

Назаретян Нелли Вартановна

### **АЛГОРИТМ ИДЕНТИФИКАЦИИ РЕЧЕВЫХ ФРЕЙМОВ**

Статья предполагает изучение формально-семантических признаков, присутствующих в предложениях с прямой речью, с целью составления на их основе алгоритма для автоматической обработки таких предложений. Основное содержание статьи составляет подробное описание каждого шага составляемого алгоритма на естественном языке. Результатом работы стало построение алгоритма, с помощью которого некая автоматизированная система сможет обрабатывать заданный публицистический текст, находить в нем речевые фреймы и выводить значения определенных параметров в соответствующие слоты.

Адрес статьи: [www.gramota.net/materials/2/2018/2-2/33.html](http://www.gramota.net/materials/2/2018/2-2/33.html)

Источник

### **Филологические науки. Вопросы теории и практики**

Тамбов: Грамота, 2018. № 2(80). Ч. 2. С. 344-350. ISSN 1997-2911.

Адрес журнала: [www.gramota.net/editions/2.html](http://www.gramota.net/editions/2.html)

Содержание данного номера журнала: [www.gramota.net/materials/2/2018/2-2/](http://www.gramota.net/materials/2/2018/2-2/)

### **© Издательство "Грамота"**

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: [www.gramota.net](http://www.gramota.net)

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: [phil@gramota.net](mailto:phil@gramota.net)

УДК 81

*Статья предполагает изучение формально-семантических признаков, присутствующих в предложениях с прямой речью, с целью составления на их основе алгоритма для автоматической обработки таких предложений. Основное содержание статьи составляет подробное описание каждого шага составляемого алгоритма на естественном языке. Результатом работы стало построение алгоритма, с помощью которого некая автоматизированная система сможет обрабатывать заданный публицистический текст, находить в нем речевые фреймы и выводить значения определенных параметров в соответствующие слоты.*

*Ключевые слова и фразы:* автоматическая обработка текста; алгоритм; иллокутивный глагол; прямая речь; публицистический дискурс; семантические признаки; речевой фрейм; слот; формальные признаки.

**Назаретян Нелли Варгановна**

*Северо-Кавказский федеральный университет, г. Ставрополь  
ursa777@yandex.ru*

### АЛГОРИТМ ИДЕНТИФИКАЦИИ РЕЧЕВЫХ ФРЕЙМОВ

Стремительно возрастающие потоки языковой информации, представленные в виде электронных источников, а также желание автоматизировать процесс перевода текстов с одного языка на другой привели к идее о создании автоматизированных систем обработки больших текстовых массивов с целью извлечения из текстов семантической информации по заранее заданным параметрам.

Системы распознавания фреймов, репрезентированных в текстах, существенно облегчают процесс извлечения семантической информации, из чего следует, что создание алгоритмов идентификации и формирования фреймов представляет собой одну из наиболее актуальных задач современной когнитивной лингвистики.

Настоящая статья посвящена написанию алгоритма, с помощью которого некая автоматизированная система сможет обрабатывать заданный публицистический текст, находить в нем речевые фреймы и выводить значения определенных параметров в соответствующие слоты.

В тексте нашего исследования мы ограничимся словесным способом записи алгоритмов, последовательно описывая этапы обработки данных в произвольном изложении на естественном языке.

Алгоритм идентификации речевых фреймов должен быть ориентирован, прежде всего, на поиск иллокутивного глагола в заданном тексте. При составлении алгоритма мы будем работать только с теми предложениями, в которых есть прямая речь, так как в рамках нашего исследования мы работаем с публицистическим текстом и нас интересует иллокутивный глагол не в речи персонажа, а в речи автора статьи.

Для точной идентификации предложений с прямой речью необходимо знать формальные признаки, обозначающие прямую речь, главным из которых, безусловно, является наличие определенных знаков препинания, различающих речь автора и речь персонажа. Таким образом, составление алгоритма следует начать с поиска предложений с этими знаками препинания в заданном тексте. Постановка знаков препинания в предложениях с прямой речью зависит, во-первых, от оформления прямой речи в печатных текстах. В русском языке существует два варианта оформления прямой речи на письме.

#### 1. Выделение прямой речи в абзац.

Если запись прямой речи идет с новой строки (с абзаца), то перед прямой речью ставится знак тире (–):

– *Несовершеннолетние, независимо от возраста, могут доставляться в ОВД для выяснения всех обстоятельств произошедшего, – поясняет старший помощник прокурора Ростовской области Оксана Сухарева [2].*

#### 2. Запись прямой речи в строку, без выделения в абзац.

В случае записи прямой речи в строку в начале и в конце прямой речи ставятся кавычки (« »):

*Для лидеров это уже вторая встреча в этом году. Приоритетное внимание Владимир Путин и Милош Земан уделили экономическим связям. «В этом году общими усилиями удалось преодолеть спад во взаимной торговле», – сказал журналистам Путин [9].*

Во-вторых, постановка знаков препинания в предложениях с прямой речью зависит от соотношения прямой речи и слов автора.

