

# Автоматизированный лингвистический анализ российского грузового автомобиля КамАЗ 5490-NEO

А.А. Исакова,  
доц., к.ф.н., доц., IsakovaAA@yandex.ru  
ТИУ, г. Тюмень

Статья посвящена лингвистической компьютерной диагностике грузового автомобиля КамАЗ 5490-NEO. Аспект предложенного исследования лежит в области инженерной и когнитивной лингвистики. Поскольку инструменты компьютерной статистики и визуализации помогают не только обрабатывать числовые данные, но и прогнозировать, диагностировать их, задача данной статьи проследить жизненный цикл имени бренда автомобиля.

The article is devoted to the linguistic computer diagnostics of terrain vehicle KAMAZ 5490-NEO. The aspect of the proposed study lies in the field of engineering and cognitive linguistics. The aim of the following article is to trace the life cycle of the name of the car brand because the tools of computer statistics and visualization help not only to process numerical data, but also to predict and to diagnose them.

Информационная эра промышленных технологий и управления движет современную цивилизацию к поиску объективных истин. Инструменты компьютерной статистики и визуализации помогают не только обрабатывать числовые данные, но и прогнозировать, диагностировать их.

По определению классиков, наука управления – «это совокупность методов формирования и обоснования способов достижения цели в конкретной ситуации». Эти методы могут включать в себя физические, математические, компьютерные модели и словесные описания» [1]. По сути, это то, на чем держится и развивается экономика, всё, что окружает нас и составляет сущность и основу бытия. Информационные технологии «толкают» социум, заставляют его обучаться и стремиться к новому. Внедрение новых аспектов способствует более комфортному существованию человека в среде обитания, настраивает на дальнейшее стремление к прогрессу. Прогресс в развитии информационно-коммуникационных технологий и методов извлечения знаний определяет современный подход к менеджменту и помогает в инженерной и управленческой деятельности в процессе управления сложных инфраструктурных систем и территорий [2].

С каждым днём увеличиваются технологические потоки, появляются новые промышленные продукты и механизмы. Создавая различные приспособления, человек незаметно становится творцом иных миров, которые живут своей обособленной жизнью. Желания создателей той или иной машины, механизма, отраженные непосредственно в названии, вдохнуть жизнь в свое изделие, могут вызвать у потенциального потребителя максимально расширенную сеть ассоциативных реакций, направленных на создание привлекательного образа «бездушного» механистического средства как продукта. Обычная «штучка», «ерундовинка», «пимпочка» вдруг обретает не просто имя, но и значимость в мире физическом и ментальном.

Актуальность данного исследования заключается в том, что оно отвечает возросшему интересу к процессам использования компьютерных знаний и их моделирующих способностей. В самых фантастических предсказаниях прошлого, отраженных в футуристических романах Ж. Верна, Г. Уэллса и других писателей, механистическому миру уделено достаточно много внимания, поскольку человек, стремясь подчинить мир техники человеческому гению, пытается упорядочить его. В этой связи так называемый мир дорог и машин требует детального анализа и изучения. Лингвистическая компьютерная диагностика автомобиля становится важной частью как при выборе, покупке автомобиля, так и при создании новых рекламных баннеров на улицах города. Для анализа взят российский грузовой автомобиль КамАЗ 5490-NEO. Мы обратились к компьютерной базе данных Национального корпуса русского языка, программе автоматизированного анализа ассоциативного эксперимента, программе ВААЛ и компьютерному анализу с интегрированной в текстовой редактор Microsoft Word 7.0 подпрограммы «Статистика удобочитаемости».

Теоретическим обоснованием нашего исследования являются фундаментальные работы по инженерной и когнитивной лингвистике. Использование алгебраических методов, методов математической статистики, логики и теории множеств позволяет не просто структурировать лингвистические данные, но и сделать языковую субстанцию более доступной для наблюдения [3]. Математическое моделирование помогает найти объяснение многим явлениям языка. Компьютерные системы – основное средство обработки, хранения и передачи огромных пластов различного рода информации. Степень научной разработанности проблемы специализированных компьютерных систем достаточно высока и разнопланова. Научные результаты учёных представлены на конференциях различного уровня по следующим направлениям: организация структур технических и программных средств проектирования и управления, средства структуры данных, виртуальная реальность, компьютерная лингвистика и т.д. Интересны научно обоснованные подходы многих ученых, в частности, представлены результаты разработки специализированной компьютерной системы для создания учебных программных систем и интерактивных технических руководств и средств виртуальной реальности [4], ряд работ посвящено основным принципам компьютерного анализа текстов на трех уровнях — морфологическом, синтаксическом и семантическом с использованием словарей, программ и корпусов текстов [5]. Интеллектуальные системы широко представлены в гуманитарной сфере в работах А. В. Гладкого, в прикладной лингвистике А.Н. Баранова, Н. Н. Леонтьевой и др. Компьютерный лингвистический анализ основан на принципах инженерной (математической) лингвистики. Это особенная сфера научно-практической деятельности, связанная с применением точных методов в языке. Как известно, понятие «компьютерная лингвистика» вошло в обиход в начале 70-х годов прошлого столетия благодаря появившейся возможности превращения чисто гуманитарной науки в логически строгую дисциплину [6].

