

ВЫСШАЯ ШКОЛА
ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ПРИ УЧАСТИИ
ВСЕМИРНОГО БАНКА



XVII

**АПРЕЛЬСКАЯ
МЕЖДУНАРОДНАЯ
НАУЧНАЯ КОНФЕРЕНЦИЯ**
по проблемам развития
экономики и общества

4



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ПРИ УЧАСТИИ ВСЕМИРНОГО БАНКА

XVII АПРЕЛЬСКАЯ МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ ПО ПРОБЛЕМАМ РАЗВИТИЯ ЭКОНОМИКИ И ОБЩЕСТВА

В четырех книгах

*Ответственный редактор
Е.Г. Ясин*

4



Издательский дом
Высшей школы экономики
Москва, 2017

УДК 330.101.5(063)
ББК 65.012
С30

*Идеи и выводы авторов не обязательно отражают
позиции представляемых ими организаций*

Опубликовано Издательским домом Высшей школы экономики
<<http://id.hse.ru>>

ISBN 978-5-7598-1572-3 (кн. 4: в обл.)
ISBN 978-5-7598-1636-2 (кн. 4: e-book)
ISBN 978-5-7598-1568-6

© Национальный исследовательский
университет «Высшая школа
экономики», 2017

СОДЕРЖАНИЕ

СОЦИОКУЛЬТУРНЫЕ ПРОЦЕССЫ

Г.А. Монусова

Объективная и субъективная социальная мобильность: межстрановые сопоставления	11
--	----

Е.В. Пруцкова

Связь религиозности и базовых ценностей в России (по данным Европейского социального исследования и всероссийского исследования «Ортодокс Монитор»)	23
---	----

М.С. Фабрикант, В.С. Магун

Гордость за страну и базовые ценности россиян	31
---	----

ОБЩЕСТВО И СОЦИАЛЬНАЯ ПОЛИТИКА

В.Н. Бобков

Неустойчивость занятости — глобальная проблема современности	41
---	----

О.И. Бородкина, М.М. Иванова

Миграция и социальная политика: пример социального обслуживания мигрантов в Санкт-Петербурге	50
--	----

Т.А. Гасс

Бедность и неравенство населения России в жилищной сфере	58
---	----

В.В. Карачаровский, Е.Н. Гасюкова

Прекариат как новая зона риска в российском обществе: опыт эмпирического анализа	67
--	----

А.В. Марченко, А.Л. Абаев, В.Р. Пратусевич

Восприятие социальной рекламы, посвященной вопросам потребительского поведения москвичей на финансовом рынке	79
--	----

Т.Ю. Сидорина

Место и роль России в системе трендов мирового развития	91
--	----

Е.А. Тарасенко, В.А. Козлов, Т.Б. Рыжкова	
Социальная политика в области повышения финансовой грамотности социально незащищенных и малоресурсных групп населения: зарубежный опыт и уроки для России	100

СОЦИОЛОГИЯ

Р.Н. Акифьева	
Структурированная активность детей в контексте стратегий их воспитания русскоязычными женщинами в Мадриде	111
Е.Я. Варшавская	
Российская NEET-молодежь: характеристики и география.....	122
Ю.Л. Войнилов, Д.В. Мальцева	
Из города в село: потенциал переселения горожан в сельскую местность.....	133
Н.В. Гончарова, Я.Н. Крупец	
Молодежь и ее карьерные пути на сельском рынке труда	146
Д.В. Мальцева, Л.В. Шубина, Ю.Л. Войнилов	
Уровень медиаграмотности населения РФ: текущее состояние и перспективы изменения (на основе всероссийского мониторинга медиаповедения)	154
О.В. Майорова, Е.В. Артюхова	
Российский рынок труда: эволюция формальных и неформальных механизмов поиска работы	173
Е.Л. Омельченко, Г.А. Сабирова	
К вопросу о молодежных культурах в современной России	183
Е.С. Петренко, Е.Г. Галицкая, Е.В. Богомолова, Ю.А. Кот	
Действующие лица российского социума и их антикризисные стратегии	192
М.О. Спирина	
Факторы оппортунистического поведения заказчиков на русскоязычном электронном рынке удаленной работы: парадокс укорененности.....	201