В русском языке существует 4 варианта соотношения слов автора и прямой речи.

#### 1. Прямая речь после слов автора.

Если слова автора предшествуют прямой речи, после них ставится двоеточие, а прямая речь заключается в кавычки. Первое слово прямой речи пишется с прописной буквы, в конце прямой речи употребляется соответствующий знак конца предложения. При этом вопросительный и восклицательный знаки, а также многоточие ставятся перед кавычками, точка – после них:

*На вопрос журналиста о том, что было бы, знай он в момент ареста о своей судьбе в течение последующих десяти лет, Ходорковский ответил: «Тогда? Боюсь, застрелился бы. Для меня тогдашнего мой нынешний опыт стал бы шоком» [4].*

#### 2. Прямая речь перед словами автора.

Если прямая речь находится перед словами автора, то она заключается в кавычки и пишется с прописной буквы, после неё ставится запятая и тире. В случае если прямая речь представляет собой вопрос, восклицание

или незаконченное предложение, то перед закрывающей кавычкой ставится соответствующий знак, а после кавычек – тире. Слова автора пишутся со строчной буквы:

*«Если мы хотим быть впереди, если мы хотим побеждать, нужно быть лучше», – заявил президент участникам совещания [8].*

### 3. Слова автора внутри прямой речи.

Если слова автора разрывают прямую речь, то возможны три варианта постановки знаков препинания:

а) если прямая речь представляет собой одно предложение, то после первой её части ставятся запятая и тире, слова автора пишутся со строчной буквы, после них ставятся запятая и тире, вторая часть прямой речи пишется со строчной буквы; кавычки ставятся только в начале и в конце прямой речи и не ставятся между прямой речью и словами автора. Завершается прямая речь соответствующим знаком. При этом вопросительный и восклицательный знаки, а также многоточие ставятся перед кавычками, точка – после них:

*«Пришло время понять, – обратился Ходорковский к россиянам, – что нефтяная халва заканчивается, и мы, граждане России, должны своими руками и своим умом изменить нашу страну» [4];*

б) если прямая речь состоит из нескольких предложений и слова автора стоят между ними, после первой части прямой речи ставится запятая и тире (если в конце предложения прямой речи должна стоять точка), восклицательный, вопросительный знак или многоточие и тире; слова автора пишутся со строчной буквы, после них ставятся точка и тире; вторая часть прямой речи начинается с прописной буквы. Кавычки ставятся только в начале и в конце прямой речи. Знаки препинания в конце второй части прямой речи регулируются правилами, описанными выше:

*«Впервые за три предыдущих года мы принимаем бюджет, где у нас рост более 2%. Бюджет роста, – заявил Володин. – И мы еще подчеркиваем, что впервые в бюджете мы видим рост зарплаты, рост денежного довольствия военнослужащих, реализация указов президента, которая нацелена именно на решение этих задач» [3];*

в) если в словах автора внутри прямой речи имеются два иллокутивных глагола, из которых один относится к первой части прямой речи, а другой – ко второй, то после слов автора ставятся двоеточие и тире, причем первое слово второй части пишется с прописной буквы:

*«Нет, ничего, прекрасно, – отвечал Павел Петрович и погодя немного прибавил: – Брата не обманешь, надо будет сказать ему, что мы повздорили из-за политики» [14, с. 187].*

### 4. Прямая речь внутри слов автора.

Если прямая речь находится внутри слов автора, то перед ней после слов автора ставится двоеточие, прямая речь заключается в кавычки, а после неё ставится тире или запятая или запятая и тире (в зависимости от контекста), слова автора пишутся со строчной буквы:

*В первые часы после провала футболист вообще не стал комментировать результат матча с Грецией: «Пока мне нечего сказать», – приводит слова Аршавина его официальный сайт [1].*

Рассмотрев правила расстановки знаков препинания в предложениях с прямой речью, мы можем перечислить все знаки, которые встречаются в таких предложениях: –; « »; ; ; !; ?; ...; .; ,.

Найденные по отдельности, перечисленные знаки препинания будут неинформативны для составляемого нами алгоритма.

Вероятность нахождения интересующих нас предложений возрастет во много раз, если мы будем искать не один знак препинания, а определенное сочетание знаков, характерное для предложений с прямой речью. Рассмотрим сочетания знаков, которые мы можем встретить в предложениях с прямой речью: , –; –; : « » –; ! –; ? –; ... –; : –.