Выбранный нами аспект научного поиска лежит в области ментальной информации, Дж. Фодор определил её как внутренний код [7], Дж. Миллер и Ф. Джонсон-Лэрд назвали концептуальной структурой [8], это то, что, по убеждению Р. Джекендоффа, должно изучаться с позиций когнитивной теории [9].

Цель исследования – комплексный компьютерный анализ грузового автомобиля КамАЗ 5490-NEO с использованием методов компьютерной лингвистики с учетом региональных особенностей. Наша гипотеза заключается в том, что КАМАЗ является автомобилем номер 1 в России среди грузового транспорта. Для того, чтобы подтвердить или опровергнуть данное утверждение мы обратились к компьютерной базе данных Национального корпуса русского языка. Корпус содержит как письменные тексты (художественные, мемуары, публицистика, научная, религиозная литература, повседневная печатная продукция), так и записи устных текстов (публичной речи и частных бесед).

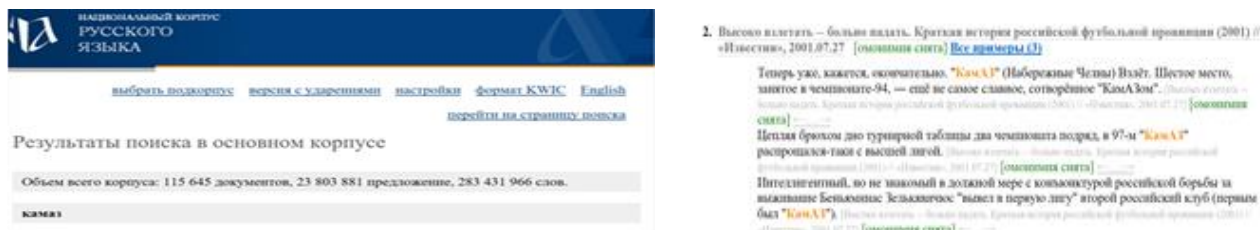


рис. 1 Программа Национального корпуса языка

Для большей наглядности все данные были объединены в таблицу. Как показано в таблице, данный номен употребляется в Национальном Корпусе 1796 раз, при этом модификация КамАЗ 5490-NEO – 1 раз. Корпусные статьи датируются 1994-2017гг., негативная коннотация – 2 газетных текста, посвященных дорожным происшествиям, что в процентном соотношении составляет 0,1 % к общему числу текстов. Популярность бренда очень высока, гипотеза о первенстве автомобиля КАМАЗ среди большегрузных машин подтверждена Национальным корпусом. Модификация Neo предполагает улучшенный вариант предыдущих моделей, практически во всех словарях толкуется как новизна, обновленность. Необходимо отметить, что написание слова латинскими буквами - это верный маркетинговый ход, поскольку слово Neo на русском языке в словаре Ушакова имеет слегка негативное значение [10.]

Таблица 1

Виды подкорпусов Национального Корпуса Русского языка	Общая характеристика	Гендерная соотносе- нность	Частота употреб- ления
Основной	Прозаические тексты	180/21	201
Акцентологический	Тексты с информацией об ударении	2	2
Газетный	Статьи из СМИ	1109/ 463	1572
Диалектный	Диалектная речь регионов России	-	-
Мультимедийный	Видео и аудио тексты	-	-
Обучающий	Произведения из школьной программы	-	-
Параллельный	Переводы	-	-
Поэтический	Стихотворные произведения	2/0	2
Синтаксический	Полная морфологическая и синтаксическая структура	1/0	1
Устный	Расшифровки записей устной публичной и частной речи	11/7	18
Итого		1305/491	1796

Следующим этапом стал автоматизированный анализ ассоциативного эксперимента. Респондентами стали согласно выборке 100 чел. Чтобы результаты были достоверны, использовалась формула для расчета объема выборочной совокупности исследования:

$$n = \frac{Nt^2 PQ}{N\Delta^2_p + t^2 PQ}$$

- n - величина выборочной совокупности;
- t - коэффициент репрезентативности;
- N - величина генеральной совокупности;
- P - частота признака в генеральной совокупности (%).

