

8. Бондарко А.В. Теория функциональной грамматики: Введение. Аспектуальность. Временная локализованность. Таксис. – Л.: Наука, ЛО. 1987. – 348с.
9. Шелякин, М.А. Категория вида и способы действия русского глагола: (Теоретические основы). – Таллин: Валгус, 1983. – 216 с.

УДК 81'33 Прикладная лингвистика

К ВОПРОСУ О МЕТОДИКЕ ОЦЕНКИ ЭФФЕКТИВНОСТИ АВТОМАТИЧЕСКОЙ РУБРИКАЦИИ ТЕКСТОВ: ПСИХОЛИНГВИСТИЧЕСКИЙ АСПЕКТ

Алексей Добров
(Санкт-Петербург, Россия)

У статті розглянуто деякі психолінгвістичні фактори, що обмежують точність прийнятої сьогодні методики оцінювання ефективності автоматичної рубрикації текстів. Запропоновано деякі вдосконалення цієї методики, спрямовані на приведення її до відповідності з правилами організації психолінгвістичного експерименту й подолання цих факторів.

Ключові слова: автоматична рубрикація текстів, психолінгвістичний експеримент, проекція тексту, польова структура, шкалювання.

В данной статье рассматриваются некоторые психолингвистические факторы, ограничивающие точность принятой на сегодняшний день методики оценки эффективности автоматической рубрикации текстов. Предлагаются некоторые усовершенствования этой методики, направленные на приведение ее в соответствие с правилами организации психолингвистического эксперимента и преодоление этих факторов.

Ключевые слова: автоматическая рубрикация текстов, психолингвистический эксперимент, проекция текста, полевая структура, шкалирование.

In this article some psycholinguistic factors are discussed, that limit accuracy of the accepted today technique to evaluate efficiency of the automatic text classification. Some improvements of this technique are proposed, intended to harmonize it with the rules of organizing psycholinguistic experiment, and to overcome these factors.

Key words: automatic text classification, psycholinguistic experiment, text projection, field structure, scaling

Постановка проблемы. Оценка эффективности автоматической рубрикации набора документов производится путём сравнения результатов автоматической рубрикации с эталонной («правильной») рубрикацией. В качестве эталона используется коллекция документов, отрубрицированных независимыми аудиторами. Проблема состоит в том, что выбор рубрик, релевантных документу, представляется сложной проблемой не только для автоматизированных систем, но и для человека. Как правило, аудиторы указывают только те рубрики, которые

характеризуют основное содержание документа, хотя документ может быть отнесен и к другим рубрикам. В результате, по данным исследования М.С.Агеева, Б.В.Доброва и Н.В.Лукашевич, «при сравнении результатов рубрикации разными экспертами одних и тех же документов процент совпадения поставленных рубрик может оказаться весьма низким – 60%» [1: 29].

Аудиторы могут не учитывать некоторые рубрики, релевантные документам, из-за сложности ориентации в больших классификаторах. Кроме того, аудиторы, компетентные в той или иной предметной области, могут испытывать трудности при рубрикации документов, относящихся к другим предметным областям. В таких случаях аудиторы могут указывать слишком широкие рубрики, некорректные рубрики или вовсе не указывать рубрик. Принятие решения об отнесении документа к той или иной рубрике может быть также затруднено отсутствием формальных критериев, позволяющих определить степень важности рубрики в документе.

Таким образом, существует ряд психолингвистических факторов, ограничивающих точность экспериментальной оценки эффективности автоматической рубрикации текстов, основанной на сравнении её результатов с результатами ручной рубрикации, выполненной аудиторами. Тем не менее, до сих пор не предпринималось попыток изучения этой методики с точки зрения психолингвистики. Кроме того, в силу трудоёмкости ручной рубрикации традиционно считается, что для проведения оценки для каждого текста достаточно наблюдать рубрики, указанные одним аудитором. В данной статье осуществляется попытка показать некорректность такого подхода.