Принимая во внимание тот факт, что в русском языке существует четыре варианта соотношения слов автора и прямой речи, мы разделили исполнение алгоритма на два блока, совместив попарно варианты, обладающие схожими формальными признаками. В первом блоке мы будем обрабатывать предложения, в которых прямая речь предшествует словам автора, и предложения, в которых слова автора находятся внутри прямой речи, а во втором блоке будем обрабатывать предложения, в которых прямая речь следует после слов автора или находится внутри слов автора.

При вводе публицистического текста на первом этапе в первом блоке задаем условие: «Если в предложении есть знаки “ ” –; “ ”, –; , –; ? –; ! –; ... –; : –, то перейти к следующему шагу алгоритма, иначе перейти ко второму блоку алгоритма и искать в предложении знаки : “ “. Если в предложении есть знаки : “ “, то перейти к следующему шагу алгоритма, иначе вернуться к первому условию и повторить команды, проверяя следующее предложение и так до тех пор, пока не будет обработан весь заданный текст. Когда все предложения заданного текста будут проверены, прекратить исполнение алгоритма, перейдя к команде “Выход”».

Далее переходим ко второму этапу – поиску иллокутивного глагола в предложении с прямой речью. Формальным признаком, ориентируясь на который можно идентифицировать иллокутивный глагол в предложении, является его положение, а также наличие окончания, характерного для глагола. В предложениях с прямой речью есть 4 варианта положения иллокутивного глагола соответственно вариантам соотношения слов автора и прямой речи.

1. Иллокутивный глагол может предшествовать прямой речи (в случае, когда слова автора предшествуют прямой речи):

*И вот, что она рассказала: «Мы живем в доме на двух хозяев, и наша половина – в критическом состоянии, разрушается фундамент...» [13].*

Иллокутивный глагол может следовать за прямой речью (в случае, когда прямая речь предшествует словам автора):

«Когда у вас больше пяти тысяч вопросов, становится бессмысленно их зубрить, легче выучить школьный курс», – пояснил господин Кузьминов [15, с. 5].

2. Иллокутивный глагол может разделять прямую речь на две части (в случае, когда слова автора находятся внутри прямой речи):

«Пришло время понять, – обратился Ходорковский к россиянам, – что нефтяная халява заканчивается и мы, граждане России, должны своими руками и своим умом изменить нашу страну» [4].

3. В предложении с прямой речью может быть два иллокутивных глагола, первый из которых относится к прямой речи, стоящей перед словами автора, второй – к прямой речи после слов автора (в случае, когда прямая речь заключена внутри слов автора):

«Нет, ничего, прекрасно, – отвечал Павел Петрович и погода немного прибавил: – Брата не обманешь, надо будет сказать ему, что мы повздорили из-за политики» [14, с. 187].

В предложениях, в которых иллокутивный глагол следует за прямой речью или разделяет ее на две части, перед иллокутивным глаголом ставится знак «тире». Таким образом, при поиске иллокутивного глагола в первом блоке нам нужно вывести в слот «иллокутивный глагол» слово, которое следует за знаком «тире», предварительно убедившись, что это слово является глаголом. В публицистическом дискурсе, в предложениях с прямой речью глаголы употреблены преимущественно в форме прошедшего времени, а также в форме настоящего времени. Употребление иллокутивного глагола в форме будущего времени крайне редко и характерно для художественного дискурса, поэтому при написании алгоритма мы не будем принимать во внимание этот факт. Иллокутивный глагол в словах автора может быть в форме 3 лица. Примем во внимание и тот факт, что глагол обладает категорией числа, а в прошедшем времени, в единственном числе – категорией рода. Итак, мы имеем 5 вариантов грамматических форм иллокутивного глагола, которые могут встречаться в публицистическом дискурсе в словах автора:

- 1) глагол в настоящем времени, 3-м лице, единственном числе;
- 2) глагол в настоящем времени, 3-м лице, множественном числе;
- 3) глагол в прошедшем времени, единственном числе, женском роде;
- 4) глагол в прошедшем времени, единственном числе, мужском роде;
- 5) глагол в прошедшем времени, 3-м лице, множественном числе.

Для того чтобы проверить, является ли слово после знака «тире» в обрабатываемом предложении глаголом, мы сопоставим окончание этого слова с возможными вариантами окончаний глагола. Приведем окончания, характерные для глагола в перечисленных грамматических формах: *ет, ёт, ит, ут, ют, ат, ят*; суффикс *л*, суффикс *л* + окончание *а*, суффикс *л* + окончание *и*.

