

К задаче определения функционального стиля документа на естественном языке

Волкова Л. Л.

*Научно-исследовательский университет «Высшая школа экономики»,
Московский институт электроники и математики
liliyavolkova@itas.miem.edu.ru*

Аннотация. Рассмотрены проблемы определения функционального стиля текста на естественном языке. Приведена классификация и описаны некоторые причины её вариации. Подробно описаны характерные особенности разных функциональных стилей на лексическом и синтаксическом уровнях представления языка, представляющие собой почву для классификации текстов по функциональному стилю на основе вектора маркеров стиля.

Ключевые слова: машинная лингвистика, обработка естественного языка, функциональный стиль, установление стиля, автоматическая обработка текстов.

1 Введение

Данная статья посвящена определению стиля текста на естественном языке. Будет проведено детальное рассмотрение функциональных стилей и их характеристик, которые могут быть использованы как маркеры при классификации текста по признаку стиля. Задача определения стилистики не является краеугольной, но это не отменяет ее важности, в частности, в вопросах поиска. Когда пользователь ищет информацию определенного рода, при сортировке результатов поисковой системой представляется полезной для повышения релевантности их категоризация в зависимости от стиля. Также фактор стиля документа применим при анализе авторского стиля, выявлении плагиата (если подвергнуть анализу различные фрагменты исследуемого документа), реферировании (с соблюдением стиля), аннотировании текстов на естественном языке.

2 Функциональные стили

Функциональные стили русского литературного языка – это разновидности языка, определяемые сферами деятельности человека и имеющие свои нормы отбора и сочетания языковых средств в зависимости от целей и задач текста. Каждый функциональный стиль (далее ФС) отличается от других следующими признаками: сферой использования, целью общения; формами, в которых он существует; набором языковых средств. Функциональный аспект стиля предполагает единство формы и содержания, а также экстралингвистическую обусловленность текста: «наряду с видами деятельности и формами общественного сознания – функции языка; типовое содержание, характерное для соответствующей

сферы общения; цели общения, связанные с назначением в социуме той или иной формы сознания и вида деятельности и пр.» (цитируется по [Кожина, 2006], формулировка опирается на понимание стиля В.В. Виноградовым, М.М. Бахтиным и Б.Н. Головиным). Там же полагается правомочным объединение двух точек зрения касательно ФС – рассмотрение как явления речевого (текстуального) и как литературного (по Виноградову). Особняком стоит точка зрения О.А. Лаптевой о существовании устной публичной речи вне зависимости от ФС, т. е. объединяющей ряд стилей в их устной форме.

2.1 Классификация функциональных стилей

Согласно классификации, приведенной в [Кожина, 2006], в русской лингвистической и чехословацкой традиции выделяют следующие ФС, которые представлены почти во всех современных европейских (и не только) языках: *художественный* (беллетристический), *научный* (научно-технический), *официально-деловой* (законодательный), *публицистический* (газетно-публицистический, политический), *разговорный* (общедиалектный, разговорный), *религиозный*. Далее: «Закономерности функционально-стилевой системы носят вероятностно-статистический характер: они создаются не столько за счёт готовых в языке стилистически маркированных средств, сколько за счёт различной частоты употребления тех или иных языковых единиц и их организаций, речевой системности, образующейся под действием своей в каждом Ф. с. доминанты (или иначе: конструктивного принципа). Напр., для оф.-дел. Ф. с. доминантой является императивность». Основные особенности каждого из стилей на основе [Кожина, 2006; Скаженик, 2006; Козеренко, 2006], различия между ними с точки зрения смыслового содержимого и, как следствие, используемых конструкций русского языка, приведены в табл. 1.