Е. Kovaleva, М. Spirina

Visual Representation of the Female Body in Yoga: New Femininity, Power Technology and Technology of the Self.....	222
--	-----

РАЗВИТИЕ ЗДРАВООХРАНЕНИЯ

М.А. Канева, Н.А. Авксентьев, В.М. Байдин

Регрессионный анализ факторов, влияющих на готовность платить за кооперативное медицинское страхование.....	233
---	-----

Е.В. Селезнева

Добровольное медицинское страхование как фактор доступности медицинской помощи для жителей г. Москвы.....	245
---	-----

Е.А. Тарасенко

Интеграция медицинских услуг и услуг по социальному обслуживанию для людей с хроническими заболеваниями: западный опыт и уроки для России	259
---	-----

А.Л. Темницкий

Структура и динамика определяющих мотивов труда врачей.....	268
---	-----

ОБРАЗОВАТЕЛЬНЫЕ ТРАЕКТОРИИ И СОЦИАЛЬНАЯ ДИФФЕРЕНЦИАЦИЯ

И.П. Попова

Профессиональное соответствие в трудовых карьерах работников	279
---	-----

Ф.Г. Сулейманова, И.И. Харченко

Рост возможностей для непрерывного образования городского и сельского населения региона как ресурс развития человеческого потенциала	289
--	-----

РОЛЬ ОБРАЗОВАНИЯ В ВОСПРОИЗВОДСТВЕ/УМЕНЬШЕНИИ СОЦИАЛЬНОГО НЕРАВЕНСТВА

Е. Tovar-García, А. Kamaev

Educational Achievements of Migrant Schoolchildren in Moscow	303
---	-----

ФАКТОРЫ КАЧЕСТВА ОБРАЗОВАНИЯ

М.В. Шеина

Взгляд российского учительства — «отличники», показатель успеваемости и средний балл.....	313
--	-----

ОБРАЗОВАТЕЛЬНЫЕ ИНСТИТУТЫ В УСЛОВИЯХ ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ ИЗМЕНЕНИЙ

А.Р. Андреева, Н.В. Родина

Взгляд различных участников системы образования на реформу оплаты труда педагогов и его влияние на результаты этой реформы.....	331
--	-----

О.Н. Волкова

Прозрачность в российском академическом сообществе: доверие или «общество спектакля»?	339
---	-----

Е.А. Смирнова, Е.С. Гордеева, М.В. Шеина

Эффекты сообучения в университете при изучении английского языка	350
---	-----

М.М. Степанова, А.Р. Денисов

Анализ возможности внедрения методологии Lean в управление образовательными процессами	359
--	-----

РУССКИЙ ЯЗЫК В МНОГОЯЗЫЧНОМ МИРЕ

В.Ю. Апресян

Псевдосинонимичные русские конструкции у <i>X-a</i> <i>У</i> и у <i>X-a есть У</i> в контексте изучения русского языка.....	369
---	-----

А. Бергманн, У. Яжинова

Работа с корпусом в развитии профессиональной компетенции преподавателя русского языка как иностранного: предпосылки и факторы успешного применения ресурса.....	380
--	-----

Ф. Бьяджини

Выражение уступительного межфразового отношения на продвинутом этапе обучения русскому языку как иностранному при помощи Русско-итальянского параллельного корпуса.....	390
--	-----

О.С. Еремина

Об особенностях обучения эритажников высокого уровня владения языком в смешанных группах.....	399
---	-----

О.Н. Ляшевская

К определению сложности русских текстов.....	408
--	-----

Е. Протасова, Е. Хейсканен

Особенности сохранения русского языка в Финляндии	419
--	-----

С. Славкова

Использование данных Национального корпуса русского языка в преподавании современного русского языка в иностранной аудитории	427
--	-----

Е.Л. Шнитке, З.Д. Иванова

Особенности употребления союзов в речи эритажных носителей русского языка и изучающих русский язык как иностранный	437
--	-----

A. Vyrenkova, E. Rakhilina

Learner Corpora Supporting Lexical Typology	450
---	-----

ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ

А.Л. Бекларян, А.С. Акопов

К вопросу об алгоритмах кластеризации толпы в условиях чрезвычайных ситуаций	463
---	-----