Таким образом, второе условие в первом блоке алгоритма будет звучать так: «Если слово после знака “тире” заканчивается на: *ет, ёт, ит, ут, ют, ат, ят, л, ла* или *ли*, то вывести это слово в слот “иллокутивный глагол”, иначе перейти к проверке первого условия во втором блоке».

В предложениях, где иллокутивный глагол предшествует прямой речи, а также в предложениях, где есть два иллокутивных глагола, и один из них относится к прямой речи, которая следует после слов автора, идентифицировать его сложнее. Дело в том, что иллокутивный глагол не всегда стоит непосредственно перед знаком «двоеточие», часто он может быть отделен от знака «двоеточие» другими словами, например:

*О своём заключении Ходорковский говорит без негатива: «Тюрьмы не надо бояться, абсолютное большинство здесь живёт “нормально” – без голода и побоев»* [4].

Перед нами встает вопрос о том, как искать иллокутивный глагол в таких случаях. Решение этого вопроса мы видим в том, чтобы при исполнении алгоритма во втором блоке искать иллокутивный глагол не непосредственно перед знаком «двоеточие», а в определенной части предложения. Так же как и в первом блоке, искать глагол мы будем, ориентируясь на возможные варианты окончаний.

Второе условие во втором блоке алгоритма звучит так: «Если в части предложения от начала до знака “двоеточие” или от второго слова после знака “тире” до знака “двоеточие” есть слово с окончанием *ет, ёт, ит, ут, ют, ат, ят, ла, л, ли*, то вывести это слово в слот “иллокутивный глагол”, иначе перейти к проверке следующего предложения или завершить исполнение алгоритма».

Перейдем к следующему этапу алгоритма – поиску слов, обозначающих участников речевого акта. Мы снова просмотрели отобранные нами в случайном порядке 100 статей из качественных изданий и 100 статей из «желтой» прессы с целью найти и описать формальные признаки, с помощью которых автоматизированная система сможет находить в тексте слова, обозначающие автора прямой речи. Мы выделили из текстов только предложения, содержащие прямую речь, затем мы выделили в каждом предложении слово или слова, обозначающие автора прямой речи, и на основании этих слов разделили предложения на категории.

Общее количество предложений с прямой речью в качественных изданиях составило 250 предложений. Ниже приведем категории, которые мы обозначили в качественных изданиях, по убыванию количества предложений.

1. Автор прямой речи обозначен именем, фамилией или именем и фамилией – 64 предложения (25,5%).
2. Автор прямой речи обозначен должностью, профессией и фамилией или должностью, профессией, именем и фамилией – 55 предложений (22%).

3. Автор прямой речи обозначен профессией, должностью или другим социальным статусом – 52 предложения (21%).

4. Автор прямой речи обозначен местоимением – 37 предложений (14,5%).

5. Автор прямой речи обозначен словом «господин» или «госпожа» в сочетании с фамилией – 22 предложения (9%).

6. Автор прямой речи анонимен – 10 предложений (4%).

7. Автор прямой речи обозначен названием учреждения – 10 предложений (4%).

Среди предложений, выделенных из статей качественных изданий, наибольшее число составили предложения, в которых автор прямой речи обозначен именем, фамилией или именем и фамилией.

Общее количество предложений, содержащих прямую речь в «желтой» прессе, составило 160 предложений. Приведем категории, которые мы обозначили, в порядке убывания количества предложений.

1. Автор прямой речи обозначен именем, фамилией или именем и фамилией – 61 предложение (38,5%).

2. Автор прямой речи обозначен должностью, профессией и фамилией или должностью, профессией, именем и фамилией – 42 предложения (26,5%).

3. Автор прямой речи обозначен профессией, должностью или другим социальным статусом – 33 предложения (20,5%).

4. Автор прямой речи обозначен местоимением – 19 предложений (11,5%).

5. Автор прямой речи обозначен названием учреждения – 4 предложения (2,5%).

6. Автор прямой речи анонимен – 1 предложение (0,5%).

В «желтой» прессе, так же как и в качественных изданиях, наибольшее число составили предложения, в которых автор прямой речи обозначен именем, фамилией или именем и фамилией.

Итак, первые четыре категории предложений расположены в одинаковом порядке как в качественных изданиях, так и в «желтой» прессе. Существенные различия мы видим, прежде всего, в общем количестве предложений, содержащих прямую речь. В 100 статьях из качественных изданий мы выявили 250 предложений с прямой речью, а в «желтой» прессе таких предложений оказалось 160. Это различие легко объяснить, принимая во внимание тот факт, что в «желтой» прессе все больше проявляется тенденция замещать текстовую информацию визуальной. «Желтые» газеты ориентированы на массовую аудиторию и поэтому стараются составлять максимально простые для восприятия тексты развлекательного характера, заменяя текст информационной графикой.