Актуальность исследования. В последние годы возрос интерес к проблеме оценки эффективности автоматической рубрикации текстов. Непрерывный рост объёмов информации в средствах массовой коммуникации и, в частности, в сети Интернет, приводит, с одной стороны, к постоянному повышению трудоёмкости этой деятельности, а, с другой стороны, к нарастающей необходимости в совершенствовании качества и скорости поиска информации и организации ее хранения. Для решения этих задач создается компьютерный инструментарий, позволяющий в той или иной мере автоматизировать и, тем самым, упростить процесс рубрикации текстов, но, по данным ряда исследований (М.С.Агеева, Б.В.Доброва, Н.В.Лукашевич и др.), уровень эффективности существующих на сегодняшний день систем автоматической рубрикации текстов недостаточно высок для того, чтобы упростить деятельность экспертов. Поскольку, как отмечалось выше, точность оценки эффективности автоматической рубрикации текстов ограничена рядом психолингвистических факторов, особую актуальность приобретает вопрос об усовершенствовании экспериментальных методов оценки эффективности автоматической рубрикации с учётом этих факторов.

Анализ последних исследований и публикаций. Для оценки эффективности систем автоматической рубрикации текстов принято использовать метрики, сходные с теми, которые используются для оценки эффективности работы информационно-поисковых систем (ИПС).

Меры точности и полноты были введены и описаны в 1955 году А.Кентом и его коллегами, разработавшими «систему оценки» ИПС, включающую в себя методы статистической выборки для оценки числа не найденных релевантных документов [6].

Точность автоматической рубрикации – это нормированная мера, определяющая для одного текста отношение количества корректно привязанных к нему рубрик к общему количеству рубрик, объективно релевантных данному тексту, или количество истинных положительных значений, деленное на общее количество элементов, отнесенных к положительному классу. Полнота автоматической рубрикации – это нормированная мера, определяющая для одного текста отношение количества корректно привязанных к нему рубрик к общему количеству рубрик, объективно релевантных данному тексту, или, в общем случае, количество истинных положительных значений, деленное на общее количество элементов положительного класса. Точность и полнота могут быть измерены для реакции одного текста, причем количество рубрик должно быть большим нуля. Формулы для определения данных параметров таковы:

$$Precision = \frac{|D_{rel} \cap D_{retr}|}{|D_{retr}|}, Recall = \frac{|D_{rel} \cap D_{retr}|}{|D_{rel}|}$$

где Precision – точность, Recall – полнота, D_{rel} – множество рубрик, релевантных тексту, D_{retr} – множество выданных рубрик.

Наиболее известной формулой для оценки эффективности автоматической рубрикации является формула Ван Рисбергена (также известная как F-мера или F_1 -мера):

$$F_1 = \frac{2 * P * R}{P + R}, \text{ где } P - \text{ мера точности, а } R - \text{ мера полноты.}$$

Проблему при определении параметров точности и полноты составляет множество D_{rel} : предполагается, что во всей совокупности рубрик можно выделить «строго» релевантные и «строго» нерелевантные анализируемому документу. Тем не менее, как отмечалось в [Добров 2010], релевантность рубрики документу – это мера, зависящая от множества параметров, каждый из которых может иметь разный вес. При составлении «эталонной» рубрикации аудитор сталкивается именно с этой проблемой.

В книге [4: 74-81] предпринято разграничение лингвистического и психолингвистического экспериментов. «Граница между психолингвистическим и лингвистическим экспериментом определяется тем, какая именно модель верифицируется в данном эксперименте. Если это – языковая модель (модель языкового стандарта), то мы имеем дело с лингвистическим экспериментом (но он может служить и для верификации функционально-речевых моделей). Если же мы верифицируем модель языковой способности или модель речевой деятельности, то в этом случае эксперимент является психолингвистическим» [4: 81]. В соответствии с этим утверждением, эксперимент, направленный на верификацию автоматической рубрикации текстов, то есть некоторой модели, относящейся не к языковым стандартам, а к речевой деятельности, должен быть именно психолингвистическим и быть организованным в соответствии с правилами проведения психолингвистического эксперимента.

Как отмечает А.А. Залевская, эксперимент – это лишь средство проверки той или иной рабочей гипотезы, которая формулируется на основе хорошо обоснованной теории. В строгом соответствии с этой теорией и поставленными задачами производится выбор исследовательских процедур (в число которых наряду с экспериментом могут входить и другие средства научных изысканий),

определяется необходимый исходный материал, отрабатывается инструкция для испытуемых или формулировки заданий, выясняется необходимое и достаточное количество участников эксперимента, требования к их образовательному уровню, половозрастным характеристикам и т.д., а также вырабатываются критерии обработки полученных материалов, их анализа и интерпретации с позиций принятой исходной теории» [3: 40].