Т.К. Богданова, А.С. Есаулова

Анализ и прогнозирование расторжения договоров накопительного страхования жизни.....	476
--	-----

Д.В. Думский, В.В. Корнилов, С.В. Соловьёва, Е.А. Исаев

Построение рейтинговой системы для оценки инновационной активности регионов Российской Федерации	488
--	-----

А.А. Иванов, В.И. Якуба	
Быстрый алгоритм оценки манипулируемости правил коллективного выбора.....	495
Д.В. Исаев	
Имитационное моделирование результатов внедрения систем управления эффективностью.....	502
П.А. Колданов, В.А. Калягин, А.П. Колданов	
Устойчивые процедуры построения сетевых структур фондовых рынков.....	510
А.А. Рубчинский	
Новая схема применения автоматической классификации для анализа социально-экономических систем	517
М. Komarov, A. Nemeth	
Customer Behavior Management Using Big Data and Location Data.....	527

Введение

Интуитивное понятие сложности/легкости текста для чтения и связанной с этим скорости чтения и понимания текста в лингвистике XX в. было формализовано в виде индексов удобочитаемости (readability). В их основе лежит ряд презумпций — например, о том, что

- короткие предложения читать легче, чем длинные;
- длинные слова затрудняют чтение;
- читатель замедляется или спотыкается, встречая низкочастотные и (или) незнакомые ему слова, и т.п.

В наиболее общепринятом варианте оценка сложности текста не вытекает из незнания читателем темы, сложности материала как такового (информационная сложность), не зависит от способа шрифтового оформления, верстки блоков текста и других аспектов графического дизайна (визуальная сложность), но связана с особенностями языка текста (подбором лексики, выбором грамматических форм, строением предложений и т.п.), логической организацией текста, выстроенностью аргументации, использованием ожидаемых риторических приемов и другими аспектами организации дискурса².

Представляется, что оценка сложности также связана с субъективными факторами, такими как:

- языковой опыт (по-разному воспринимают текст носитель языка (L1), изучающий язык как неродной (L2), «несовершенный» носитель, например, носитель эритажного языка (LH), или студент, осваивающий плохо знакомый ему жанр академического письма);
- возраст носителя (например, если читающий — ребенок или пожилой человек);

¹ Исследование выполнено в рамках проекта «Стандарты оценки методов автоматического извлечения информации из текстов» при финансовой поддержке РФФИ, грант № 15-07-09306.

² Обратим внимание, что наличие таблиц и иллюстраций может также существенно облегчать понимание текста, однако взаимодействие текстовых и графических стимулов выходит за рамки рассматриваемой нами проблемы.

- мотивированность читателя;
- индивидуальные когнитивные, психологические, неврологические особенности читателя.

В связи с разнообразием ситуаций, в которых встречаются Текст и его Читатель, проводятся исследования в отдельных областях:

- оценка удобочитаемости упражнений и учебных текстов для иностранцев, изучающих язык как неродной (L2); экспертиза школьных учебников, экзаменационных тестов и других материалов (для носителей L1) — для русского языка см., например, работы [Невдах, 2010; Ткачук, Кучма, 2014; Солнышкина, Кисельников, 2015];
- оценка читабельности деловой документации [Гращенко, Мизернов, 2015]; рекламных материалов; медицинской документации;
- оценка текстов веб-сайтов с точки зрения привлекательности для целевой аудитории (например, для детей-подростков) [Рогушина, 2007];
- создание рекомендательных систем для библиотек и ряд других.

Данная работа представляет краткий обзор методов измерения сложности в свете разработки онлайн-ресурсов, позволяющих пользователю оценить любой выбранный им русскоязычный текст.