При компьютерной обработке данных опроса основным критерием в пользу грузового КамАЗ 5490-NEO стали - мощь, сила, эффективность и производительность.

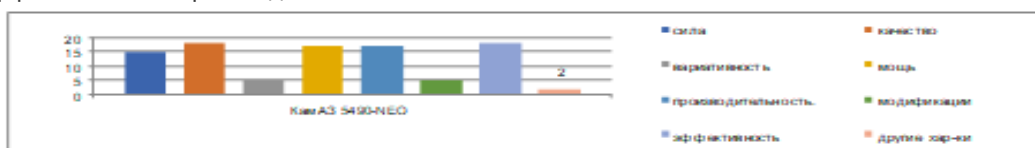


рис 2 Сравнение статистических данных

В центре нашего внимания – интернет и газетные тексты, в которых упоминается автомобиль КамАЗ 5490-NEO. Целевая группа данных текстов – в основном, профессиональные водители и автолюбители. Нами был выбран метод анализа с интегрированной в текстовой редактор Microsoft Word 7.0 подпрограммы «Статистика удобочитаемости». Графическая и статистическая обработка анализа была проведена с поддержкой Microsoft Excel 7.0. Подпро-

грамма "Статистика удобочитаемости", которая дает среднее число символов, слов, предложений и оценивает легкость чтения текста. Эти показатели характеризуют текст на подготовленность читателя для его восприятия [11].

**Уровень образования** базируется на образовательном индексе Флеша-Кинсайда (Flesch-Kincaid Grade Level) и указывает степень образования, необходимого для чтения предложенного текста. Подсчет основан на вычислении среднего числа слогов в слове и слов в предложении. Диапазон – от 0 до 20. 0-10 – школьный класс. 11 - 15 – вузовский курс. 16-20 – сложные научные тексты. Рекомендуемый диапазон значений – 8-10.

**Благозвучие.** Удобочитаемость с точки зрения фонетики. Подсчет основан на вычислении среднего количества шипящих и свистящих согласных – 0-100. Рекомендуемый диапазон – 80-100.

**Легкость чтения.** Данный показатель основан на индексе легкости чтения Флеша (Flesch Reading Ease). Подсчитывается среднее число слогов в слове и слов в предложении – 0-100. Чем выше значение, тем легче прочесть текст, тем он будет более понят читателем. Рекомендуемый диапазон – 70.

Таблица 2

Статистическое сопоставление изученных текстов о КамАЗ, КамАЗ 5490-NEO

Среднее значение (n=7)	Количество слов в тексте	Уровень образованности	Легкость чтения
	52.1	4.4	94.8

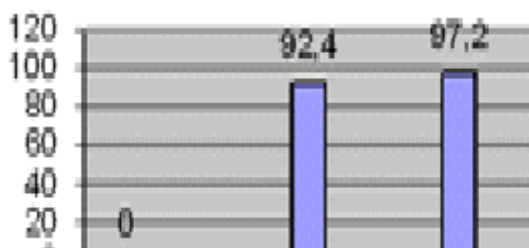


рис 3 Лёгкость восприятия текстов (КАМАЗ, КАМАЗ NEO)

Таблица 3

Средние показатели удобочитаемости

условное название	уровень образования	показатель легкости восприятия	Благозвучие
КАМАЗ, КамАЗ 5490-NEO	4.6	95.2	89.8

Итак, проанализируем полученные данные. Благозвучие текста определяется количеством шипящих и свистящих согласных. Уровень благозвучия текстов, предлагающих сведения о грузовом автомобиле КАМАЗ очень высок – оптимальный диапазон значений от 80 до 100. Распределение параметра легкости чтения-восприятия текста, приведенное в таблице 3, также отличается линейностью. Тексты воспринимаются адекватно, нет трудностей с пониманием и сохранением информации, реклама четко структурирована и нацелена не только на потенциальных покупателей, но и представляет собой так называемую имидж-рекламу, свойственную обычно зарубежным автомобилям таким как Ferrari, Oldsmobile и др. Автомобиль «КАМАЗ» выпускается на Камском автобусном заводе, его товарная марка получила свое название в честь этого завода. Один из немногих грузовых автомобилей, который имеет эмоциональную рекламу. Рекламные слоганы: «Танки грязи не боятся», «КАМАЗ – король дорог». При обработке текста компьютерная программа выдала достаточно высокий показатель – 95.2. Измерение уровня образованности, требующегося для наиболее эффективного восприятия текстов, показало, что он соответствует примерно 4, 5 классу средней общеобразовательной школы, легкость чтения тоже высока и находится вблизи верхней границы рекомендуемых значений, то есть читабельность и воспринимаемость текстов на высоте. Значение параметра благозвучия для всех обработанных текстов практически идеально соответствует оптимальному значению (90), что свидетельствует о сбалансированности языка по шипящим и свистящим согласным.