Таблица 18. Функциональные стили и их характерные признаки

Наименование	Назначение	Общие характеристики	Языковые конструкции
Художественный	Рисование словами картины, выражение отношения к предмету, воздействие на чувства и воображение	Высокая образность, разнообразие оттенков и тонкостей речи	Использование эпитетов, описательные конструкции, индивидуальность и неповторимость сочетаний, словесный повтор в рамках одного предложения, метонимический перенос названия (белый запах), синонимы и антонимы, замена

			придаточных предложений инфинитивными конструкциями с кратким не отглагольным прилагательным (чужд любить)
Деловой	Выражение предписаний, констатация статуса, положения дел в законодательной, правовой, административной, коммерческой сферах	Точность, не допускающая инонотолкования, императивность, стандартизация речи, безэмоциональность, недопущение образности, «мучительная обстоятельность слога» (Д.С. Лихачёв)	Канцелярит, сложные предложения, императив и инфинитивы, редкость местоименно-глагольных форм 1-2 л., собирательные, отглагольные сущ., страдательные конструкции, сложные предложения, безличные и неопределенноподличные предложения
Научный	Сообщение научной информации, её объяснение с представлением стройной научной аргументации	Официальность, логичность, объективность, смысловая точность, широкое использование стереотипов специальной лексики, исключение эмоциональных эпитетов (в отличие от английского языка [Козеренко, 2006])	Редкое использование личных местоимений, частое использование причастий и деепричастий и оборотов с ними, сложные предложения, безличные и неопределенноподличные предложения; номинативность – более широкое использование существительных (что чётче выражено в русском, чем в английском [Козеренко, 2006])
Публицистический	Воздействие на массовое сознание посредством общественно значимой информации	Логичность, образность, оценочность, призываность, сочетание экспрессии и стандарта, часто установка на новизну выражения	Широкое использование эпитетов, восклицательные предложения, редкое использование причастных и деепричастных оборотов

Религиозный	Обслуживание сферы церковно-религиозной общественной деятельности согласно с религиозной формой общественного сознания, явно выраженная коммуникативная цель	Эксплицитный характер, не допускающий иностранный терминологий; архаическая стилистическая окраска речи; вкрапления публицистич. стиля	Церковнославянские вкрапления, богословская терминология, средства усиления экспрессии, модальная рамка из поздравлений, призывов, наставлений, восхвалений деятельности Церкви
Разговорный	Персонально адресованное неофициальное общение, не связанное с профессиональной деятельностью (впрочем, допускающее её упоминание), по большей части бытовая сфера; превалирование коммуникативной функции языка	Наибольшая свобода в выражении мыслей и чувств, минимум заботы о форме выражения, часто низкий стиль и варваризмы, жаргон	Эмоциональная лексика, парцелляция, тире и многоточия, восклицательные и оборванные предложения, преобладание неосложнённых предложений, широкое использование местоимений и частич, простых предлогов

Обычно различают пять функциональных стилей речи [Браславский, 2003] (за вычетом религиозного), в порядке убывания "нормативности": официально-деловой, научный, публицистический, художественный, разговорный. Эти ФС сохраняются в большинстве иностранных языков [Кожина, 2006; Trosborg, 1997]. В зависимости от прикладного характера задачи, возможно рассматривать объединение стилей или растворение одного стиля в остальных, например, художественный стиль подчас рассматривается наряду с другими, без учёта его особого статуса в системе функциональных стилей [Браславский, 2003]. Эти макростили (по А.Н. Васильевой) могут подразделяться на подстили.

2.2 Характерные черты функциональных стилей

Стоит отметить, что разговорный стиль подчас выделяют в отдельный класс в противоположность остальным ФС, либо разделяют устную и письменную разновидность языка. С одной стороны, в разговорном стиле проявляется окказиональная и персонально-личностная составляющая автора, с другой же, велик автоматизм речи, сглаживающий «идиостильные» различия между людьми и выражающийся в бедности

лексикона и обилии стереотипов и «штампов». Такая речь богата «узкоколлективными» словечками типа названий семейных блюд и мест, отмечается обилие калькированных иноязычных слов [Кожина, 2006]. Впрочем, разговорная речь интеллигенции более схожа с другими ФС. Наблюдается и обратное явление: использование разговорных элементов как средств имитации разговорной речи характерно как для публицистики, так и для художественной речи – стилей, осуществляющих функцию воздействия [Ливнева, 2004].