Метрики сложности

Простые метрики удобочитаемости, такие как

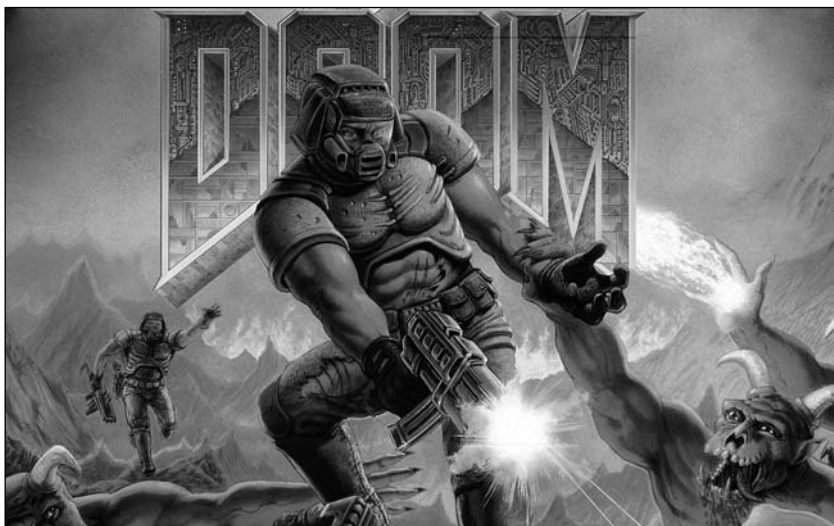
- индекс Флеша — Кинкейда [Kincaid et al., 1975];
- индекс Колмана — Лиану [Coleman, Liau, 1975];
- индекс SMOG [McLaughlin, 1969],

строятся на характеристиках, которые легко получить из текста без привлечения дополнительных лингвистических ресурсов или разметки. К таким характеристикам относятся, например, средняя длина слова в словах или слогах, количество слов длиной более пяти символов, средняя длина предложения в словах или слогах, количество знаков препинания и т.д.

В большинстве случаев метрики удобочитаемости представляют собой формулу линейной регрессии вида:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots,$$

где X_1 , X_2 , X_3 и т.д. — отдельные характеристики текста; Y — индекс удобочитаемости, который соответствует либо возрасту читателя («текст для детей возраста N лет»), либо уровню образования, либо уровню владения языком (L2), см. пример на рис. 1. В некоторых случаях индекс не имеет однозначной интерпретации, но задает известный диапазон значений, кото-



Разработчики анонсировали следующий Doom еще в 2008 г. В конце прошлого года из компании ушел «отец» серии Джон Кармак, чтобы заниматься созданием очков виртуальной реальности, так что завершают проект уже без него.

Источник: <https://vc.ru/p/doom-4>.

Индекс: 7.27, 7–9-е классы, 12–14 лет.

Рис. 1. Пример оценки удобочитаемости веб-сайта

рый позволяет сравнивать между собой текстовые материалы как в пределах одного языка, так и между языками. Коэффициенты β_0 , β_1 и т.д. подбираются таким образом, чтобы на заданной выборке текстов оценка Y наилучшим образом соответствовала оценкам, поставленным экспертами.

К текстам на русском языке обычно применяются формулы, первоначально разработанные для английского, но с адаптацией коэффициентов. Ср., например, индекс удобочитаемости Флеша для английского и русского языков:

$$\text{Flesch-ENG} = 206,835 - (1,015 \times \text{средняя длина предложения}) - (84,6 \times \text{среднее число слогов в слове});$$

$$\text{Flesch-RUS} = 206,836 - (1,52 \times \text{средняя длина предложения}) - (65,14 \times \text{среднее число слогов в слове}) \text{ [Оборнева, 2006]}.$$

Другие характеристики, которые привлекаются при оценке сложности текстов, можно разделить на лексические, морфологические, синтаксические и дискурсивные. Например, индекс Дейла — Челла [Chall, Dale, 1995]

анализирует, входят ли слова, встречающиеся в тексте, в 3 тыс. наиболее частотных слов. Аналогично может оцениваться доля слов за пределами лексических минимумов для иностранцев (A1, A2 и т.п.), средняя частота лемм (вычисляемая на основе частотных словарей), доля абстрактных слов, аббревиатур, служебных слов, лексическое разнообразие в конкретном тексте (чем чаще повторяются отдельные словоформы или леммы, тем предположительно легче текст).