Для определения фонетического состава слова мы использовали программу SPSS Statistics (компьютерная программа для статистической обработки данных). Номен КАМАЗ четко сбалансирован, по фонетическому составу в общем не отличается от разговорной речи. Если взять звуковой состав согласных в слове, то в процентном соотношении это выглядит следующим образом:

Таблица 4

Анализ фонетического состава слова

Звуки	КАМАЗ	КАМАЗ НЕО
Гласные	40%	50%
Звонкие согласные	20%	12.5%
Глухие согласные	20%	12.5%
Сонорные согласные	20%	25%
Итого:	100%	100%

В обычной речи процент употребления сонорных звуков составляет примерно 29. В слове КАМАЗ процент гласных звуков одинаков, то есть соответствует литературно-речевой норме, в слове КАМАЗ НЕО сонорный звук «н» придает названию большей звучности, что в конечном итоге говорит о правильности подобранного словосочетания.

Для оценки фоносемантики слов и текстов мы выбрали программу ВААП, которая на основе 20 шкал звукоряда дает эмоциональную оценку заданного текста, и программу Диатон для определения ритмов головного мозга. Данная программа выдает влияние текста статьи на подсознание человека с учетом его жизненного опыта.

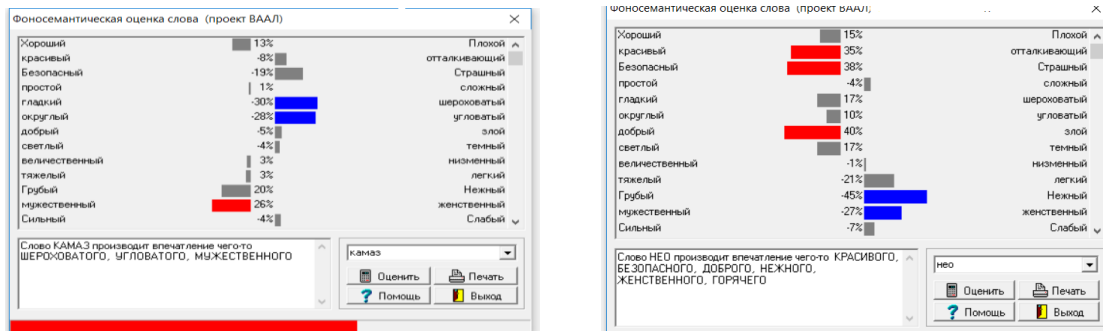


рис 4 Фоносемантический анализ слов

Обычно разграничиваются два аспекта значения слова: значение, замещающий психический феномен и значение как субъективный факт, зависящий от сознания носителя языка своей речи. Анализ качества языковых моделей в ракурсе фоносемантики дал следующие результаты: слово КАМАЗ производит впечатление чего-то шероховатого, угловатого, мужественного, слово КАМАЗ НЕО – красивого, безопасного, доброго, нежного, женственного, горячего. Таким образом, оба бренда положительно влияют на сознание человека и вызывают в воображении перцептивный образ, которым, по сути, и является российский грузовой автомобиль КАМАЗ, КАМАЗ НЕО. Более того, бренд КАМАЗ НЕО имеет более мягкую суггестию, отлично гармонирует с предыдущим брендом и дополняет его.

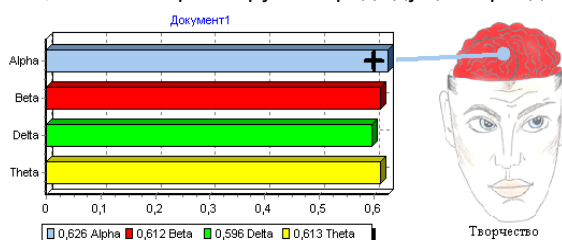


рис 5 Анализ ритмов мозга

Альфа-плюс текст. Погружает человека в состояние расслабленности, медитации; пробуждает интуицию, творчество; способствует нестандартным формам общения.