Рубрикация относится к области смыслового восприятия текста, исследованной с различных точек зрения во множестве работ, относящихся к психолингвистике, лингвистике текста, математической лингвистике. Обзор работ в этой области мог бы составить отдельную монографию. С психолингвистической точки зрения, по всей видимости, рубрикация относится к проекции текста у реципиента по А.А.Залевской (см. [2]), входящей в систему из пяти составляющих: «автор – авторская проекция текста – тело текста – реципиент – проекция текста у реципиента». В работе [3] А.А.Залевская соотносит понятие проекции текста с образом содержания текста у А.А.Леонтьева [4: 142], трактующего понимание текста как процесс перевода смысла текста в любую другую форму его закрепления [4: 141]. К числу таких форм закрепления относится в частности смысловая компрессия – реферирование и рубрикация текста.

В рамках диссертационного исследования Н.В.Рафиковой [5] был проведен эксперимент, показавший полевою структуру проекции текста: в составе смыслового поля текста выделяется ядро, периферия и маргинальные элементы, которые у различных реципиентов могут входить как в ядро, так и в периферию. В применении к рубрикации это означает, что для каждого текста существуют ядерные и периферийные рубрики, а также множество маргинальных рубрик, которые различные аудиторы могут указывать по-разному.

Цель статьи состоит в изложении некоторых усовершенствований экспериментальной методики оценки эффективности автоматической рубрикации текстов путём сравнения её результатов с эталонной рубрикацией, направленных на преодоление тех психолингвистических факторов, которые ограничивают точность этой оценки.

Изложение основного материала. Работа одного эксперта вряд ли может считаться эталоном при оценке качества работы системы автоматической рубрикации. Как отмечалось в [Добров 2010], релевантность той или иной рубрики документу представляет собой шкалируемую величину, поэтому эталонная рубрикация должна содержать в себе оценку этой величины для каждого отнесения документа к рубрике. Как было показано выше, среди рубрик текста выделяются ядерные, периферийные и маргинальные, различающиеся у разных аудиторов. Поэтому экспериментальной методикой, наиболее целесообразной для выявления релевантности рубрики документу, представляется метод шкалирования. Тем не менее, до сих пор не предпринималось попыток применять этот метод при составлении эталонной рубрикации, более того, в качестве эталона, как правило, используется совокупность рубрик, указанная одним экспертом.

Эталонная рубрикация одной и той же коллекции документов может быть произведена несколькими экспертами. При этом для каждого отнесения документа к рубрике можно произвести расчет математического ожидания и дисперсии указанных экспертами значений релевантности, что позволяет, в

соответствии с Т-критерием Стьюдента, оценить статистическую значимость различий между результатами работы САРТ и эталонной рубрикации. При этом должно применяться сравнение выборочного среднего с заданным значением. Релевантными выданными рубриками можно считать те рубрики, для которых Т-критерий не показывает статистически значимых различий между выборочным средним значением релевантности, установленным экспертами, и значением релевантности, вычисленным САРТ. Такое усовершенствование методики оценки эффективности автоматической рубрикации позволяет повысить точность этой оценки: если при применении традиционной методики в качестве эталона используется субъективное выделение релевантных рубрик одним экспертом, то при применении предлагаемого подхода оценивается статистическая значимость различий между проверяемой расстановкой рубрик и мнениями нескольких независимых аудиторов. При этом, чем в большей мере мнения аудиторов расходятся, тем меньше вероятность выявления таких статистических различий, что как представляется, в полной мере отвечает разделению рубрик на ядерные, периферийные и маргинальные.

Для применения вышеизложенного подхода необходимо, чтобы одни и те же тексты рубрицировались одновременно разными аудиторами, специалистами в различных предметных областях, желательно, разных возрастных групп и разного пола.