Морфологические факторы включают долю разных частеречных классов в тексте, а также присутствие слов с определенной словообразовательной структурой (например, словосложений, существительных с суффиксами *-ние*, *-ость* и т.п.).

Синтаксические факторы оценивают сложность синтаксической структуры предложений, в частности, среднюю долю подчиненных, сочиненных и т.п. клауз, причастных и деепричастных оборотов.

Дискурсивные характеристики, которые используются реже всего из-за сложности их автоматического распознавания, могут учитывать среднюю долю диалогических единиц на предложение, количество анафорических местоимений и других слов, требующих для понимания разрешения кореференции, сложность риторической структуры и т.д.

Помимо линейных регрессионных формул, предлагаются и более сложные модели (ср. индекс SMOG, в котором характеристики перемножаются друг на друга, и из этого значения извлекается квадратный корень). В пользу того, что функция $Y = f(\beta, X)$ должна быть устроена нелинейно, говорят и соображения, что языковая компетенция развивается неравномерно для разных возрастных периодов и даже деградирует у пожилых и больных носителей, а важность таких факторов, как длина слова, может меняться — например, становиться малорелевантной для взрослых образованных носителей языка [Микк, 1974].

Очевидно и то, что многие перечисленные факторы не независимы друг от друга. В связи с этим неудивительно, что в последнее время появляется все больше моделей-классификаторов, которые обучаются на размеченных коллекциях текстов на большом наборе предлагаемых исследователем факторов, которых может быть несколько сотен. Так, в работе [Reynolds, 2016] модель Random Forest, работающая на 179 факторах, смогла аккуратно предсказать сложность текстов РКИ согласно принятым уровням (A1, A2, B1, B2, C1, C2) в 66% случаев, причем только в 8% она ошиблась более чем на один уровень (рис. 2).

Вместе с тем в этой же работе отмечается, что результаты машинного обучения сильно зависят от коллекции, на которой проводится оценка, и требуется разрабатывать модели, более устойчивые к ошибкам и шумам в

	A1	A2	B1	B2	C1	C2
A1	234	120	48	0	0	0
A2	41	553	192	17	0	0
B1	16	76	1130	90	5	5
B2	1	57	311	478	83	4
C1	1	20	66	98	394	6
C2	0	3	40	58	9	78

Рис. 2. Кросс-валидация предсказания уровня текстов РКИ в [Reynolds, 2016, p. 135]: предсказанные системой ответы сопоставляются с классом, приписанным тексту в тестовой коллекции

обучающей выборке (например, если текст атрибутирован экспертом неверно или если в тексте представлено одно, но очень длинное слово). Вклад каждого из предлагаемых исследователем факторов в работу классификатора также оценивается по-разному, в зависимости от выбранного способа ранжирования. Например, в одних случаях системы ранжирования факторов рейтингуют выше лексические факторы, в других — морфологические.

Экспертная оценка удобочитаемости

Как уже было сказано, метрики удобочитаемости настраиваются, как правило, по коллекциям, размеченным экспертами. Например, в [Кагров, 2014; Reynolds, 2016] использовались коллекции текстов для чтения в курсах РКИ, такие как CIE (МГУ), Red Kalinka: «Russian books with audio», TORFL (тексты на понимание), «Златоуст» (тексты для чтения). Коллекция LingQ, собранная методом краудсорсинга, включает оценки пользователей — как правило, преподавателей РКИ. Однако из-за того, что оценку в этом случае ставит один эксперт (человек, выкладывающий текст на сайт) и критерии оценки четко не определены и могут варьировать от эксперта к эксперту, степень доверия к таким оценкам будет ниже.

В [Tanaka-Ishii et al., 2010] (для английского и японского языков) применен другой стандарт оценки, когда тексты в тесте попарно сравниваются друг с другом носителями языка. В результате создается граф с попарно упорядоченными текстами, и классификатор обучается и тестируется на его оценках.

Наконец, имеет смысл понимать разброс оценок носителей языка и изучающих язык как неродной для отдельных текстов и однородных коллекций текстов. Например, в [Кошелева, 2015] (на материале польских текстов для студентов-иностранцев) проводится сопоставление между множеством индексов удобочитаемости, разработанных для польского языка, и множеством оценок носителей. Не исключено, что разные лингвистические факторы оказываются по-разному важны для разных носителей. На материале русского языка разнообразие экспертных оценок анализируется в [Петрова, Окладникова, 2009; Шпаковский, 2012; Reynolds, 2016].

