

А.А. Кретов

Воронежский государственный университет, г. Воронеж

МАРКЕМЫ И КЛЮЧЕВЫЕ СЛОВА В НАУЧНЫХ ТЕКСТАХ MARKEMS AND KEY-WORDS IN SCIENTIFIC DISCOURSE

Ключевые слова: автоматический анализ текста, ключевые слова, маркемный анализ лингвистических текстов

Keywords: automatic processing of the text, key-words, markemic analysis of the linguistic texts

Нормой современного оформления статей в научных журналах является указание ключевых слов.

Цель данного исследования – проверить, в каких отношениях находятся интуитивно выделяемые авторами ключевые слова к маркемам (словоформам с максимальным Индексом текстуальной маркированности – ИнТеМом; определение маркем и способы обнаружения ИнТеМа описаны в (Кретов, 2010)). Такое сравнение позволит, с одной стороны, объективно оценить интуицию авторов, а с другой – исследовать возможность автоматического выделения в научных статьях ключевых слов с помощью маркемного анализа.

Объектом исследования является 20 научных статей по лингвистике, опубликованные в научных журналах, а предметом – маркемы, выделяемые в этих статьях.

Инструментом обработки текстов для получения маркем являются программы: «Программа выделения тематически маркированной лексики – *ProTeMaL*», авторы Попова И.В., Воронина И.Е., Кретов А.А. (зарегистрирована «Центром информационных технологий и систем органов исполнительной власти», № 50201000004 от 11.01.2010) и программа автоматизированной лемматизации словоформ «*Add2*», автор М.В. Катов (ВГУ).

Алгоритм выявления маркем посредством указанных программ таков: 1) текстовый файл статьи обрабатывается программой *ProTeMaL*, результата обработки сохраняется в файле КС; 2) полученный файл обрабатывается программой *Add2*, результат обработки сохраняется в файле КСМ в текстовом формате. Дальнейшая обработка файла КСМ осуществляется в электронных таблицах Microsoft Excel; 3) материал файла сортируется по следующим параметрами: по частям речи и по величине ИнТеМа (в порядке убывания); 4) в последующем анализе участвуют только существительные, прилагательные и глаголы с положительными значениями ИнТеМа, причем опыт показывает, что глагольные маркемы малочисленны и малоинформативны: в лучшем случае они подтверждают свидетельства субстантивных и адъективных маркем; 5) существительные, прилагательные и глаголы с положительными значениями ИнТеМа анализируются на разных листах книги Excel; 6) подсчитывается число словоформ каждого слова и

суммируются их ИнТеМы; 7) леммы ранжируются в порядке убывания значений суммы ИнТеМов словоформ одного слова (СумИнТеМа); 8) сравниваются леммы существительных, прилагательных и глаголов с СумИнТеМом > 0,175. Предпочтение отдаётся однокоренным леммам, а также леммам, имеющим синонимичные или близкие по значению леммы; 9) леммы с чрезмерно широкой (общенаучной или общелингвистической) семантикой, не отражающей специфики данной статьи (например, *рассмотрение, исследование, анализ, описание, исследователь, автор* и т.п.), отсеиваются.

Затем полученные маркемы сравниваются с ключевыми словами, представленными в опубликованных статьях, названиями статей и (при необходимости – с аннотацией статьи).

Мерой согласия (*Мсогл*) интуиции и маркемного анализа можно считать относительную величину пересечения двух множеств: ключевых слов и маркем. Эта величина вычисляется по формуле:

$$M_{согл} = \frac{0,5 (N_{кc} + N_{мк}) - (N_{кc}^e + N_{мк}^e)}{0,5 (N_{кc} + N_{мк})},$$

где $N_{кc}$ – количество ключевых слов, $N_{мк}$ – количество маркем, $N_{кc}^e$ – количество единичных (не имеющих соответствия в маркемах) ключевых слов, $N_{мк}^e$ – количество единичных (не имеющих соответствия в ключевых словах) маркем.

Рассмотрим в качестве примера статью Л.Г. Кузьминой КОНЦЕПЦИЯ КАЧЕСТВА ПРОФЕССИОНАЛЬНО-МЕТОДИЧЕСКОЙ ПОДГОТОВКИ ПРЕПОДАВАТЕЛЯ ИНОСТРАННЫХ ЯЗЫКОВ (Воронежский государственный университет), поступила 16 апреля 2009 г.