Разговорный стиль часто противопоставляется высокому стилю «Высокий штиль» первоначально представлял собою стилевое единство темы, содержания, жанра и определенных языковых средств – с окраской приподнятости, торжественности, что достигалось посредством книжной лексики, «пышной» метафоричности, сложности построения конструкций, риторических вопросов и восклицаний, использования инверсий. Типичным примером являются оды М.В. Ломоносова, разработавшего теорию трёх стилей. С сер. XIX в. высокий стиль существует в применении к отдельным (большей частью стилизованным) произведениям: торжественная речь и др., либо в частях текста (художественного, публицистического) для подчеркивания значимости содержания конкретного контекста посредством обращения к приподнятым тону изложения [Кожина, 2006]. В художественном тексте вообще наблюдается большое разнообразие синонимов, что не наблюдается, в частности, в деловом стиле. Пример из [Лихачёв, 2014]: «”Да” и “нет” в нашем языке... Но как много синонимов мы можем подобрать к “да” (“конечно”, “безусловно” и т. д. и т. п.) и как мало к “нет”. Соглашаемся всевозможными способами, а отвергаем немногими». Количество показатель употребления синонимов и антонимов (на материале соответствующих словарей либо лингвистических онтологий, в частности, тезаурусов) может являться одним из маркеров художественного или, в зависимости от количественного показателя этого маркера, научного ФС [Ефремова и др., 2010].

В противоположность «высокому штилю», в разговорной речи и её имитации (соц. сети, дневниковые сервисы и пр.) имеет место системное искажение языка. Д.С. Лихачёв писал: «Меня интересует психология сознательной порчи языка» [Лихачёв, 2014]. Так, намеренное искажение слов присуще авторам, использующим «олбанский» ненормативный язык, а также для имитации неграмотности либо её воспроизведения на письме. Сюда стоит отнести частотные междометия и характерные слова [Богданова-Бегларян, 2014]. В современном письменном языке (разговорный ФС) наблюдаются такие новые явления, как повторяющиеся словесные элементы [Иомдин Л., 2013] и, в частности, редупликация [Гилярова, 2013], смысловые зачёркивания [Пиперски, Сомин, 2013]

(впрочем, происходящие из художественной литературы XIX-XX вв.) и другие изменения в бытовой терминологии [Иомдин Б., 2013].

Для научного и публицистического ФС, а также для некоторого подмножества делового ФС характерно употребление профессионального сленга. Одно из определений, упоминаемых в [Лихачёв, 1964], толкует «арго как результат некоего “contrat social”, заключаемого арготирующей группой с целью скрытия своих замыслов и действий от могущих их подслушать представителей чуждых слоёв населения». В противоположность этому, там же приводится социологическое толкование: «Арго рассматривается как явление языка социального коллектива, не выходящее за пределы обычных лингвистических норм, как явление первичной, основной и естественной для данного социального коллектива речи». Жаргонные слова являются характерными, их можно найти в специализированных словарях либо выделить методами автоматического выделения терминов в тексте [Клышинский, Кочеткова, 2014; Шелманов, 2012] на материале корпуса текстов из узкоспециальной области. Отличие термина от жаргонизма на уровне таких методов не делается: это понятие, свойственное лингвистам, чётко различающим жargon и технические выражения, термины, впрочем, в настоящее время в узкоспециальных областях жаргонизмы тесно переплетаются с собственно проф. терминами, например, «отдебажить», «залить» против «отладить», «загрузить». Для целей автоматического определения ФС использование определённой категории лексики, профессионального сленга, будет маркером без разделения на устоявшиеся термины и жаргон. Наличие таких терминов будет свидетельствовать в пользу того или иного жанра; низкочастотные употребления также могут иметь место в разговорном ФС, если автор является носителем профессионального арго.

