
Головко Н.В. Опыт решения практической задачи...

6. Petrova N.E., Ratsiburskaya L.V. Yazyk sovremennykh SMI: sredstva rechevoj agressii: uchebnoe posobie. M.: Flinta: Nauka, 2011.
7. Shcherbinina Yu.V. Rechevaya agressiya. Territoriya vrazhdy. M.: Forum, 2015.
8. Fairclough N. Media Discourse. L.; N.Y.; Sydney; Auckland: E. Arnold, 1995.
9. Hayakawa S.I., Hayakawa A.R. Language in Thought and Action. 5th ed. San Diego; N.Y.; L.: Harvest Original Harcourt, Inc., 1992.
10. Lutz W. The New Doublespeak. N.Y.: Harper Collins Publishers, 1996.

DOI: 10.25586/RNU.V925X.20.03.P.053

УДК 81'33

Н.В. ГОЛОВКО

ОПЫТ РЕШЕНИЯ ПРАКТИЧЕСКОЙ ЗАДАЧИ
ПО АВТОМАТИЗИРОВАННОЙ РАССТАНОВКЕ БУКВЫ Ё
В РУССКОЯЗЫЧНЫХ ТЕКСТАХ

Описывается сценарий частичной автоматизации замены буквы Е на букву Ё в конкретных условиях редакционно-издательской деятельности. В целях решения практической задачи проведено статистическое исследование возможных вариантов графического окружения буквы Ё на материале тематического словаря и проанализированы полученные результаты для создания простого механизма, учитывающего специфику и возможности применяемых программных инструментов, который не предполагал бы длительной и трудоёмкой разработки плагинов, расширений или баз данных. Требуемая функциональность была успешно реализована с помощью текстового поиска с применением регулярных выражений.

Ключевые слова: прикладная лингвистика, автоматизированная обработка текстов, графика, редакционно-издательская деятельность, статистические методы обработки языкового материала, русский язык.

N.V. Golovko

AN EXPERIENCE OF SOLVING A PRACTICAL CASE
OF COMPUTER-ASSISTED PLACEMENT OF LETTER Ё
IN RUSSIAN TEXTS

In Russian editorial practice, it is sometimes necessary to replace letter E by letter Ё in texts prepared to be published. The article represents a custom scenario of partial computer-based assistance in this replacement, developed under concrete circumstances. In order to solve this practical problem the author performed a statistical study of possible graphical contexts for letter Ё, based upon a specific dictionary, and analyzed the results of this study for the purpose of further creation of a simple mechanism that would have accounted for specifics and available features of certain software used in their editorial practice – a mechanism that would not have required long-term complicated development of plugins, extensions or databases. The desired functionality was successfully achieved by means of text search with regular expressions.

Keywords: applied linguistics, natural language processing, spelling, editorial practice, statistical methods of language processing, the Russian language.

Одним из практических аспектов литературного редактирования и корректуры в издательской деятельности является решение вопроса о расстановке буквы Ё, или, как иногда говорят отдельные авторы, «ёфикации» текста. Следует заметить, что сама по себе идея «ёфикации» является предметом споров. Хорошо известно, что «Правила русской орфографии и пунктуации» [6] предписывают употреблять Ё только в тех случаях, когда это необходимо для избегания разночтений. Эти указания имеют два следствия. С одной стороны, «факультативность» буквы Ё кажется неочевидной и склоняет авторов и редакторов к дискуссиям о корректности этого правила. С другой стороны, расстановка буквы Ё в текстах превращается в прикладную задачу, которая требует решения на том или ином уровне в зависимости от того, какая политика принята в конкретной редакции или издательстве.

Прежде чем говорить о способах решения упомянутой задачи, хотелось бы кратко описать известные нам точки зрения на проблему «ёфикации» как таковую. Полагаем, что это поможет составить более ясное представление о контексте исследования.