Аннотация. В статье рассматриваются современные подходы к определению целей, содержания и форм организации профессионально-методической подготовки будущих преподавателей иностранных языков, составляющие концептуальную основу методической системы, сложившейся на факультете РГФ ВГУ. Утверждается, что комплексная реализация данных положений свидетельствует о качественных изменениях в подготовке преподавателя, отвечающих современным запросам общества.

Ключевые слова. Профессиональное образование, подготовка преподавателя иностранных языков, методическая компетенция будущего преподавателя, качество высшего образования.

Маркемный анализ дал следующие результаты:

МаркемыС	СумИнТеМ	СлФ	МаркемыП	СумИнТеМ	СлФ
преподаватель	0,77967	3	профессиональный	2,02165	6
специальность	0,39862	2	методический	0,62505	4
студент	0,32071	3	педагогический	0,57313	3
качество	0,29923	3	образовательный	0,38880	2
образование	0,28566	1	иностраннный	0,35148	3
преподавание	0,28561	1			

$N_{\text{КС}} = 13, N_{\text{МК}} = 12, N_{\text{КС}}^c = 3$ (компетенция, языков, высшего), $N_{\text{МК}}^c = 0$.

$M_{\text{согл}} = [0,5 (13+12) - (3+0)] : [0,5 (13+12)] = (12,5 - 3) : 12,5 = 9,5 : 12,5 = 0,8$.

Лучшие значения $M_{\text{согл}}$ будут при равенстве $N_{\text{КС}}$ и $N_{\text{МК}}$. ($N_{\text{КС}} - N_{\text{МК}} = 0$). Увеличение разности $N_{\text{КС}} - N_{\text{МК}}$ по модулю ведёт к уменьшению $M_{\text{согл}}$. Следовательно, чем больше ключевых слов выделено автором статьи, тем (при прочих равных) больше значения $M_{\text{согл}}$.

Можно было бы высчитать $M_{\text{согл}}$ для каждого из 20 обследованных текстов и средний $M_{\text{согл}}$ по всем 20 текстам. Единственная причина, по которой мы этого не делаем заключается в том, что, взятая в отвлечении от содержания, величина $M_{\text{согл}}$ мало информативна: она может интерпретироваться и как свидетельство несостоятельности алгоритма, и как показатель неудачного выбора ключевых слов автором статьи.

Поэтому более адекватной представляется содержательная оценка работы алгоритма на фоне ключевых слов и названия статьи.

Например, Карпова О.М. РОЛЬ ПИСАТЕЛЬСКИХ СЛОВАРЕЙ В ФОРМИРОВАНИИ АНГЛИЙСКОЙ НАЦИОНАЛЬНОЙ ЛЕКСИКОГРАФИИ (Ивановский государственный университет) Поступила в редакцию 30.09.2008 г. *Аннотация:* В статье рассматриваются тенденции развития английской национальной лексикографии в ракурсе историко-типологического подхода, выработанного в отечественной теоретической лексикографии. Особое внимание уделяется специальным словарям, оказавшим существенное влияние на формирование толковых словарей английского языка. Впервые выдвигается авторская гипотеза о том, что именно писательские словари (главным образом, конкордансы и глоссарии Шекспира, Чосера, Мильтона) и конкордансы к английскому переводу Библии стали базой иллюстративных примеров для микроструктуры первых толковых словарей английского языка, включая авторитетный словарь С.Джонсона. *Ключевые слова:* лексикография, конкорданс, глоссарий.