Выводы. Точность традиционной экспериментальной оценки эффективности автоматической рубрикации текстов ограничивается рядом психолингвистических факторов. Эталонная рубрикация, производимая независимыми аудиторами, должна проводиться в соответствии с правилами психолингвистического эксперимента. Полевая структура рубрик, к которым может относиться текст, включает в себя наряду с ядерными и периферийными маргинальные рубрики, выделяемые по-разному различными аудиторами. Поэтому каждый текст должен рубрицироваться разными аудиторами, специалистами в различных предметных областях, желательно, разных возрастных групп и разного пола. Степень близости рубрики к ядру поля шкалируется, поэтому для рубрикации текста недостаточно только указания списка рубрик, необходимо также указание степени их значимости для текста. Сравнение результатов шкалирования рубрик различными аудиторами с величиной, полученной в результате автоматической рубрикации, может производиться в соответствии с Т-критерием Стьюдента путем сравнения выборочного среднего с заданным значением.

Перспективы дальнейших исследований. Предлагаемую усовершенствованную методику планируется применить при оценке эффективности работы разрабатываемой автором системы автоматической рубрикации новостных сообщений, основанной на комплексном лингвистическом анализе текста. В дальнейшем планируется установить необходимое количество испытуемых, перечень предметных областей, в которых должны быть компетентны аудиторы, их социокультурный статус и структуру необходимых возрастных групп. Планируется также произвести анализ возможных зависимостей между перечисленными параметрами и мерами точности и полноты рубрикации, производимой аудиторами.

ЛІТЕРАТУРА

1. Агеев М.С. Автоматическая рубрикация текстов: методы и проблемы / М.С.Агеев, Б.В.Добров, Н.В.Лукашевич // Уч. Зап. Казанского Государственного Университета. – Т. 150. – Кн. 4., 2008.
2. Залевская А.А. Метафора и формирование проекций текста / А.А.Залевская // Текст в коммуникации: [Сб. науч. тр.] / Ин-т языкознания; Тверской сельхоз. ин-т. – М., 1991.
3. Залевская А.А. Введение в психолінгвістику: [Учебник] / А.А.Залевская. – М.: Российский гос. гуманит. ун-т, 1999. – 382 с.
4. Леонтьев А.А. Основы психолінгвістики / А.А.Леонтьев. – М.: Смысл, 1997.
5. Рафикова Н.В. Динамика ядра и периферии семантического поля текста: Автореф. дис. ... канд. филол. наук / Н.В. Рафикова. – Тверь, 1994.
6. Kent A., Berry Madeline M., Luehrs Jr. Fr. U., Perry J.W. Machine literature searching VIII. Operational criteria for designing information retrieval systems // American Documentation, Vol 6, Issue 2, pp 93-101. – April 1955.

УДК 811.111(075.8)

**ПСИХОЛІНГВІСТИКА – ПЕРЕКЛАДОЗНАВСТВО:
ПОЛІТИКА ПАРТНЕРСТВА**

**Сергій Засекін
(Луцьк, Україна)**

У статті розглянуто становлення й розвиток міждисциплінарних контактів психолінгвістики й перекладознавства. Психолінгвістика як порівняно молода галузь науки спромоглася намітити орієнтири новітніх досліджень у перекладознавстві.

Ключові слова: психолінгвістика, переклад, перекладознавство.

Статья предлагает рассмотрение становления и развития междисциплинарных контактов психолінгвістики и переводоведения. Психолінгвістике, как сравнительно молодой сфере науки, удалось наметить ориентиры новых исследований в переводоведении.

Ключевые слова: психолінгвістика, перевод, переводоведение.

The article focuses on emergence and development of interdisciplinary contacts between Psycholinguistics and Translation Studies. The former as a comparatively young field of science has been able to draw new directions in Translation Studies' research.

Key words: Psycholinguistics, translation, Translation Studies.

Постановка проблеми й актуальність дослідження. Потреба в перекладі викликана розмаїттям мов у світі й міжкультурною комунікацією, а вона, як слушно зауважує Т. Слама-Казак, – у фокусі інтересу психолінгвістики [24]. Підхід цієї науки до перекладу як виду комунікації тлумачить перекладача як “сеmiotичну особистість” [9], покликану долати труднощі незбігу значень, різниці “мовних картин світу” та іменованих об’єктів, знаходити еквіваленти