Представляется возможным утверждать, что идея об отказе от использования буквы Ё или о её употреблении только в целях разрешения омографии в конечном счёте сводится к принципу языковой экономии – передавать информацию с приложением как можно меньшего количества усилий. Сложившаяся практика, отражённая в «Правилах...» 1956 г., опирается, по-видимому, на стремление отказаться от избыточности в условиях, когда читатель на основании опыта и контекста способен различить Е и Ё даже при их одинаковой записи – по аналогии, скажем, с дифференциацией омографов, различающихся

ударением. Во главу угла, таким образом, ставится узус: «важен сам факт стихийно сложившегося и естественно сохраняющегося неразличения двух гласных букв», который «говорит о востребованности и легитимности такого упрощения системы букв. Оно не препятствует взаимопониманию: иначе выработались бы нормы различения данных букв» [1]. Также в процитированной работе обращают внимание на то, что таким образом проще сохранять единообразие морфем. Наконец, последовательное употребление Ё создавало бы неразрешимые противоречия при печати старых текстов, где невозможно определить, какое именно произношение имел в виду автор (см. об этом: [5; 8]).

В противовес существующему правилу высказываются, например, следующие суждения:

1. Буква Ё не является вариантом буквы Е, несмотря на аналогичное написание, поэтому смешивать их на письме странно. С графической точки зрения Е соответствует звуку [э], Ё – звуку [о], включая йотированные варианты. В то же время не являются факультативными, например, Й или Щ, хотя, в частности, в последнем случае нетрудно было бы предложить аргумент о том, что Щ – это не более чем долгое мягкое Ш, а твёрдые и мягкие согласные в русской графике, как известно, отдельных букв не имеют (см., например: [2; 4]).

2. Содержащееся в «Правилах...» 1956 г. предписание употреблять Ё постольку, поскольку это устраняет смысловые неоднозначности, создаёт эффект произвольности: само представление о неоднозначности зависит от пишущего, чьё восприятие может отличаться от идей читающего. Грубо говоря, автору текста может показаться, что всё и так ясно, в то время как читатель столкнётся с ситуацией, когда даже широкий контекст не помогает извлечь из текста

правильный смысл (см. пример со стихотворением А. Барто: [2]).

3. Неупотребление буквы Ё является ошибкой, поскольку в орфографическом словаре все соответствующие словоформы записаны именно с Ё.

Лично мы считаем аргумент № 3 достаточно весомым для того, чтобы им руководствоваться: если слово в нормативном словаре записано определённым образом, то именно так оно и должно писаться. В противном случае само понятие нормы выглядит пошатнувшимся: норма существует, но в то же время ей словно бы и необязательно следовать. Поэтому мы, начав личную редакторскую практику в сетевом средстве массовой информации, решили пользоваться оговоркой, содержащейся в академическом справочнике «Правила русской орфографии и пунктуации» 2006 г. [7]: автор или редактор конкретного издания вправе использовать букву Ё во всех возможных случаях по своему усмотрению.

Заметим, впрочем, что в целом принцип различения омографов может быть хорошей и удобной альтернативой. В отличие от задачи по расстановке буквы Ё во всех возможных случаях, он предполагает составление заведомо конечного списка слов, требующих проверки. Однако в открытом доступе отсутствуют полные списки омографов с Ё (хотя нам известны некоторые попытки в смежных областях; см., например: [3; 9]), самостоятельное составление такого списка представляется весьма трудоёмким и не гарантирующим его полноты предприятием, к тому же задача полной «ёфикации» более универсальна: в некоторых изданиях, таких как книги для детей младшего возраста или учебные издания для изучающих русский язык как иностранный, обойтись разрешением омографических неоднозначностей нельзя. Наконец, не следует забывать и о других аспектах

предписаний «Правил русской орфографии и пунктуации», например об обязательном употреблении Ё в малоизвестных наименованиях вроде «река Олёкма».

Руководствуясь изложенными рассуждениями, мы пришли к необходимости решения практической задачи по выделению в русскоязычных текстах тех мест, где может требоваться замена буквы Е на букву Ё. Эта задача рассматривается нами как альтернатива полному прочтению текста с самостоятельным ручным поиском слов, нуждающихся в «ёфикации». Несомненно, что при должной сноровке и орфографической зоркости обнаружение слов, которые пишутся с буквой Ё, не является непреодолимым препятствием, однако нам хотелось бы обеспечить этому процессу частичную техническую (вычислительную) поддержку, которая позволила бы избежать случайного пропуска таких слов и ускорить соответствующую часть вычитки за счёт концентрации на тех местах, где буква Ё действительно может присутствовать. Также мы хотели обойтись минимальным инструментарием, доступным в нашей рабочей среде, – программным комплексом G Suite и его компонентом Google Docs, т.е. сформировать некий минимальный критерий, который можно было бы реализовать через встроенную функциональность текстового поиска, не разрабатывая программных расширений и не прибегая к обширным базам данных. Очевидно, в частности, что можно просто взять орфографический словарь и извлечь из него все слова с Ё; однако ввести в поисковую строку тысячи словоформ невозможно, и этот список опять же не был бы гарантированно исчерпывающим (т.е. универсальным).