КС	МКС	Сл Ф	Сум ИнТеМ	МКП	Сл Ф	Сум ИнТеМ
лексикография	лексикография	3	1,09961	английский	6	1,29752
конкорданс	произведение	2	0,69252	лексикографический	1	0,81414
глоссарий	словоупотребление	1	0,64238	иллюстративный	2	0,69099
	словарь	4	0,62490	национальный	2	0,50893
	формирование	1	0,40312	дифференциальный	1	0,47042

справочник	2	0,39102	литературный	1	0,45697
становление	1	0,28969	академический	1	0,43049
создание	1	0,23743	дополнительный	1	0,23135
лексикограф	1	0,23135	орфоэпический	1	0,23135
составитель	1	0,22797	писательский	1	0,22797
гlossарий	2	0,22086	специальный	1	0,22797
писатель	1	0,20832	толковый	1	0,19891

Сравнение показывает, что ключевые слово **лексикография**, поддержанное маркемами *лексикография*, *лексикограф*, *лексикографический*, выделено автором статьи обосновано. То же, хотя и с меньшей категоричностью можно сказать и о КС **гlossарий**, поддержанном маркемой *гlossарий*. Менее удачным следует признать выделение в качестве КС лексемы **конкорданс**, поскольку его ИнТеМ = 0,10134 недостаточен для придания ему статуса маркемы.

Рассмотрим ещё один случай. Авторы *Кретов А.А.*, *Фененко Н.А.* Реноминация как проблема переводоведения // Вестник ВГУ. Серия лингвистика и межкультурная коммуникация, 2011, № 2, с. 154-158.

КС	МаркемыС	Сл Ф	СумИнТе М	МаркемыП	Сл Ф	СумИнТе М
реноминация	эквивалентность	2	0,70825	французский	3	0,59716
теория реноминации	реноминация	2	0,61096	языковой	3	0,50062
денотат	действительность	1	0,55549	денотативный	1	0,18369
сигнификат	сигнификат	2	0,41236			
преобразование сигнификата	перевод	2	0,37699			
	преобразование	1	0,33095			
	значение	2	0,31834			
	реалия	4	0,31795			
	выражение	2	0,27115			
	информация	2	0,24487			
	процесс	2	0,24360			
	номинация	2	0,20835			
	денотат	2	0,18342			
	отношение	1	0,18112			

Маркемы свидетельствуют, что КС **реноминация** и **сигнификат**, выбраны удачно, но содержание статьи ими отражается не полностью, поскольку в ней также обсуждается проблема *информационной эквивалентности* в передаче *реалий* при *переводе* с французского или на *французский язык*.

Проведённое исследование позволяет заключить, что предложенный выше алгоритм маркемного анализа научных текстов вполне способен давать достоверную информацию как о ключевых словах текста, так и об их тематике.

Литература

1. Кретов А.А. Понятие маркемы: методика выявления и практика использования // Универсалии русской литературы 2. Сборник статей / Под ред. А.А. Фаустова. Воронежский государственный университет. – Воронеж: НАУКА-ЮНИПРЕСС, 2010. - С.138 - 153.

(0,2 п.л.)

ПРИЛОЖЕНИЕ

Сопоставление ключевых слов и маркем в научных текстах

1. Борискина О.О. КРИПТОКЛАССНЫЕ **ПРОЕКЦИИ МИРА** НЕПРЕДМЕТНЫХ СУЩНОСТЕЙ: **опыт** криптоклассного **анализа** словосочетаемости

КС	МкС	СумИнТ еМ	МкП	СумИнТ еМ
коллокация	криптокласс	1,54412	межкриптоклас сный	1,33964
индекс периферийности	избирательность	0,91950	сочетательный	0,70575
межкриптоклассный портрет имени	словосочетание	0,87211	языковой	0,51225
	исследование	0,79156	существительн ый	0,48022
	предрасположен ность	0,75000	распределитель ный	0,47604
	сочетаемость	0,63249	англоязычный	0,46830
	валентность	0,61611	английский	0,42307
	распределенност ь	0,56472	непредметный	0,35347
	периферийность	0,53334	экстенциальн ый	0,23415
	представление	0,52693	лексический	0,21870
	принадлежность	0,48022	настоящий	0,21870
	активность	0,42875		
	распределение	0,31264		
	сознание	0,23896		
	употребление	0,23415		
	возможность	0,21870		
	описание	0,20881		
	сущность	0,20881		
	количество	0,19172		
	сравнение	0,19172		
	портрет	0,18326		