Интересны попытки включить более объективные критерии в сопоставление текстов по сложности. В пионерной работе [Микк, 1974] было предложено использовать для оценки сложности время реакции носителя при восприятии слов. В [Петрова, Окладникова, 2009], наряду с субъективными оценками сложности, получаемыми от испытуемых, используется мера времени чтения текста. В [Шпаковский, 2012] описано, что наряду с просьбой оценить сложность по шкале испытуемые получали контрольные задания: ответить на вопросы к тексту, заполнить пропуски, упорядочить абзацы по порядку.

Открытые онлайн-сервисы оценки сложности текста

В 2011 г. появился первый открытый сервис для определения уровня сложности русскоязычных текстов readability.io, который был разработан И. Бегтиным (в настоящее время размещен на странице <https://plainrussian.ru/>). Сервис использует простые индикаторы (количество символов, пробелов, слов, предложений, количество слов с более чем четырьмя слогами и т.п.). Кроме того, используются два высокоуровневых параметра: доля сложных слов и доля слов с тем или иным числом слогов в оцениваемом тексте. Сервис выдает оценку текста по пяти метрикам: индексы Флеша — Кинкейда, Дейла — Челла, Колмана — Лиану, SMOG, автоматический индекс удобочитаемости (Automated Readability Index, ARI).

Значительно больше высокоуровневых критериев используется в онлайн-сервисе, разработанном в магистерской программе «Компьютерная лингвистика» НИУ ВШЭ [Дружкин, 2015, 2016]. В частности, здесь в число критериев входит доля слов различных частей речи в тексте, доля слов в родительном падеже, доля слов на *-ция*, доля слов разных частотных рангов и многое другое. Помимо этого, используется графическая визуализация положения анализируемого текста среди других в виде гистограмм (рис. 3).



Рис. 3. Фрагмент онлайн-сервиса [Дружкин, 2015]

Тексты «Тотального диктанта» в зеркале readability

Мы проанализировали тексты, которые зачитываются участникам ежегодной всероссийской акции «Тотальный диктант», с помощью онлайн-сервиса К. Дружкина. Как и во всех ранее упомянутых случаях, нас интересовала не орфографическая и пунктуационная сложность диктантов, но их удобочитаемость (табл. 1).

Как видно, индексы SMOG (учитывает долю многосложных слов) и Флеша — Кинкейда (см. выше) хорошо согласуются друг с другом. Самыми сложными оказались тексты Р. Киплинга, З. Прилепина, Д. Рубиной и Д. Быкова, самыми простыми — тексты В. Быкова и Л. Толстого. Общий индекс сервиса строится на нескольких десятках критериев и рассчитывается как нормированное значение первой компоненты в анализе главных компонент (PCA); он примерно соответствует классу средней общеобразовательной школы. Согласно данному индексу, самыми сложными являются тексты Д. Рубиной и И. Соколова-Микитова, самыми простыми — тексты Л. Толстого, Е. Водолазкина, А. Усачева и В. Быкова. Как показывают правые столбцы таблицы, тексты, определенные как простые по многоуровневому индексу сервиса, имеют большую плотность личных форм глагола (а значит,

Таблица 1. Результаты отдельных метрик для текстов «Тотального диктанта», прочитанных в 2004–2016 гг.