Мы сформулировали и решили проверить простую гипотезу: возможно, существует конечный набор буквосочетаний,

в которых возможна буква Ё. Кажется очевидным, что в конкретном реальном языке могут использоваться отнюдь не все возможные сочетания звуков (а следовательно, и букв); таким образом, не исключено, что спектр допустимых окружений для букв Е и Ё различается.

В качестве источника материала мы взяли «Словарь употребления буквы Ё» (2009 г., около 20 000 слов), составленный известным сторонником последовательной «ёфикации» В.Т. Чумаковым [10].

Убеждения составителя словаря позволяют рассчитывать на то, что в последнем представлен как можно более широкий спектр случаев использования буквы Ё, включающий жаргонизмы, диалектную лексику и авторские неологизмы. Из списка слов полуавтоматическим способом (с использованием механизмов поиска в текстовом процессоре Microsoft Word 2016) были извлечены и проанализированы сочетания вида «символ + Ё» и «Ё + символ» (табл. 1–2).

Таблица 1

Сочетания вида «символ + Ё» и статистика по ним, ед.

№ п/п	Сочетание	Количество	№ п/п	Сочетание	Количество	№ п/п	Сочетание	Количество
1	Начало слова	1852	12	КЁ	63	23	ХЁ	2
2	АЁ	187	13	ЛЁ	2455	24	ЦЁ	2
3	БЁ	202	14	МЁ	511	25	ЧЁ	1032
4	ВЁ	546	15	НЁ	1366	26	ШЁ	356
5	ГЁ	32	16	ОЁ	206	27	ЩЁ	435
6	ДЁ	780	17	ПЁ	310	28	ЪЁ	214
7	ЕЁ	82	18	РЁ	1663	29	ЫЁ	3
8	ЖЁ	600	19	СЁ	425	30	ЬЁ	535
9	ЗЁ	291	20	ТЁ	1071	31	ЭЁ	1
10	ИЁ	123	21	УЁ	87	32	ЮЁ	25
11	ЙЁ	4	22	ФЁ	31	33	ЯЁ	0

Таблица 2

Сочетания вида «Ё + символ» и статистика по ним, ед.

№ п/п	Сочетание	Количество	№ п/п	Сочетание	Количество	№ п/п	Сочетание	Количество
1	Конец слова	419	12	ЁК	579	23	ЁХ	537
2	ЁА	0	13	ЁЛ	531	24	ЁЦ	2
3	ЁБ	262	14	ЁМ	853	25	ЁЧ	82
4	ЁВ	847	15	ЁН	5153	26	ЁШ	290
5	ЁГ	226	16	ЁО	0	27	ЁЩ	8
6	ЁД	133	17	ЁП	264	28	ЁЪ	0
7	ЁЕ	0	18	ЁР	1718	29	ЁЫ	0
8	ЁЖ	371	19	ЁС	521	30	ЁЬ	0
9	ЁЗ	245	20	ЁТ	2363	31	ЁЭ	0
10	ЁИ	0	21	ЁУ	1	32	ЁЮ	3
11	ЁЙ	73	22	ЁФ	18	33	ЁЯ	0

Как видно из полученных данных, поиск сочетаний вида «символ + Ё» менее перспективен с точки зрения выделения буквосочетаний: предшествовать букве Ё может почти любая другая буква алфавита. В свою очередь, комбинации вида «Ё + + символ» оказались более пригодными для наших целей, поскольку даже при самом тщательном подходе к поиску максимально экзотических слов (включая придуманные самим составителем словаря) не удалось найти ни одной лексемы для 9 различных сочетаний. Целесообразно при этом заметить, что ни разу не встретившиеся сочетания буквы Ё с последующими гласными вполне возможны в случае буквы Е: *театр, длиннее, идеи* и т.п.