Много исследований посвящено обнаружению количественных закономерностей, позволяющих различать между собой разные литературные жанры [Fucks, 1955]. В статистических исследованиях количественных параметров авторского стиля выявляются "значимые различия" авторов, однако вопрос выбора таких различий составляет трудность статистического подхода к установлению стиля [Фоменко, Фоменко, 1996]. Тогда как есть устоявшаяся тенденция использования спектров отдельных распорядительных слов [Морозов, 1915; Meier, 1964], при больших объемах выборок результаты такого подхода должны "колебаться около среднего числа, подчиняясь общим законам языка", что затрудняет различение авторов [Марков, 1916]. Впрочем, по исследовании на основе большого числа произведений писателей XVIII-XX вв. статистически доказано, что доля всех служебных слов в длинном прозаическом произведении является т.н. авторским инвариантом [Фоменко, Фоменко, 1996].

Для выделения отличительных особенностей ФС проводится статистический анализ морфологии и лексики (в частности, на основе данных [Сиротинина, 1983; Сиротинина, 1974] для разговорной речи). Так, проводят частотный анализ русского языка [Salton, Buckley, 1988; Засорина, 1977]. Морфологические параметры (частеречный спектр текста) принадлежат к маркерам функционального стиля и поэтому могут быть использованы для автоматической классификации текстов по стилям [Браславский, 2003]. Полезным оказывается автоматический классификатор функционального стиля текста на основе спектров длин слов – Худломер (важным представляется аспект зависимости успешности работы последнего метода от длины текста) [Делицын, 2003]. Анализ стилистических погрешностей, например, паронимии [Кирсанов, 1995], позволяет получить маркер не художественной литературы, вопрос различия по максимальному показателю для разговорной речи и делового стиля требует исследования. Богатство используемого словаря само по себе является маркером стиля [Stamatatos et al., 2000].

2.3 Синтаксические особенности функциональных стилей

Именно текст... является релевантной единицей стилистических исследований. М.А.К. Хэллидей.

Структура текста с синтаксической точки зрения также представляет интерес. Так, характерным для русскоязычного научного и делового текста является употребление безличных и неопределенно-личных предложений¹ [Козеренко, 2006]. Анализа требует калькирование на русский герундия английского языка.

При рассмотрении предложения как цепочки слов информация о его синтаксическом строении представляется набором сведений о «главенствовании» одних слов цепочки над другими, что описывается некоторым графом на множестве точек цепочки (при этом при учёте проективности это не обязательно граф подчинительных связей [Большакова и др., 2011]). То, что этот граф является деревом, есть эмпирический факт: практически всегда, когда изображение структуры предложения в виде набора сведений о главенствовании одних вхождений слова над другими достаточно адекватно, соответствующий граф оказывается деревом [Гладкий, 2007]. Однако при развёртке графа в линейную структуру порядок слов не всегда соблюдается в русском языке, в отличие от английского. Разорванные и непроективные конструкции [Касевич, 1977] достаточно часто возникают в языках со свободным порядком слов, анализ нарушений проективности и глубины гнездования [Гладкий, 2007] представляется полезным при разделении

¹ При анализе иноязычных текстов может быть полезным тот факт, что такие конструкции совпадают с типичными переводами людей, изучающих английский язык на начальных стадиях – имеются в виду попытки сделать грамматические кальки безличных конструкций русского языка [Козеренко, 2006].

функциональных стилей (так, большая глубина гнездования более характерна для художественного стиля). Индивидуальные особенности построения речи могут анализироваться путём сравнения деревьев зависимостей в текстах разного ФС [Wang, Zhang, 2008].