Далее нам необходимо выяснить, есть ли в каждой выявленной категории какие-либо формальные признаки, по которым мы можем составить команду алгоритма, которая будет идентифицировать и выводить в определенный слот слово или слова, обозначающие автора прямой речи.

Во всех категориях предложений, содержащих прямую речь, при поиске слов, обозначающих автора прямой речи, помимо других формальных признаков можно ориентироваться на порядок слов в предложении. Такой порядок слов, при котором слово или слова, обозначающие автора прямой речи, стоят непосредственно рядом с иллокутивным глаголом, относящимся к прямой речи (перед иллокутивным глаголом, если иллокутивный глагол предшествует прямой речи, либо после иллокутивного глагола, если он следует за прямой речью или разделяет ее на две части), встречается в подавляющем большинстве случаев, что позволяет нам использовать этот факт при поиске данного слова или слов. Но тем не менее встречаются предложения, в которых слова, обозначающие автора прямой речи, отделены от иллокутивного глагола другим словом. В качестве примера рассмотрим следующее предложение:

*Когда В. Путин слушал доклад о причинах наводнения, он несколько раз переспрашивал: «Так был сброс из водохранилища или не был?», «Было оповещение о паводке или не было?» [5, с. 6].*

Учитывая этот факт, мы будем задавать условие так, чтобы искать автора прямой речи во всем отрезке от иллокутивного глагола до запятой или точки или от начала предложения до иллокутивного глагола, в зависимости от положения иллокутивного глагола.

При написании алгоритма на этом этапе расположим команды таким образом, чтобы максимально полно и правильно заполнить слот «автор прямой речи» и свести возможные ошибки и погрешности к минимуму.

Первое условие на данном этапе в первом блоке будет звучать так: «Если в отрезке предложения от иллокутивного глагола до знака “запятая” или “точка” есть слово “он”, “она” или “они”, то вывести это слово в слот “Автор прямой речи”, иначе перейти к следующему условию».

Во втором блоке условие будет звучать так: «Если в отрезке от начала предложения до иллокутивного глагола есть слово “он”, “она” или “они”, то вывести это слово в слот “Автор прямой речи”, иначе перейти к следующему условию».

Далее напишем условие для тех предложений, в которых автор прямой речи обозначен словом «господин» или «госпожа» в сочетании с фамилией.

Учтем, что в некоторых случаях слово «господин» или «госпожа» заменяются сокращением «г-н» или «г-жа». Итак, следующее условие в первом блоке будет звучать так: «Если в отрезке предложения от иллокутивного глагола до знака “запятая” или “точка” есть слово “г-н”, “г-жа”, “господин” или “госпожа”, то вывести это слово и слово после него в слот “Автор прямой речи”, иначе перейти к следующему условию».

Аналогичное условие зададим для предложений, обрабатываемых во втором блоке, переместив поиск нужных слов в отрезок предложения от его начала до иллокутивного глагола.

Следующее условие будет направлено на поиск имени и фамилии, обозначающих автора прямой речи. Формальным признаком, который поможет выявить автора прямой речи, будет заглавная буква, с которой начинается имя или фамилия. Для того чтобы избежать ошибок и потери части информации при заполнении слота «Автор прямой речи», данное условие следует написать таким образом, чтобы искать слова с заглавной буквы непосредственно перед иллокутивным глаголом или после него.

В качестве примера рассмотрим следующее предложение:

*«– Сейчас вовсю идет строительство центра ядерной высокотехнологичной медицины в Димитровграде, поскольку рядом находится предприятие НИИ атомных реакторов, – рассказывает декан факультета экспериментальной и теоретической физики Национального исследовательского ядерного университета “МИФИ” профессор Владимир Беляев» [6, с. 38].*

Если мы зададим условие так, чтобы искать слово с заглавной буквы на всем отрезке от иллокутивного глагола до знака «запятая» или «точка», то при обработке приведенного выше предложения слот «Автор прямой речи» будет заполнен словом «Национального».

Таким образом, в первом блоке третье условие будет звучать так: «Если слово после иллокутивного глагола начинается с заглавной буквы, то вывести его в слот “Автор прямой речи” и проверить следующее за ним слово – если оно начинается с заглавной буквы, то и его вывести в слот “Автор прямой речи”, иначе перейти к следующему условию».

Во втором блоке зададим аналогичное условие, но искать слова с заглавной буквы будем перед иллокутивным глаголом.