С учётом статистики, а также возможностей поискового механизма текстового процессора Google Docs, которые изначально были частью задачи, мы приняли решение использовать в качестве индикатора набор вариантов сочетаний вида «Е + символ» (в данном случае именно Е, поскольку обрабатываются тексты, в которых Е необходимо заменить на Ё). При этом узкая тематическая специфика конкретного СМИ, тексты которого подвергались редактированию, позволила нам дополнительно отсеять некоторые другие сочетания, такие как ЁЮ (все три варианта, встретившиеся в словаре, – это устаревшие окончания существительных типа *землёю*, которые не могут появиться в современном тексте) или ЁУ (единственное вхождение – название финского музыкального инструмента *ёухикко*, вероятность упоминания которого в текстах по информационной безопасности близка к нулевой). Для некоторых буквосочетаний мы добавили проверку комбинаций вида «символ + Е + символ».

На вопросе о проверке трёхсимвольных сочетаний, как представляется, следует остановиться отдельно. Наш анализ

показал, что некоторые сочетания вида «Ё + символ» имеют крайне ограниченный набор возможных предшествующих символов: например, комбинация ЁП существует только в вариантах ЛЁП, РЁП, ТЁП и ШЁП. Следовательно, вспомогательное ограничение предшествующих символов могло бы сузить поиск и уменьшить количество отображаемых единиц. Мы, однако, использовали трёхсимвольные сочетания лишь частично в связи с фундаментальным ограничением поисковой функциональности текстового процессора Google Docs: при поиске по регулярному выражению каждая буква может входить в состав только одной отображаемой комбинации. Например, если обрабатывается написание «дешевый», то при поиске по трёхсимвольным сочетаниям будет выделено только «деш», в то время как вторая буква Е (которая как раз и подлежит замене на Ё) останется без маркировки, поскольку буква Ш уже рассмотрена как входящая в «деш» и, следовательно, не может входить в состав комбинации «шев». В силу этого применение трёхбуквенных сочетаний возможно лишь там, где среди возможных предшествующих букв нет тех, которые отобраны в набор основных комбинаций вида «Ё + символ». Кроме того, мы не обнаружили способа добавить во множество предшествующих символов обозначение начала слова, а следовательно, лексемы наподобие «ёмкость» также не будут выделены. Учитывая конкретные условия задачи, мы решили ограничиться главным образом двухсимвольными конструкциями, добавив трёхсимвольные там, где это не вызывает пересечений.

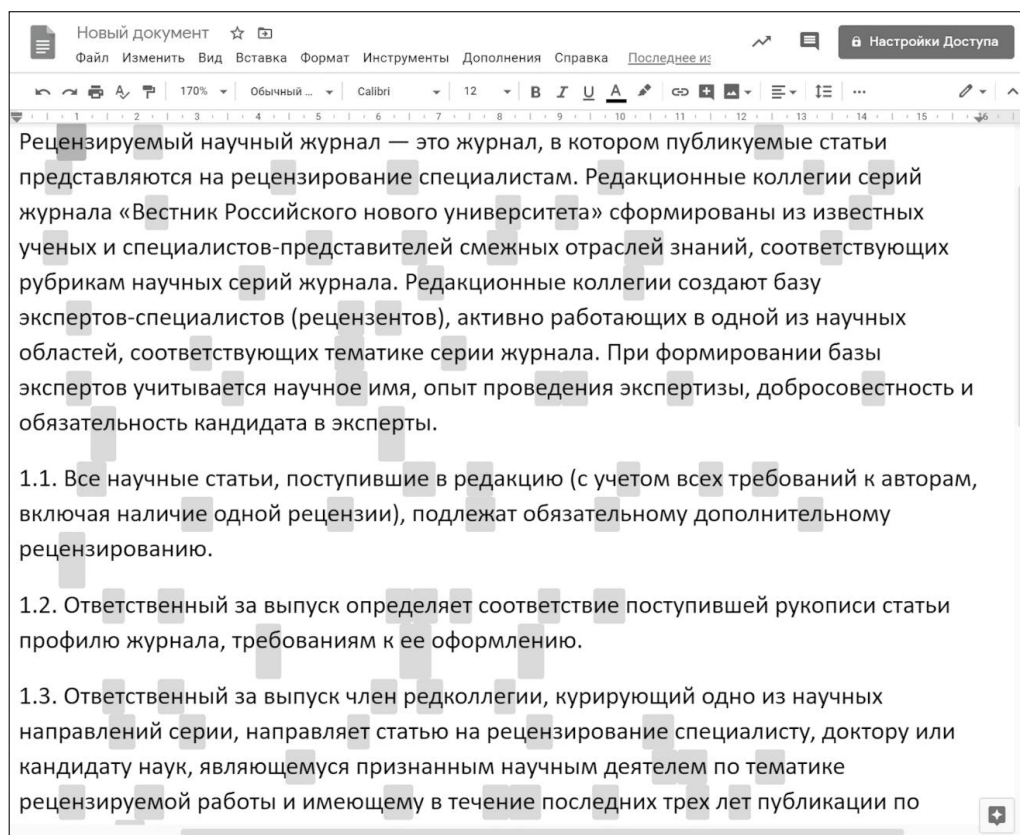
В то же время мы провели дополнительную проверку, сравнив результаты поиска в одном из отредактированных текстов при использовании только трёхсимвольных комбинаций, только двухсимвольных соче-