Год	Автор	Индекс сервиса	Индекс SMOG	Индекс Флеша — Кинкейда	Существительные	Личные формы глагола	Родительный падеж	Леммы на -ние	1000 частотных слов
					%				
2016	А. Усачев, ч. 1	6,1	9,12	9,93	0,26	0,16	0,29	0,02	0,63
2015	Е. Водолакин, ч. 2	6,1	8,41	7,98	0,24	0,11	0,31	0,01	0,61
2014	А. Иванов, ч. 3	7,3	10,35	9,60	0,23	0,11	0,18	0,01	0,66
2013	Д. Рубина, ч. 2	10,1	13,35	14,73	0,34	0,08	0,40	0,03	0,61
2012	З. Прилепин	8,4	13,42	11,82	0,24	0,10	0,38	0,04	0,72
2011	Д. Быков	7,8	13,27	14,03	0,28	0,09	0,34	0,03	0,59
2010	Б. Стругацкий	7,8	11,65	10,48	0,28	0,07	0,39	0,04	0,67
2009	Н. Гоголь	8,1	8,64	7,84	0,28	0,07	0,35	0,01	0,59
2008	Р. Киплинг	8,5	16,19	15,93	0,32	0,07	0,37	0,03	0,60
2007	В. Быков	6,2	7,36	7,12	0,27	0,13	0,38	0,01	0,58
2006	И. Соколов-Микитов	10,1	8,54	11,19	0,36	0,12	0,39	0,03	0,46
2005	А. Бек	7,7	8,89	8,34	0,30	0,10	0,35	0,01	0,61
2004	Л. Толстой	4,8	8,32	5,59	0,23	0,12	0,26	0,00	0,70

меньше причастных, деепричастных и инфинитивных конструкций), меньше существительных в родительном падеже, меньше лемм на *-ние*, а также малое лексическое разнообразие (слова из списка топ-1000 самых частотных покрывают от 58 до 70% текста).

Заключение

Завершая статью, мы хотели бы обратить внимание на две тенденции в анализе удобочитаемости, в том числе с помощью онлайн-сервисов. Во-первых, это привлечение к анализу не только низкоуровневых единиц (длина слова, длина предложения), но и единиц более высокого уровня (например, морфологических и синтаксических характеристик). Во-вторых, видна тенденция к представлению отдельных критериев — в дополнение к единому индексу удобочитаемости. Последнее особенно полезно в случае, если онлайн-сервис используется самим автором текста для того, чтобы проанализировать текст и сделать его более доступным для чтения, улучшая отдельные параметры.

По мере настройки индексов удобочитаемости на разных коллекциях текстов и с помощью различных классификаторов становится ясно, что вес тех или иных критериев нельзя определить раз и навсегда (ср. коэффициенты в формуле Флеша — Кинкейда) и что предлагаемые модели пока слишком зависят от обучающей коллекции, состава текстов, схемы определения экспертных оценок и случайных «шумных» данных в самих текстах.

Дальнейшее направление исследований видится в том, чтобы понять, как модели оценки удобочитаемости зависят от жанрового состава текстов. Кроме того, требуется исследовать более подробно, какие лингвистические факторы важны для определения отдельных уровней текста и для отдельных категорий читателей (например, студентов РКИ, достигших уровня B2, и т.п.).

Можно ставить и обратную задачу: с помощью экспериментальных исследований пытаться определить, что говорят оценки сложности текста о характеристиках самого эксперта (носителя языка или изучающего язык).

Очень важный прикладной аспект оценки удобочитаемости — потенциальное массовое применение индексов удобочитаемости для экспертизы лингвистических корпусов. Как известно, тексты НКРЯ и других общих корпусов в целом сложно использовать в обучении РКИ, особенно на младших ступенях (A1–B2). Сплошная оценка удобочитаемости текстов, с настройкой на РКИ, помогла бы выделить кластеры текстов, пригодных для использования в обучении. Наконец, полезно было бы вычислить индекс удобочи-

таемости всего корпуса как некоторый показатель, по которому этот корпус может быть противопоставлен другим корпусам.

Источники

Гращенко Л.А., Мизернов И.Ю. Анализ методов оценки сложности текста // Новые инф. технологии в автоматизир. системах. 2015. № 18. С. 572–581.

Дружкин К.Ю. Метрики удобочитаемости для русского языка [Электронный ресурс]: ВКР / НИУ ВШЭ. М., 2016. URL: <https://www.hse.ru/edu/vkr/184791276>.

Дружкин К.Ю. Readability: онлайн-сервис [Электронный ресурс]. 2015. URL: <http://web-corpora.net/wsgi3/readability/index>.

Кошелева Д. Определение уровня языковой сложности текстов для изучающих польский язык как иностранный: диплом. работа. М.: НИУ ВШЭ, 2015.