Следующее условие на данном этапе будет направлено на обработку предложений, в которых автор прямой речи анонимен. Среди рассмотренных нами примеров газетных статей анонимность автора отмечается словами «собеседник», «собеседник в (название учреждения или организации)», «собеседник издания», «собеседник (“название издания)”», «источник», «источник в (название учреждения или организации)», «источник (“название издания)”», «один из источников», «один из источников в (название учреждения или организации)», «один из источников (“название издания)”».

Для данной категории предложений мы напишем ряд условий, которые поочередно будут проверять наличие одного из слов или фраз, перечисленных выше в нужной части предложения, в котором была идентифицирована прямая речь.

В первом блоке условие будет звучать следующим образом: «Если после иллокутивного глагола есть слово “собеседник”, “источник” или фраза “один из источников”, то вывести их в слот “Автор прямой речи”, иначе перейти к следующему условию».

«Далее проверить, если после этих слов есть предлог “в”, то вывести этот предлог и слово после него в слот “Автор прямой речи”, иначе проверить, есть ли после этих слов слово “издания”, если есть, то вывести это слово в слот “Автор прямой речи”, иначе проверить есть ли знак “кавычки” после этих слов, если есть, то вывести слово в кавычках в слот “Автор прямой речи” и перейти к следующему шагу алгоритма».

Во втором блоке зададим аналогичные условия для проверки части предложения от начала до иллокутивного глагола.

Далее напишем условие для обработки предложений, в которых автор прямой речи обозначен названием учреждения или организации. В качестве ключевого формального признака используем предлог «в», который в большинстве случаев связывает иллокутивный глагол и название учреждения:

*«– У пострадавших черепно-мозговые травмы, ушибы, рваные раны, у одного мужчины перелом конечностей, состояние средней тяжести, – сообщили в областном Центре медицины катастроф» [7].*

При написании предыдущих условий мы могли прогнозировать, сколько слов нужно вывести в слот «Автор прямой речи». Если автор прямой речи обозначен местоимением, очевидно, что это будет одно слово, если словами «господин» или «госпожа», нам ясно, что в рассматриваемый слот нужно вывести два слова, так как далее всегда следует фамилия. Что касается данной категории предложений, то мы не можем заранее предугадать, сколько слов нужно вывести в слот «Автор прямой речи», поэтому мы выведем всю фразу после иллокутивного глагола или всю фразу до иллокутивного глагола (в зависимости от его положения).

Итак, следующее условие в первом блоке: «Если после иллокутивного глагола есть предлог “в”, то вывести всю фразу после иллокутивного глагола до знака “точка” или “запятая”, иначе перейти к следующему условию».

Во втором блоке мы будем искать предлог «в» в части предложения от начала до иллокутивного глагола, а в слот «Автор прямой речи» будем выводить фразу от предлога «в» до иллокутивного глагола.

В рассматриваемой категории в качестве автора прямой речи часто выступают такие организации, как компания, пресс-служба, прокуратура или информационные агентства:

*«Но все же мы расстроены и находимся в неловком положении, мы удвоим усилия, чтобы гарантировать, что подобное никогда не повторится», – заявляет компания, уже рассылающая пострадавшим пользователям уведомления о случившемся [12].*

Поэтому для данной категории предложений мы напишем еще одно условие. Для первого блока: «Если после иллокутивного глагола есть слово “компания”, “прокуратура”, “пресс-служба”, то вывести это слово в слот “Автор прямой речи”, иначе перейти к следующему условию».

Для второго блока: «Если перед иллокутивным глаголом есть слово “компания”, “прокуратура”, “пресс-служба”, то вывести это слово в слот “Автор прямой речи”, иначе перейти к следующему условию».

Итак, мы рассмотрели те категории предложений, в которых найти необходимую информацию об авторе прямой речи и вывести ее в соответствующий слот можно с помощью определенных формально-семантических признаков, присутствующих в этих предложениях. Далее рассмотрим случаи, в которых будет сложнее найти слова, обозначающие автора прямой речи, и вывести их в соответствующий слот.

В случае, когда автор прямой речи обозначен профессией, должностью, статусом или профессией и именем с фамилией, мы не можем предсказать, сколько слов будет предшествовать или следовать за иллокутивным глаголом. В предыдущих категориях при поиске нужной информации мы руководствовались такими формальными признаками, как заглавная буква, наличие определенного предлога или слова, но слова, обозначающие профессии, не обладают такими формальными признаками, которые помогли бы составить условие для идентификации этих слов. Исключение могут составить лишь некоторые профессии, заканчивающиеся на характерные окончания, такие как *-ор*, *-ист*, *-ик*. Но такой алгоритм, ориентированный на поиск слова по его окончанию, был бы неэффективным, так как обрабатывал бы лишь некоторые предложения и те частично, так как мы видим, что в большинстве предложений среди отобранных нами примеров название профессии не ограничивается одним словом, а представляет собой ряд слов:

*«– Сейчас всюду идет строительство центра ядерной высокотехнологичной медицины в Димитровграде, поскольку рядом находится предприятие НИИ атомных реакторов, – рассказывает декан факультета экспериментальной и теоретической физики Национального исследовательского ядерного университета “МИФИ” профессор Владимир Беляев» [6, с. 38].*

По этой же причине нецелесообразным было бы и составление списка профессий. В данной категории предложений правильным действием было бы выводить в слот «Автор прямой речи» всю фразу до иллокутивного глагола (если он предшествует прямой речи или разделяет ее) или всю фразу после него до знака «запятая» или «точка» (если иллокутивный глагол следует за прямой речью). При таком подходе, конечно, не избежать вывода в соответствующий слот лишней информации. Для примера рассмотрим следующее предложение:

*«Сейчас мы не ждём три месяца, а точно знаем, есть туберкулёз или нет, в течение одного-двух дней. При некоторых методиках это всего полтора часа. Кроме того, мы получаем информацию об устойчивости болезни к препаратам. Более детальный анализ можно получить с помощью ускоренной диагностики в пробирках. Результаты придут за 2-3 недели. Это позволяет лечить заболевание адресно», – рассказала о прорыве в этой области заведующая отделением фтизиатрии [11].*

Следуя заданной нами команде, при обработке этого предложения в слот «Автор прямой речи» будет выведена следующая фраза «о прорыве в этой области заведующая отделением фтизиатрии», где слова «о прорыве в этой области» будут неактуальной информацией для данного слота, а слова «заведующая отделением фтизиатрии» актуальной информацией.

О сложности глубокого и точного анализа М. Г. Мальковский пишет: «...по причине специфичности подязыка различных областей человеческой деятельности достаточно глубокий анализ достигается пока только для определенных узких предметных областей. Во многих коммерческих системах, существующих на сегодняшний день, ввиду того, что предназначаются они для большого количества пользователей и разных предметных областей, принята концепция поверхностного анализа. Положительной стороной подобного анализа является быстрота» [10, с. 7].

Итак, в результате написания статьи мы разработали алгоритм, с помощью которого можно обрабатывать заданный публицистический текст, находить в нем речевые фреймы, а также находить и выводить в соответствующие слоты такие параметры речевого фрейма, как иллокутивный глагол и автор прямой речи.

Дальнейшее усовершенствование разработанного нами алгоритма подразумевает постоянное обеспечение его описаниями и моделями для более полного анализа.

#### Список источников

1. Аршавин: Если мы не оправдали ваши ожидания – это не наши проблемы! [Электронный ресурс]. URL: <http://www.eg.ru/sport/32335/> (дата обращения: 18.07.2012).
2. Варламова И. подростку вкатили «Пятерочку» [Электронный ресурс]. URL: <https://rg.ru/2017/11/20/v-rostove-nadonu-ohranniki-magazina-posadili-rebenka-pod-strazhu.html> (дата обращения: 21.11.2017).
3. Замахина Т. Ключевое слово – рост [Электронный ресурс]. URL: <https://rg.ru/2017/11/24/gosduma-zapisala-v-biudzhet-rost-zarplat-i-ekonomiki.html> (дата обращения: 27.11.2017).
4. Звонова О. К юбилею Михаила Ходорковского: 5 высказываний о себе, тюрьме и политике [Электронный ресурс]. URL: [http://www.aif.ru/dontknows/k\\_yubileyu\\_mihaila\\_hodorkovskogo\\_5\\_vyskazyvaniy\\_o\\_sebe\\_tyurme\\_i\\_politike](http://www.aif.ru/dontknows/k_yubileyu_mihaila_hodorkovskogo_5_vyskazyvaniy_o_sebe_tyurme_i_politike) (дата обращения: 06.07.2017).
5. Костиков В. Когда сошла вода // Аргументы и Факты. 2012. 18 июля.
6. Кудрявцева Е. Пристальное излучение // Огонёк. 2013. № 28.
7. Куликов А. Спасатель в Саратове остановил «взбесившуюся» фуру [Электронный ресурс]. URL: <https://rg.ru/2013/06/25/reg-pfo/lukianov-anons.html> (дата обращения: 06.07.2017).
8. Латухина К. Азбука победы [Электронный ресурс]. URL: <https://rg.profkiosk.ru/article.aspx?aid=605316> (дата обращения: 21.11.2017).
9. Латухина К. Путин пошутил про пиво после переговоров с Земаном [Электронный ресурс]. URL: <https://rg.ru/2017/11/21/putin-poshutit-pro-pivo-posle-peregovorov-s-zemanom.html> (дата обращения: 21.11.2017).
10. Мальковский М. Г., Грацианова Т. Ю., Полякова И. Н. Прикладное программное обеспечение: системы автоматической обработки текстов. М.: МГУ, 2000. 52 с.