2. Галева Н.Л. Перевод как способ экономии мышления культуры

КС	МкС		МкП	
перевод	интертекстуальност ь	0,81604	переводческий	0,94696
культура	заимствование	0,66547	культурный	0,71512
культурообразующ ая функция перевода	перевод	0,63905	культурообразующий	0,44254
взаимодействие культур	культура	0,63220	непосредственный	0,42490
способы заимствования культурных форм	текстопостроение	0,62906	преимущественный	0,42490

культурная лакуна	мыследеятельность	0,49974	лингвистический	0,36461
развитие культуры	эквивалентность	0,41473	существенный	0,35582
	оригинал	0,35278	собственный	0,31861
	текст	0,32698	культурозначимый	0,23837
	деятельность	0,32697		
	содержание	0,29943		
	формирование	0,24103		
	взаимодействие	0,23837		

3. Кашкин В.Б. Языковые **контрасты** в обыденном сознании и в научной лингвистике

КС	МкС	СумИнТе М	МкП	СумИнТе М
сопоставление языков	представление	1,18808	эпистемологический	1,50461
наивная лингвистика	сопоставление	0,60670	лингвистический	0,67320
обыденное сознание	пользователь	0,56355	обыденный	0,30254
история языкознания	концептуализация	0,53120	иностраный	0,25550
	противоречивость	0,40091	наивный	0,22571
	язык	0,36580	материальный	0,21694
	лингвистика	0,29399		
	исследование	0,27393		
	формирование	0,27393		
	мотивированность	0,27229		
	восприятие	0,21870		
	исследователь	0,20727		

4. Кашкин В.Б., Княжева Е. А. Концептуализация перевода в обыденном сознании и в научном представлении

КС	МкС	СумИнТе М	МкП	СумИнТе М
перевод	переводчик	2,25855	профессиональный	4,21396
обыденное сознание	представление	1,81520	метапереводческий	1,35332
метапереводческий дискурс	представитель	1,56025	переводческий	1,31694
public image профессии переводчика	деятельность	0,66526	общественный	1,26968
	гетеростереотип	0,51248	метакогнитивный	0,43226
	преподаватель	0,49671	обыденный	0,41406
	перевод	0,47716		

5. Рудницкая Е.Л., Новичков П.Ю. Английские номинации болевых и неприятных ощущений с позиций лексико-семантической типологии

КС	МкС		МкП	
лексико-семантическая типология	употребление	0,85426	существительный	1,75781
стативный	значение	0,51665	семантический	0,84504
(не)переходный	валентность	0,49346	неопределенный	0,72997
семантическая/синтаксическая валентность	экспериенцер	0,45739	лексический	0,62913
тема	параметр	0,43959	множественный	0,50143
пациенс	ощущение	0,37831	антропоморфный	0,27540
причина	возникновение	0,36498	продолжительный	0,27540
каузатор	обозначение	0,35833	каузативный	0,23141
экспериенцер	глагол	0,30031	прилагательный	0,20451
лексический параметр	существительное	0,27540	синтаксически	0,20451
лексическое поле	функциональность	0,27540	переносный	0,17733
донорское поле	конструкция	0,26423	переходный	0,17266
метафорическое употребление	дополнение	0,24360	числительный	0,17266
	разрушение	0,23522	единственный	0,13645
	воздействие	0,23141	непереходный	0,13645
	ситуация	0,21440	английский	0,12696
	малефактив	0,20634		
	деформация	0,17266		
	структура	0,15212		
	пациенс	0,14932		
	исчисляемость	0,13645		
	преграждение	0,13645		
	причина	0,12368		
	движение	0,12311		
	случай	0,11008		
	каузатор	0,10580		

6. Алпатов В.М. История европейской лингвистической традиции

КС	МкС		МкП	
история языкознания	структурализм	0,76976	лингвистический	1,68708
античная традиция	лингвистика	0,72379	европейский	0,72109
лингвистические методы	язык	0,51835	исторический	0,65466
	языкознание	0,45125	древнееврейский	0,37739
	традиция	0,44385	древнегреческий	0,30153
	последователь	0,38617	латинский	0,29872
	дескриптивизм	0,36439	сравнительный	0,27859
	представление	0,36439	неодушевленный	0,22849
	словоизменение	0,30153	старописьменный	0,22849
	возрождение	0,28495	действительный	0,22751
	изучение	0,28226	исключительный	0,22751
	человек	0,24475	французский	0,19769
	дескриптивист	0,22849	совершенный	0,19502
	происхождение	0,22849	греческий	0,19498
	глоссематика	0,22751		
	предложение	0,22751		
	человечество	0,22751		
	объяснение	0,22423		
	государство	0,21861		
	развитие	0,20113		
	грамматика	0,19562		