таний и смешанного подхода, описанного в предыдущем предложении. Объём текста составлял около 25 000 знаков с пробелами и около 3000 слов. При максимально широком подходе (только «Ё + символ») поиск подсветил 1529 буквосочетаний, при максимально узком (только трёхбуквенные комбинации) – 1167, при смешанном – 1373. Таким образом, применение трёхсимвольных сочетаний уменьшило количество вхождений для проверки на 362 единицы (23,7%), смешанный подход – на 156 единиц (10,2%). Соответственно, с учётом проблематичности реализации максимально узкого варианта в конкретных условиях решаемой практической задачи – каждое сочетание пришлось бы

проверять отдельно, чтобы они не создавали помех друг другу, – смешанный подход представляется удовлетворительным компромиссом.

По итогам проведённой статистической и аналитической работы мы сформировали регулярное выражение для поисковой функциональности текстового процессора Google Docs, которое позволяет нам подсвечивать в редактируемом тексте возможные места замены буквы Е на букву Ё с учётом тематической специфики того конкретного сетевого издания, в интересах которого проводится это редактирование (рис.):

`е|ен|ер|ем|ев|ед|еш|ек|ес|еж|[еиосшь е[-,?!>]);]|ез|ег|ед|[елнруь]ей|ех|еб|еп`



Пример подсветки буквосочетаний с помощью поискового запроса

Как видно из рисунка, регулярное выражение позволит обратить внимание при вычитке на слова *ученых, учетом, ее, трех*, в которых букву Е необходимо заменить на Ё.

Таким образом, мы провели статистическое исследование возможных графических окружений, в которых встречается буква Ё, и в результате последующей аналитической работы решили практическую задачу по частичной технической поддержке процесса расстановки буквы Ё при редак-

тировании текста с учётом конкретных условий. Полученные статистические данные и описанный опыт решения задачи могут быть полезны при разработке более сложных алгоритмов и механизмов «ёфикации», в том числе с использованием более точных трёхсимвольных буквосочетаний при отсутствии ограничений, налагаемых особенностями работы конкретного текстового процессора, средствами которого реализовывалась интересовавшая нас функциональность.

Литература

1. Голев Н.Д. Коммуникативная орфография русского языка (на примере неразличения на письме букв е и ё). URL: <http://lingvo.asu.ru/golev/articles/v99.html> (дата обращения: 17.05.2020).
2. Гутентог М. Буква Ё – графема-омоним. Причины и следствия // Фигуры речи. Филология. URL: http://figur.ir2.ru/bukva_io.html (дата обращения: 17.05.2020).
3. Лобанов Б.М. Проблема разрешения «Ё»-омографов при синтезе речи по тексту // Международная конференция «Диалог». URL: <http://www.dialog-21.ru/digests/dialog2009/materials/html/45.htm> (дата обращения: 17.05.2020).
4. Нисман Л. Русский язык. Трудности, тайны, тонкости и не только... М.: Литео, 2018. 300 с.
5. Пахомов В.М. Буква Ё // Справочно-информационный интернет-портал «Русский язык». URL: http://gramota.ru/class/istiny/istiny_7_jo/ (дата обращения: 17.05.2020).
6. Правила русской орфографии и пунктуации. М.: Учпедгиз, 1956. 176 с.
7. Правила русской орфографии и пунктуации. Полный академический справочник / под ред. В.В. Лопатина. М.: Эксмо, 2006. 480 с.
8. Суперанская А.В. Вновь о букве Ё // Наука и жизнь. 2008. № 1. С. 32–34.
9. Чемерилов В.В., Фадеев А.С. Система автоматического разрешения омографии на основе семантической связи слов смежных предложений в текстовом отрывке // Доклады ТУСУР. 2018. Т. 21, № 2. С. 42–48.
10. Чумаков В.Т. Словарь употребления буквы Ё. URL: <http://www.yomaker.ru/Slov2009.doc> (дата обращения: 17.05.2020).