Интерес представляет теория актуального членения предложения, или теория о правильной расстановке акцентов в предложении [Мухортов, 2012]. Если в устной речи рему можно выделить интонацией, то на письме это часто выражается в эмфазе на последнем слове предложения (через порядок слов) или через слова-маркеры, в том числе эмфатические обороты, например, *и только X, it is X that* (англ.), *c'est X que* (фр.) и т. п. Выделение ремы как коммуникативного центра высказывания в сочетании с анализом проективности может дать почву для выделения маркера стиля.

Стилевые особенности речи могут определяться литературным жанром – типом литературного произведения (художественного, публицистического, научного и др.), напр., роман, монография, репортаж и т.д. [Кожина, 2006]. Разделение жанров позволяет конкретизировать задачу, например, выделить текст, преимущественно состоящий из диалогов. В этом частном случае следующим этапом будет выбор между пьесой и интервью, то есть сужение области выбора до художественного и газетного стилей. Особняком стоят энциклопедии и справочники, что определяется по пунктуации, выделениям в тексте и доле сокращений.

3 О задаче определения стиля

Задача определения стиля по итогам выбора набора маркеров сводится к задаче классификации [Kessler et al., 1997; Kahlgren, Cutting, 1994; Biber, 1995; Stamatatos et al., 2000], для которой текст представляется вектором признаков (маркеров двоичных и/или количественных). Можно применить бинарные меры близости к документам, в частности ассоциативные и корреляционные меры [Волкова, Строганов, 2014]. Оценка может быть качественной (принадлежность или не принадлежность конкретному ФС), а может быть нечёткой [Тэрано и др., 1993]. Это не только позволит дать более подробную оценку принадлежности пограничного по стилю текста, но и перейти к жанровой подзадаче, то есть усложнению системы стилей. Со временем система претерпевает изменение, так, в современной разговорной речи могут оформиться отдельные подстили (см. раздел 2.2). Так, «при изменении взаимоотношений между разными стилями, когда трихотомическое членение литературного языка на стили (высокий, средний, низкий) не противоречит другому возможному членению, а накладывается на него, условно можно говорить о замещении "высокого штиля" средним, среднего - низким, а на место низкого приходит просторечие. В такой ситуации можно говорить о преобразовании, усложнении трех штилей» [Семёнов 2002]. Если разрабатывать систему машинного обучения, многое

зависит от корпуса текстов, поэтому корпус должен быть репрезентативен. В дальнейшем при развитии семантического анализа могут быть использованы смысловые структуры и их соотношения, а также характеристики когнитивной функции текста.

4 Список литературы

[Богданова-Бегларян, 2014] Богданова-Бегларян Н. В. Об одной из самых частых единиц русской спонтанной речи: блин с лингвистической и социолингвистической точек зрения. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2014 г.). Вып. 13 (20). М.: Изд-во РГГУ, 2014. С. 76–82.

[Большакова и др., 2011] Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие. М.: МИЭМ, 2011.

[Браславский, 2003] Браславский П. Опыт автоматической классификации текстов по стилям (на материале документов Internet)//Русский язык в Интернете. Сб. статей. Казань, 2003. С. 6–15.

[Браславский, Соколов, 2008] Браславский П.И., Соколов Е.А. Сравнение пяти методов извлечения терминов произвольной длины. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (4–8 июня 2008 г.). Вып. 7 (14). М.: РГГУ, 2008. С. 67–74.

[Волкова, Строганов, 2014] Волкова Л.Л., Строганов Ю.В. Об ассоциативных бинарных мерах близости документов: классификация и приложение к кластеризации. / Новые информационные технологии в автоматизированных системах: материалы семнадцатого научно-практического семинара. М.: ИПМ им. М.В. Келдыша, 2014. С. 421–432.

[Гилярова, 2013] Гилярова К.А. Статья такая статья. Об одном типе редупликации в современном русском языке. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). Т. 1: Основная программа конференции. М.: Изд-во РГГУ, 2013. С. 256–269.