7. Бобунов А.М. Сопоставительный анализ наименований одежды в английском и русском песенном фольклоре

КС	МкС		МкП	
язык фольклора	наименование	1,12415	английский	0,98888
словник	существительное	0,77873	русский	0,46961
наименования одежды	словоупотребление	0,71409	синтагматический	0,46266
русский язык	характеристика	0,31773	шелковый	0,34116
английский язык	традиция	0,31591	необрядовый	0,29929
национальная специфика	исследование	0,25038	специфический	0,25038
	эстематоним	0,25038		
	фольклор	0,24645		
	предмет	0,20889		
	одеяние	0,20509		
	одежда	0,20381		
	лексема	0,20323		

8. Вдовиченко А.В. Странный "язык" новозаветного корпуса: в поисках аутентичной традиции

КС	МкС		МкП	
лингвистический статус текста	язык	0,2979	лингвистически й	2,13153
корпус Нового Завета	интерференция	0,2723	нелитературный	1,28657
некогерентность объяснительных концепций языка Нового Завета	классицист	0,2669	новозаветный	1,06180
аутентичная традиция	литературность	0,2515	грекоговорящий	0,86182
грекоговорящая иудейская диаспора	материал	0,2477	коммуникативн ый	0,81108
лингвистическая практика православной литургии	текст	0,2423	греческий	0,60297
	направление	0,2232	языковой	0,42142
			литературный	0,41932
			ненормализован ный	0,34177
			современный	0,33102
			иудейский	0,32347
			повседневный	0,31585
			эллинистически й	0,29203
			литургический	0,28385
			традиционный	0,25884
			семитоязычный	0,25152
			русскоязычный	0,23375
			невозможный	0,23155
			древнееврейский	0,20926

9. Зырянова М.В. Переводческий метатекст как способ реноминации реалий

КС	МкС		МкП	
перевод	комментарий	1,10047	французский	0,55112
реноминация	метатекст	0,80934	переводческий	0,29391
реалия	реалия	0,51381	когнитивный	0,24885
метатекст	эквивалентность	0,58013	культурный	0,24779
	реноминация	0,36771	эквивалентный	0,22001
	переводчик	0,44537	чеховский	0,19141
	перезначивание	0,45452	русский	0,09743
			языковой	0,04060

10. Никонов В.М. Теоретические и прикладные аспекты динамики коннотаций в русском языке на рубеже XX – XXI вв.

КС	МкС		МкП	
коннотация в динамике	исследование	0,65286	современный	0,71070
СМИ	коннотация	0,32191	аксиологический	0,41400
речевая культура	действительность	0,31226	языковой	0,37713
словари разных типов	характеристика	0,31226	литературный	0,34687
стилевые и стилистические пометы в словарях	ответственность	0,25461	телевизионный	0,30196
лингвистическая прогностика	принципиальность	0,25461	русский	0,22908
	личность	0,24374		
	формирование	0,20022		
	выражение	0,19120		

11. Никонов В.М. Смысл как коннотация

КС	МкС		МкП	
смысл и коннотация	микрокомпонент	0,97313	грамматический	0,66047
коннотация и внутренняя форма слова	значение	0,67422	внутренний	0,40805
ФЕ	смыслорождение	0,65594	коннотативный	0,22440
коннотация и концепт	смыслообразование	0,60725	смысловой	0,21068
	коннотация	0,60107	языковой	0,19103

12. Пименова М.В. Развитие правовых метафор совести в русском и английском языках

КС	МкС		МкП	
концепт	ответственность	1,16132	нравственный	1,60777
языковая картина мира	сознательность	0,74366	внутренний	0,55945
метафора	понятие	0,63652	русский	0,45187
символы культуры	совесть	0,56509	неправильный	0,39570
	поведение	0,54812	религиозный	0,39570
	правда	0,40069		
	представление	0,39040		
	сознание	0,38111		