Literatura

1. Golev N.D. Kommunikativnaya orfografiya russkogo yazyka (na primere nerazlicheniya na pis'me bukv e i yo). URL: <http://lingvo.asu.ru/golev/articles/v99.html> (data obrashcheniya: 17.05.2020).
2. Gutentog M. Bukva Yo – grafema-omonim. Prichiny i sledstviya // Figury rechi. Filologiya. URL: http://figur.ir2.ru/bukva_io.html (data obrashcheniya: 17.05.2020).
3. Lobanov B.M. Problema razresheniya "Yo"-omografov pri sinteze rechi po tekstu // Mezhdunarodnaya konferentsiya "Dialog". URL: <http://www.dialog-21.ru/digests/dialog2009/materials/html/45.htm> (data obrashcheniya: 17.05.2020).

4. Nisman L. Russkij yazyk. Trudnosti, tajny, tonkosti i ne tol'ko... M.: Liteo, 2018. 300 s.
5. Pakhomov V.M. Bukva Yo // Spravochno-informatsionnyj internet-portal "Russkij yazyk". URL: http://gramota.ru/class/istiny/istiny_7_jo/ (data obrashcheniya: 17.05.2020).
6. Pravila russkoj orfografii i punktuatsii. M.: Uchpedgiz, 1956. 176 s.
7. Pravila russkoj orfografii i punktuatsii. Polnyj akademicheskij spravochnik / pod red. V.V. Lopatina. M.: Eksmo, 2006. 480 s.
8. Superanskaya A.V. Vnov' o bukve Yo // Nauka i zhizn'. 2008. № 1. S. 32–34.
9. Chemerilov V.V., Fadeev A.S. Sistema avtomaticheskogo razresheniya omografii na osnove semanticheskoy svyazi slov smezhnykh predlozhenij v tekstovom otryvke // Doklady TUSUR. 2018. T. 21, № 2. S. 42–48.
10. Chumakov V.T. Slovar' upotrebleniya bukvy Yo. URL: <http://www.yomaker.ru/Slov2009.doc> (data obrashcheniya: 17.05.2020).

DOI: 10.25586/RNU.V925X.20.03.P.060

УДК 811.163.41

И. Голушин

СЛОВООБРАЗОВАТЕЛЬНЫЕ СРЕДСТВА ВЫРАЖЕНИЯ ЭМОТИВНОСТИ В СЕРБСКОМ ЯЗЫКЕ

Рассматриваются словообразовательные средства выражения эмотивности в сербском языке. Особое внимание уделяется аффиксации (префиксации и суффиксации), а также универбации и композиции как наиболее распространенным способам словообразования в сербском языке. Проводится комплексный анализ словообразовательных ресурсов сербского языка с точки зрения их эмотивного потенциала. В результате делается вывод, что словообразовательные средства играют важную роль в процессе выражения эмотивности и наиболее продуктивным способом передачи эмотивного значения в сербском языке является суффиксация.

Ключевые слова: эмотивность, средства выражения эмотивности, словообразование, аффиксация, универбация, композиция.

I. Golushin

WORD-FORMING MEANS OF EXPRESSING EMOTIVITY IN THE SERBIAN LANGUAGE

The word-formation means of expressing emotivity in the Serbian language are considered. Special attention is paid to affixation (prefixation and suffixation), as well as univerbation and composition as the most common ways of word formation in the Serbian language. A comprehensive analysis of the word-formation resources of the Serbian language from the point of view of their emotive potential is carried out. As a result it is proved that word-forming means play an important role in the process of expressing emotivity and the most productive way of transmitting emotional meaning in the Serbian language is suffixation.

Keywords: emotivity, means of expressing emotivity, word formation, affixation, univerbation, composition.