoff K. Factored language models and generalized parallel backoff. In Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Stroudsburg, PA, USA, Vol. 2, 2003, pp. 4–6.
5. Vergyri D., Kirchoff K., Duh K., Stolcke A. Morphology-Based Language Modeling for Arabic Speech Recognition. In Proceedings of ICSLP-2004, 2004, pp. 2245-2248.
6. Tachbelie M.Y., Teferra Abate S., Menzel W. Morpheme-based language modeling for Amharic speech recognition. In Proc. 4th Language and Technology Conference LTC-2009, Poznan, Poland, 2009, pp. 114-118.
7. Young S. et al. The HTK book, Cambridge Univ. Press, 2009. 384 p.
8. Alumae T. Sentence-adapted factored language model for transcribing Estonian speech. In Proc. ICASSP-2006, Toulouse, France, 2006, pp. 429–432.
9. Карпов А.А., Верходанова В.О. Речевые технологии для малоресурсных языков мира // Вопросы языкознания. № 3. М.: Наука, 2015, С. 117-135.
10. Adel H., Kirchoff K., Telaar D., Vu N. T., Schlippe T., Schultz T. Features for factored language models for code-switching speech. In Proc. 4th International Workshop on Spoken Language Technologies for Under-resourced languages (SLTU-2014), St. Petersburg, Russia, 2014, pp. 32-38.
11. 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. BioKIT - Real-time decoder for bi-signal processing. In Proceedings of Interspeech-2014, Singapore, 2014, pp. 2650-2654.
12. Adel H., Vu N. T., Schultz T. Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria, 2013.
13. Vazhenina D., Markov K. Factored Language Modeling for Russian LVCSR. In Proc. International Joint Conference on Awareness Science and Technology & Ubi-Media Computing, Aizu-Wakamatsu, Japan, 2013, pp. 205-210.

14. Schmid H. Probabilistic part-of-speech tagging using decision trees. In Proc. International Conference on New Methods of Language Processing, Manchester, UK, 1994, pp. 44–49.
15. Kirchhoff K., Bilmes J., Duh K. Factored Language Models Tutorial. Tech. Report UWEEETR-2007-0003, Department of Electrical Engineering, University of Washington, June 2007.
16. Кипяткова И.С., Карпов А.А.. Автоматическая обработка и статистический анализ новостного текстового корпуса для модели языка системы распознавания русской речи // Информационно-управляющие системы, № 4(47), 2010, С. 2-8.
17. Сокирко А.В. Морфологические модули на сайте [www.aot.ru](http://www.aot.ru) // Труды Международной конференции «Диалог-2004», Москва, 2004. С. 559-564.
18. Stolcke A., Zheng J., Wang W., Abrash V. SRILM at Sixteen: Update and Outlook. In Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop ASRU'2011. Waikoloa, Hawaii, USA, 2011.
19. Kipyatkova I., Verkhodanova V., Karpov A. Rescoring N-best lists for Russian speech recognition using factored language models. In Proc. 4th International Workshop on Spoken Language Technologies for Under-resourced languages (SLTU-2014), St. Petersburg, Russia, 2014, pp. 81-86.
20. Kipyatkova I., Karpov A. Development of Factored Language Models for Automatic Russian Speech Recognition // Компьютерная лингвистика и интеллектуальные технологии: По материалам Международной конференции «Диалог», Вып. 14, М., 2015, С. 234-246.
21. Whittaker E.W.D., Statistical Language Modelling for Automatic Speech Recognition of Russian and English. PhD thesis, Cambridge University, 2000, 140 p.
22. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. 2-е изд. М.: Физматлит, 2006, 320 с.
23. Kipyatkova I., Karpov A. Study of Morphological Factors of Factored Language Models for Russian ASR // Springer International Publishing Switzerland. A. Ronzhin et al. (Eds.): SPECOM 2014, LNAI 8773, 2014, pp. 451–458.
24. Karpov A., Markov K., Kipyatkova I., Vazhenina D., Ronzhin A. Large vocabulary Russian speech recognition using syntactico-statistical language modeling // Speech Communication. Elsevier, 2014, Vol. 56, pp. 213-228.
25. Lee A., Kawahara T. Recent Development of Open-Source Speech Recognition Engine Julius. In Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2009), Sapporo, Japan, 2009, pp.131–137.

**Кипяткова Ирина Сергеевна.** Старший научный сотрудник лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Окончила Санкт-Петербургский государственный университет аэрокосмического приборостроения в 2008 году. Кандидат технических наук. Автор 55 научных публикаций. Область научных интересов: автоматическое распознавание речи, человеко-машинное взаимодействие. E-mail: [kipyatkova@iias.spb.su](mailto:kipyatkova@iias.spb.su)

**Карпов Алексей Анатольевич.** Заведующий лабораторией речевых и многомодальных интерфейсов СПИИРАН. Окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения в 2002 году. Доктор технических наук. Автор более 200 научных публикаций и трех монографий. Область научных интересов: речевые технологии, аудио-визуальная обработка речи, многомодальные человеко-машинные интерфейсы. E-mail: [karpov@iias.spb.su](mailto:karpov@iias.spb.su)