[Гладкий, 2007] Гладкий А.В. Синтаксические структуры естественного языка. Изд. 2 М.: ЛКИ, 2007.

[Делицын, 2003] Делицын Л. Худломер. Автоматическая классификация стиля текстов. URL: <http://teneta.rinet.ru/hudlomer/article.html> (дата обращения 25.03.2003).

[Ефремова и др., 2010] Ефремова Н.Э., Большакова Е.И., Носков А.А., Антонов В.Ю. Терминологический анализ текста на основе лексико-синтаксических шаблонов. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26–30 мая 2010 г.). Вып. 9 (16). М.: Изд-во РГГУ, 2010. С. 124–129.

[Засорина, 1977] Частотный словарь русского языка / Под. ред. Л. Н. Засориной. М.:

Русский язык, 1977.

[Иомдин Б., 2013] Иомдин Б. Л., Лопухина А. А., Панина М. Ф., Носырев Г. В., Вилл М. В., Зайдельман Л. Я., Матиссен-Рожкова В. И., Винокуров Ф. Г., Выборнова А. Н. Маг вел мот: изменения в языке на материале бытовой терминологии. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). Т. 1: Основная программа конференции. М.: Изд-во РГГУ, 2013. С. 311–324.

[Иомдин Л., 2013] Иомдин Л. Л. Читать не читал, но.... об одной русской конструкции с повторяющимися словесными элементами. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). Т. 1: Основная программа конференции. М.: Изд-во РГГУ, 2013. С. 297–310.

[Касевич, 1977] Касевич В.Б. Элементы общей лингвистики. М., Наука, 1977.

[Кирсанов, 1995] Кирсанов Д. Свежий взгляд/Fresh Eye, версия 1.21, 1995. URL: <http://www.kirсанов.com/fresheye/> (дата обращения 25.03.2015).

[Клышинский, Кочеткова, 2014] Клышинский Э.С., Кочеткова Н.А. Метод извлечения технических терминов с использованием меры странности. / Новые информационные технологии в автоматизированных системах: материалы семнадцатого научно-практического семинара. М.: ИПМ им. М.В. Келдыша, 2014. С. 365–370.

[Кожина, 2006] Стилистический энциклопедический словарь русского языка / под ред. М.Н. Кожиной; члены редколлегии: Е.А. Баженова, М.П. Котюрова, А.П. Сковородников. 2-е изд., испр. и доп. М.: Флинта, Наука, 2006.

[Козеренко, 2006] Козеренко Е.Б. Проблема эквивалентности языковых структур при переводе и семантическом выравнивании параллельных текстов // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» (Бекасово, 31 мая – 4 июня 2006) / Под ред. Л.Л. Иомдина, Н.И. Лауфера, А.С. Нариньяни, В.П. Селегея. М.: Изд-во РГГУ, 2006. С. 252–258.

[Ливнева, 2004] Ливнева Л. Средства имитации разговорной речи в художественном произведении. URL: <http://www.stihi.ru/2004/01/05-20> (дата обращения: 25.12.2004).

[Лихачёв, 1964] Лихачёв Д.С. Арготические слова профессиональной речи // Развитие грамматики и лексики современного русского языка / АН СССР: Ин-т рус. яз. М.: Наука, 1964. С. 311–359.

[Лихачёв, 2014] Лихачёв Д.С. Заметки о русском. М.: Колибри, Азбука-Аттикус, 2014.

[Логичев, 2013] Логичев С.В. Русская виртуальная библиотека: Каталог лингвистических программ и ресурсов в Сети. URL: <http://www.rvb.ru/soft/catalogue/catalogue.html> (дата обращения: 19.03.2013).

[Лукашевич, 2010] Лукашевич Н.В. Квазисинонимы в лингвистических онтологиях. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26–30 мая 2010 г.). Вып. 9 (16). М.: Изд-во РГГУ, 2010. С. 307–312.