11. Семенова Е. Как обезопасить себя от туберкулёза: советы фтизиатра [Электронный ресурс]. URL: <http://www.aif.ru/health/life/1474875> (дата обращения: 24.07.2013).
12. Смирнов С. Facebook извинилась за утечку личных данных [Электронный ресурс]. URL: [https://www.vedomosti.ru/technology/articles/2013/06/22/facebook\\_priznala\\_utechku\\_lichnyh\\_dannyh\\_polzovatelej](https://www.vedomosti.ru/technology/articles/2013/06/22/facebook_priznala_utechku_lichnyh_dannyh_polzovatelej) (дата обращения: 24.06.2013).
13. Сюткина Е. Инвалидам из Башкирии предложили убежище в США [Электронный ресурс]. URL: <https://mirmov.ru/obshchestvo/invalidam-iz-bashkirii-predlozhili-ubezhishe-v-ssha.html> (дата обращения: 25.01.2013).
14. Тургенев И. С. Отцы и дети. М.: Астрель, 2004. 240 с.
15. Черных А. Общественная палата подготовила сочинение на тему ЕГЭ // Коммерсантъ. 2013. 23 июля.

## ALGORITHM OF SPEECH FRAMES IDENTIFICATION

**Nazaretyan Nelli Vartanovna**  
North-Caucasus Federal University, Stavropol  
[ursa777@yandex.ru](mailto:ursa777@yandex.ru)

The article presupposes the study of the formal semantic features represented in the sentences with direct speech with the aim of making the algorithm on their basis for automatic processing of such sentences. The main content of the article is a detailed description of each step of the compiled algorithm in a natural language. The result of the work is the construction of an algorithm by means of which an automated system can process the specified publicistic text, find speech frames in it and show the values of certain parameters in the corresponding slots.

*Key words and phrases:* automatic text processing; algorithm; illocutionary verb; direct speech; publicistic discourse; semantic features; speech frame; slot; formal features.

УДК 811.161.1:811.163.41

*Предлагаемая статья посвящена анализу системы терминов родства и свойства в современном сербском языке, объем которых составил 300 единиц, полученных методом сплошной выборки из словарей современного сербского языка и литературных примеров из произведений сербских авторов, а также выявлению универсальных и культурно-специфичных признаков терминов родства в рамках лингвокультурного и лингвокогнитивного описания. Объектом исследования являются лексические и паремиологические единицы современного сербского литературного языка, социолингвистический обзор которого также представлен в статье.*

*Ключевые слова и фразы:* номинация; термины родства; сербский язык; культурно-специфичные лексические единицы; паремии.

**Новоспасская Наталья Викторовна**, к. филол. н.  
**Лазарева Олеся Викторовна**, к. филол. н.  
Российский университет дружбы народов, г. Москва  
[novospasskaya72@mail.ru](mailto:novospasskaya72@mail.ru); [0lesikk@mail.ru](mailto:0lesikk@mail.ru)

**Николич Милина**, к. филол. н.  
Нефтяная индустрия Сербии  
[milina89@gmail.com](mailto:milina89@gmail.com)

## ОСОБЕННОСТИ НОМИНАЦИИ РОДСТВА-СВОЙСТВА В СОВРЕМЕННОМ СЕРБСКОМ ЛИТЕРАТУРНОМ ЯЗЫКЕ

### Введение

Вопросу о функционировании языка как средства общения и как совокупности знаний о социальных, культурных, исторических богатствах различных народов в современной лингвистической литературе уделяется всё больше внимания. Система номинаций родственных и свойственных связей не раз являлась объектом описания языковедов-славистов, однако данное исследование актуально на современном этапе развития науки за счёт отсутствия единого лингвокультурологического анализа рассматриваемых единиц в сербском языке и включения в объект исследования паремий с компонентами-терминами родства и свойства сербского языка, что даёт возможность раскрыть языковую специфику тематических групп *кровное родство*, *некровное родство*, *духовное родство*, *свойство*, *свадьба*, *семья*.

Предметом проведенного исследования является выражение идеи родства и свойства системой терминов и паремиологическими оборотами в сербском языке. Цель исследования состоит в определении универсальных и культурно-специфичных черт в функционировании терминологии родства и свойства сербского языка, т.к. исследуемый массив лексики и паремиологии является преимущественно относимым к общеславянской лексике. Методологическую базу исследования составили научные труды лингвистов