13. Попова З.Д. Федоров В.А. Русские высказывания со словоформой «у кого» и их французские эквиваленты

КлСлова	МкС	СумИнТеМ	МкП	СумИнТеМ
синтаксическая позиция	высказывание	1,23940	французский	1,16624
словоформа	словоформа	1,10987	русский	0,48695
структурная схема	глагол	0,27617	структурный	0,37872
структурная схема	местоимение	0,38382	синтаксический	0,36502
эквивалент	перевод	0,24521	притяжательный	0,22406
эквивалент	переводчик	0,38110		

14. Правда Е.А. Межъязыковые проксиматы в параллелях сербского и русского, словацкого и русского языков: типы проксиматов

КС	МкС	МкП		
интерференция	предложение	0,32649	межъязыковой	0,36167
ложные ассоциации	язык	0,25623	близкородственный	0,29085
неполное сходство	выражение	0,24078	эквивалентный	0,12949
сходно-различное языковое явление	обозначение	0,23196	различный	0,12613
дифференцирующий признак	словосочетание	0,18207	подобный	0,09045
	конструкция	0,13520		

15. Правда Е.А. Межъязыковые проксиматы в параллелях сербского и русского, словацкого и русского языков: отношения между проксиматами

КС	МкС	МкП		
интерференция	существительное	2,08962	словацкий	1,11798
ложные ассоциации	проксимат	1,03686	сербский	1,06022
неполное сходство	прилагательное	0,57915	русский	0,87129
сходно-различное языковое явление	антитранспонат	0,56949	стилистический	0,54077
дифференцирующий признак	язык	0,53818	грамматический	0,48539
	соответствие	0,50075	валентностный	0,40247
	принадлежность	0,48539	симметричный	0,32406
	псевдотавтоним	0,48539	регулярный	0,30960
	исследование	0,46237	лексический	0,24471
	ассоциация	0,37672	межъязыковой	0,24471
	значение	0,37633		
	эквивалент	0,32406		
	слово	0,31320		
	окказионат	0,30960		
	повышение	0,30960		
	обозначение	0,24471		
	параллель	0,23833		
	глагол	0,20356		

16. Титов В.Т., Кретов А.А. Роль скрытых категорий в типологическом описании грамматики романских языков // Вестник ВГУ. Серия лингвистика и межкультурная коммуникация, 2010, № 1, С. 7-12.

КС	МкС		МкП	
скрытые категории грамматики	классификатор	0,83581	грамматический	2,60114
лингвистическая типология	категория	0,74701	синтаксический	0,45883
романские языки	классификация	0,65046	сочетаемый	0,22942
	существительное	0,61928	скрытый	0,20637
	противопоставление	0,61565	французский	0,20023
	определение	0,45361	именной	0,19615
	языкознание	0,37215	английский	0,19589
	язык	0,35240		
	местоимение	0,32354		
	неодушевленность	0,29006		
	класс	0,26613		
	сопоставление	0,25338		
	словосочетание	0,22942		
	грамматика	0,20791		
	кондиционал	0,20477		
	криптокласс	0,20477		

17. Щербакова М. В. Повалюхина Д.А. Формирование ценностного отношения к языковому образованию у студентов медицинского вуза

КС	МкС		МкП	
ценностное отношение студентов к языковому образованию	образование	0,7439	профессиональный	3,4643
профессионально-ориентированное обучение иностранному языку	формирование	0,7305	иностраннный	1,6705
уровни иноязычной коммуникативной компетенции	специалист	0,6093	коммуникативный	0,9527
	потребность	0,4856	личный	0,7479
	отношение	0,4629	медицинский	0,5645
	деятельность	0,3856	ценностный	0,4667
	студент	0,3189	самостоятельный	0,3678
	приобретение	0,2918	иноязычный	0,3218
	язык	0,2839	эмоциональный	0,2214
	коммуникация	0,2752	воспитательный	0,2140
	информация	0,2479	педагогический	0,2140
	компетенция	0,2454	языковой	0,1846
	средство	0,2345		
	общение	0,2296		
	обучение	0,2228		