[Марков, 1916] Марков А.А. Об одном применении статистического метода. / Известия Академии Наук. Сер. 6, т. X, вып. 4, 1916.

[Морозов, 1915] Морозов Н.А. Лингвистические спектры / Известия Академии Наук. Отделение русского языка и словесности. Кн. 1–4, т. XX, 1915.

[Мухортов, 2012] Мухортов Д.С. Практика перевода: английский – русский. Учебное пособие по теории и практике перевода. М.: Книжный дом «ЛИБРОКОМ», 2012.

[Пиперски, Сомин, 2013] Пиперски А.Ч., Сомин А.А. Литуративы в русском интернете: семантика, синтаксис и технические особенности бытования. / Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). Т. 1: Основная программа конференции. М.: Изд-во РГГУ, 2013. С. 605–617.

[Семёнов, 2002] Семёнов П.А. Эволюция стилистических категорий в русском литературном языке предпушкинского периода: На материале драматургических текстов. Дисс. докт. филол. н. М., 2002.

[Сиротинина, 1983] Разговорная речь в системе функциональных стилей современного русского языка. Лексика / Под ред. О. Б. Сиротининой. Саратов: Изд-во Сарат. ун-та, 1983.

[Сиротинина, 1974] Сиротинина О. Б. Современная разговорная речь и ее особенности. М.: Знание, 1974.

[Скаженик, 2006] Скаженик Е.Н. Деловое общение. Учебное пособие. Таганрог: Изд-во ТРТУ, 2006.

[Тэрано и др., 1993] Тэрано, Т., Асаи, К., Сугэно, М. Прикладные нёчеткие системы. М.: Мир, 1993.

[Фоменко, Фоменко, 1996] Фоменко В.П., Фоменко Т.Г. Авторский инвариант русских литературных текстов. Предисловие А.Т. Фоменко. // Фоменко А.Т. Новая хронология Греции: Античность в средневековье. Т. 2. М.: Изд-во МГУ, 1996. С. 768–820.

[Хмелёв, 2000] Хмелёв Д.В. Распознавание автора текста с использованием цепей А.А. Маркова. / Вестник МГУ, сер. 9: Филология, № 2, 2000. С. 115—126

[Шелманов, 2012] Шелманов А.О. Метод автоматического выделения многословных терминов из текстов научных публикаций // Труды тринадцатой национальной конференции по профессиональному интеллекту с международным участием КИИ–2012. Белгород: БГТУ, 2012, т. 1. С. 268–274.

[Biber, 1995] Biber D. Dimensions of register variation: a cross-linguistic comparison. Cambridge University Press, 1995.

[Fucks, 1955] Fucks W. Mathematical Theory of Word-Formation. London, 1955.

[Kahlgren, Cutting, 1994] Karlgren J., Cutting D. Recognizing text genres with simple metrics using discriminant analysis. // Proceedings of the 15th International Conference on Computational Linguistics (COLING '94), 1994. P. 1.071–1.075.

[Kessler et al., 1997] Kessler B., Nunberg G., Schütze H. Automatic detection of text genre. // Proceedings of 35th Annual Meeting. Association for Computational Linguistics, 1997. P. 32–38.

[Meier, 1964] Meier H., Deutsche Sprachstatistik. Hildesheim, 1964.

[Salton, Buckley, 1988] Salton G., Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 1988, pp. 513—523.

[Stamatatos et al., 2000] Stamatatos, E., Fakotakis, N., Kokkinakis, G. Automatic text categorization in terms of genre and author. // Comput. Linguist. 26, 2000. P. 471–495.

[Trosborg, 1997] Trosborg A. Text Typology: Register, Genre and Text Type. // Text typology and translation. John Benjamins Publishing, 1997. P. 3–23

[Wang, Zhang, 2008] Wang L., Zhang K. Space efficient algorithms for ordered tree comparison. Algorithmica, July 2008, Vol. 51, Issue 3. P. 283–297.