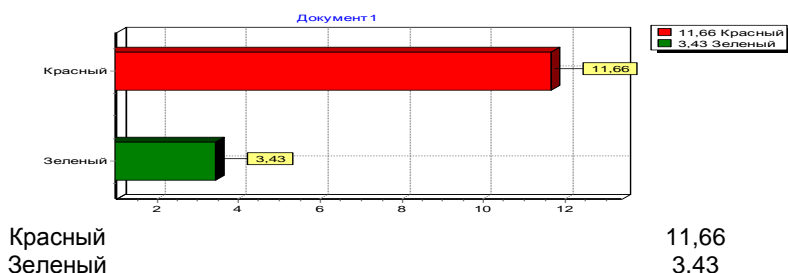


рис. 6 Цвето – звуковые ассоциации

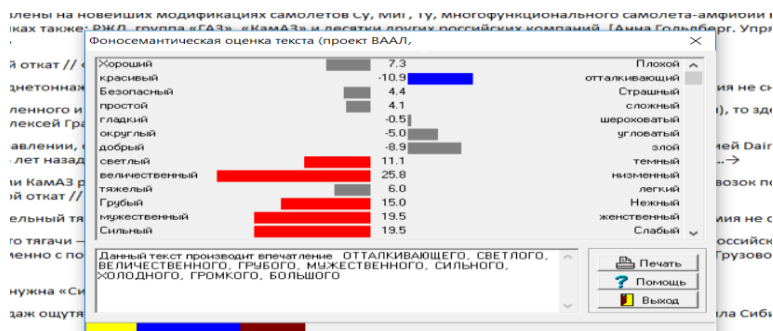


рис. 7 Анализ выборочных текстов

Основные параметры восприятия текстов, выданные программой, - величественный, светлый, сильный, мужественный, есть и отрицательная реакция – отталкивающий, грубый. Качественные характеристики преобладают и позволяют сделать вывод, что данные тексты полностью соответствуют требованиям к написанию такого рода сообщениям, а значит, эффективно выполняют свои основные функции – создание имиджа автомобиля, высокий рейтинг предприятия, положительная коннотация у респондентов и, в итоге стимулируют потенциальных клиентов к покупке именно этого грузового автомобиля. Подобные внесистемные, специфические лексические образования вполне отвечают предъявленному к тексту основному требованию - максимум информации при минимуме слов.

Следующим этапом нашего исследования стал анализ ассоциативного ряда. При проведении ассоциативного эксперимента мы выявили следующие ассоциации на слово-стимул «КамАЗ 5490-NEO» - мужские ассоциации/ женские ассоциации: труднопроходимый (0/1), мощь (3/2), большой (0/4), детская игрушка (0/1), дальнобойщик (2/0), фу-

ра (1/0), авария (0/2), слон (0/1), роботяга (1/0), дина (0/2), машина (0/2), проходимость (1/0), грязь (0/2), неповоротливый (1/0), тяжесть (0/1), на луне (0/2), грязь (0/1), сила (1/0), тяжесть (1/0), песочница (1/0), зверь(1/0), тонна (1/0), медленный (0/1), авария (0/1), новая модель (1/0), новое авто (1/0), мощный (1/0), двигатель (1/0), выносливость (0/1), строительство (1/0), мустанг (1/0), мощный (1/0), советское кино (0/1), Россия (1/0), Карабас Барабас (1/0), эпоха (1/0), только вперед (1/0), красный (1/0), таз (1/0), хороший грузовик(1/0), автомобиль (2/1), Дакар (1/0), песок(1/1), река (1/2), стройка (0/1), татарин (1/0), русская машина (1/0), КАМАЗ – король дорог (2/1), боярин (1/0), грузовик (2/1), грузовая машина (1/2), машина, собирающая мусор у дома (0/1), самосвал (1/3), новый (1/0), лидер (0/1), дорога (0/1), маза(0/1), полный водитель (0/1), лошадь (1/0), танки грязи не боятся(1/5), поезд(1/0), сильный (1/2), алмаз (0/1), Набережные челны (1/0), Париж – Дакар (1/1), Франция (1/0). Автоматизированный анализ ассоциативного эксперимента дает возможность профессионалам реально оценить востребованность и популярность бренда. На эту популярность влияют многие факторы, как экстралингвистического характера (например, репутация страны в деловом мире в целом), так и собственного лингвистического (благозвучность, ритмичность, краткость и т.п.). Так, например, создатели автомобильных средств большое внимание уделяют национальной и социокультурной специфике той страны, где будет продаваться данный автомобиль.