Микс Я.А. Методика разработки формул читабельности // Сов. педагогика и школа. 1974. № 9. С. 78–163.

Невдах М.М. Разработка методики автоматизированной проверки трудности восприятия учебного текста // Методика автоматизации отладки и тестирования реализации алгоритмов обработки изображений. 2010. С. 20–21.

Оборнева И.В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров: дис. ... канд. пед. наук. М., 2006.

Петрова И.Ю., Окладникова С.В. Методика расчета базовых показателей читабельности тестовых материалов на основе экспертных оценок // Прикасп. журн.: упр. и высокие технологии. 2009. № 3 (7). С. 85–95.

Розушина Ю.В. Использование критериев оценки удобочитаемости текста для поиска информации, соответствующей реальным потребностям пользователя // Проблемы програмування. 2007. № 3. С. 76–87.

Солнышкина М.И., Кисельников А.С. Параметры сложности экзаменационных текстов // Вестн. Волгогр. гос. ун-та. Сер. 2: Языкознание. 2015. № 1. С. 99–108.

Ткачук Е.А., Кучма В.Р. Трудность современных учебников: мифы и реальность, способы оценки // Вопр. школ. и унив. медицины и здоровья. 2014. № 4. С. 15–22.

Тотальный диктант [Электронный ресурс]. URL: <http://totaldict.ru/dictants>.

Филиппова А.В. Управление качеством учебных материалов на основе анализа трудности понимания учебных текстов: автореф. дис. ... канд. тех наук. Уфа, 2010.

Шпаковский Ю.Ф. Оценка трудности восприятия текста // Тр. Белорус. гос. технолог. ун-та. Сер. 9: Изд. дело и полиграфия. 2012. Вып. XX. С. 72–75.

Chall J.S., Dale E. Readability Revisted: The New Dale-Chall Readability Formula. Cambridge, MA: Brookline Books, 1995.

Coleman M., Liau T.L. A Computer Readability Formula Designed for Machine Scoring // J. of Applied Psychology. 1975. Vol. 60. No. 2. P. 283–284.

Karpov N., Baranova J., Vitugin F. Single-Sentence Readability Prediction in Russian // Proc. of the 3rd Intern. Conf. “Analysis of Images, Social Networks, a. Texts” (AIST) / D.I. Ignatov, M.Y. Khachay, A. Panchenko, N. Konstantinova, R.E. Yavorsky (eds). Springer, 2014. P. 91–100. (Communications in Computer a. Inform. Sci.; vol. 436).

Kincaid J.P., Fishburne R.P.Jr., Rogers R.L., Chissom B.S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel: Research Branch Rep. 8-75 / Naval Technical Training Command. Millington, TN: Inst. of Simulation a. Training, Univ. of Central Florida, 1975.

McLaughlin H. SMOG Grading — a New Readability Formula // J. of Reading. 1969. Vol. 12. No. 8. P. 639–646.

Readability.io: проект Ивана Бегтина [Электронный ресурс]. URL: <https://plainrussian.ru/>.

Reynolds R. Russian Natural Language Processing for Computer-Assisted Language Learning: PhD diss. UiT: The Arctic Univ. of Norway, 2016.

Tanaka-Ishii K., Tezuka K., Terada H. Sorting Texts by Readability // Computational Linguistics. 2010. Vol. 36. No. 2. P. 203–227.

Научное издание

**XVII Апрельская международная научная конференция
по проблемам развития экономики и общества**

В четырех книгах

Книга 4

Зав. редакцией *Е.А. Бережнова*

Редактор *О.А. Каменкова*

Художественный редактор *А.М. Павлов*

Компьютерная верстка и графика: *Л.В. Маликина*

Корректор *О.А. Каменкова*

Подписано в печать 24.03.2017. Формат 60×88 1/16

Печать офсетная. Гарнитура Newton

Усл. печ. л. 32,7. Уч.-изд. л. 29,2. Тираж 100 экз. Изд. № 2104

Национальный исследовательский университет «Высшая школа экономики»

101000, Москва, ул. Мясницкая, 20

Тел.: (495) 772-95-90, доб. 15285