Таблица 5

#### Результаты обработанных данных

Слово-стимул	Положительная коннотация		Отрицательная коннотация		Процент узнаваемости бренда		Процент эффективности бренда
	50/	/48	0/	/2	100%/	/100%	
КамаЗ 5490-NEO							98.2 %

Таким образом, на основе полного автоматизированного анализа можно сделать следующие выводы:

- бренд эффективен, процент эффективности очень высок, что способствует созданию имиджа автомобиля, высокому рейтингу предприятия, положительной коннотации у респондентов и, в итоге стимулируют потенциальных клиентов к покупке именно этого грузового автомобиля;
- распределение параметра легкости чтения-восприятия текстов отличается линейностью. Подобные внесистемные, специфические лексические образования вполне отвечают предъявленному к тексту основному требованию - максимум информации при минимуме слов;
- тексты рассчитаны на широкую целевую аудиторию; программа выдала самый высокий показатель узнаваемости бренда –100;
- бренд обладает мягкой суггестией, что благоприятно влияет на водителя данного транспортного средства и на сам процесс вождения;
- основные ассоциации модификации «КАМАЗ НЕО» дополняют и практически тождественны с ассоциативным рядом КАМАЗ, т.е. дополнительное название модификации выбрано правильно. В слове КАМАЗ процент согласных звуков одинаков, то есть соответствует литературно-речевой норме, в слове КАМАЗ НЕО сонорный звук «н» придает названию большую звучность, что в конечном итоге говорит о правильности подобранного словосочетания.

Нейминг – это сложная структура, которую необходимо создателям, в лучшем случае, предугадать., поскольку жизненный цикл изделия в производстве, в нашем случае российского грузового автомобиля КамаЗ 5490-NEO, начинается с имени. И это имя подобрано правильно, адаптировано в речи согласно нормам современного русского языка, идеально для восприятия и, говоря языком рекламистов, отлично «раскручено».

Наша гипотеза подтверждена. КамаЗ 5490-NEO является товарной маркой номер один в России среди грузовых автомобилей. Следовательно, этот этап жизненного цикла КамаЗа стратегически и тактически верен.

#### Литература

1. Vassilyev S.N., Kudinovb I.Y., Kudinovb Y.I., Pashchenko F.F. Logical linguistic controllers / Procedia computer science – Elsevier (United States), № 103, с. 629-636.
2. Мильчук Я.Г., Кравец, А.Г. Подход к проектированию и реализации геоинформационных систем на основе анализа данных социальных сетей / Системы проектирования, технологической подготовки производства и управления этапами жизненного цикла промышленного продукта (CAD/CAM/PDM - 2016) С. 65-68.
3. Баранов А.Н. Введение в прикладную лингвистику: — М.: Эдиториал УРСС, 2001. — 360 с. 4.
4. Артамонов Е.И., Балабанов А.В., Ромакин В.А., Смирнов С.В. Специализированная компьютерная система для создания учебных программных систем и интерактивных технических руководств на основе объемных геометрических моделей и средств виртуальной реальности. Системы проектирования, технологической подготовки производства и управления этапами жизненного цикла промышленного продукта (CAD/CAM/PDM - 2012) Труды 12-й Международной конференции. Под редакцией Е.И. Артамонова. 2012. С. 12-17.
5. Боярский К.К., Каневский Е. А. Вега — компьютерная система классификации и анализа текстов. Lambert Academic Publishing, 2011. – 143 с.
6. Jyunichi Tsujii History of Natural Language Processing. Monthly Issue. Language & Computer. Tokyo. 2000.
7. Fodor J.A.: The language of thought. Harvard university press, Cambridge. 1975.
8. Miller G.A.: Language and perception. Harvard university press, Cambridge, 1976.
9. Jackendoff R. Languages of the mind. Essays on mental representation. MIT press, London, 1992.
10. Толковый словарь русского языка. Под ред. Д.Н. Ушакова. М.: Гос. ин-т "Сов. энцикл."; ОГИЗ; Гос. изд-во иностр. и нац. слов. 1935-1940. 4 т.
11. Isakova A.A., Manyashin S.A. Computer linguistic and technical diagnostics of the Volga car / 11<sup>th</sup> IEEE International Conference on Application of Information and Communication Technologies Russia, Moscow, 20-22 September 2017. P.336-34