

## **Тришин В.Н. Две рецензии на статью «Эволюция автономного словаря ASIS® к интерактивному общедоступному веб-словарю» и комментарий к ним**

Наша совместная с Ю.В. Куяновым статья «Эволюция автономного словаря ASIS® к интерактивному общедоступному веб-словарю» [1] была направлена на XVIII Международную конференцию: «Аналитика и управление данными в областях с интенсивным использованием данных» (DAMDID/RCDL'2016, Ершово, Москва, 11–14 октября 2016 года), см. <http://damdid2016.frccsc.ru/>.

В ответ были получены две анонимные рецензии (опечатки не исправлены):

### **Рецензия №1**

Статья описывает историю, состав и технологию большого онлайн-словаря синонимов. Статья не содержит обзора литературы (ссылки есть только на работы авторов) и сравнения с аналогами, что затрудняет объективную оценку предложенного продукта и его преимуществ. Некоторые принципы организации словаря (смещение синонимических и родо-видовых связей, отсутствие разделения значений многозначных слов и омонимов), закрытость проекта, а также невозможность интеграции словарной базы с внешним ПО выглядят как существенные недостатки продукта. Возможно, работу можно представить в демо-секции, если такая будет на конференции.

### **Рецензия №2**

Дан исторический экскурс в создание и развитие словаря синонимов. Реализация доступа к словарю на сайте потребовала трансформацию его из FoxPro в MySQL+PHP. В виду отсутствия поддержки каких-либо видов связей, кроме синонимии, другие виды связей используются ошибочно. Обобщающие термины преподносятся как синонимы. Разобраться в омонимах невозможно, так как синонимы слов, имеющих не одно значение, слиты в кучу. Преимущество словаря видно в присутствии множества разговорных вариантов употребления слов. Однако легко нашлись распространённые разговорные варианты слов, отсутствующие в словаре.

Статья вряд ли относится к тематике конференции. Преобразования, проводимые при переложении на другую программную платформу, элементарны. Возможно, словарь полезен для задач анализа данных, текстовой аналитики, но вряд ли в том виде, в котором он реализован. Какие-либо рассуждения на эту тему, кроме единственного упоминания поисковых систем, тоже отсутствует. Нет анализа подобных работ, в том числе, в списке литературы.

### **Замечание Тришина В.Н.**

Я должен сказать, что на все замечания рецензентов имеются прямые ответы в текстах моих трёх статей о словаре (рецензируемая статья является заключительной из этой серии, предыдущие две статьи приведены в списке литературы к рецензируемой статье, см. ниже раздел "Литература"), что вызывает, мягко сказать, недоумение.

Поэтому я ответил рецензентам прямыми цитатами из своих статей и дополнительным своим комментарием.

### **Замечания рецензента 1 к статье [1]**

1) «Статья не содержит обзора литературы (ссылки есть только на работы авторов) и сравнения с аналогами, что затрудняет объективную оценку предложенного продукта и его преимуществ».

Цитаты из статьи [2].

«В электронной версии Оксфордского словаря английского языка (над которой в настоящее время работают около 80 специалистов), см. <http://www.oed.com/>, <http://dic.academic.ru/dic.nsf/ruwiki/212818>, имеется около 600 тыс. слов и словосочетаний. Словарь включает все слова, бытующие или бытовавшие в английском литературном и разговорном языке начиная с 1150 года. **Онлайн-версию Оксфордского словаря** намечают выпустить в 2017 году.

**Имеется две онлайн-версии словаря английского языка Уэбстера**, в которых быстро выдаётся то, что мы понимаем под синонимами, синонимическими связями («**Related Words**»):

1) <http://www.merriam-webster.com/> (примерно на 600 тыс. слов и словосочетаний);

2) Разработка Принстонского университета

<http://www.webster-dictionary.org/>

(примерно на 240 тыс. слов и словосочетаний) с фонетическими транскрипциями слов, их звучанием, переводом на другие европейские языки.

Поиск слов в обеих онлайн-словарях осуществляется **только по началу слова**.

В нашей онлайн-версии словаря можно искать по началу, фрагменту, окончанию слова (словосочетания).

«Ссылки на синонимы в словаре начиная с версии 6.1 закодированы в качестве защиты от недобросовестных пользователей, а таких в Интернете немало, разместивших его на своих сайтах в своих программных оболочках без указания автора. Например, если поискать в Google ошибочное словосочетание «без вознаграждения» (вместо «без вознаграждения») — будет 576 ссылок (на дату 31.08.2015). Эта ошибка была допущена в размещенной на нашем сайте до 2010 года незашифрованной версии словаря. Не случайно ни количество слов, ни количество синонимов в этих словарях не указано. Длины записи в этих словарях ограничены 50 символами, а толкования обрезаны в тех же местах, как и в старой открытой версии данного словаря-справочника. А все опечатки в старой версии словаря-справочника перешли в эти словари. Не случайно ни количество слов, ни количество синонимов в этих вторичных словарях не указаны (см., например, ссылки на пять анонимных синонимических онлайн-словаря в анонимной же статье

«СИНОНИМЫ» в Википедии <https://ru.wikipedia.org/wiki/Синонимы> )».

### *Комментарий*

Онлайн-версия словаря Тришина по своим поисковым возможностям намного превосходит зарубежные аналоги, не говоря уже об анонимных российских «аналогах», словарные базы которых заимствованы из открытой версии словаря Тришина 2010 года на 326 тыс. слов. Что же касается количественного состава словаря (587 тыс. слов, 2154 тыс. синонимических слов), то ни в одном словаре русского языка и близко эти цифры не достигаются.

Это совершенно пионерская работа, которой принципиально не может быть ни в каких обзорах. Она не имеет аналогов. При кажущейся простоте нашей реализации никто не сделал ничего подобного ни у нас, ни за границей.

2) «Некоторые принципы организации словаря (смещение синонимических и родо-видовых связей, отсутствие разделения значений многозначных слов и омонимов),

закрытость проекта, а также невозможность интеграции словарной базы с внешним ПО выглядят как существенные недостатки продукта».

### *Комментарий*

В статье [1] написано, что «степень близости по смыслу является качественной категорией, поэтому мнения о том, являются ли конкретные слова (словосочетания) синонимами или нет, у разных людей могут не совпадать. Учитывая, в первую очередь, прикладную нацеленность словаря-справочника Тришина В. Н. для компьютерной обработки текстов, мы используем более общее понятие «синонимическая связь», а именно, если в какой-либо фразе одно слово (словосочетание) заменить другим и, хотя бы один смысл фразы сохраняется, то мы говорим, что эти слова (словосочетания) синонимически связаны».

Эта расширенная трактовка синонимии используется не только нами, но она используется в вышеупомянутых онлайн-словарях английского языка (вместо слова «Synonyms» используется понятие «Related Words»). Академик Ю. Д. Апресян также считает, что «главным свойством синонимов является именно то, что они не различаются своим лексическим значением» [4].

Вышеупомянутые принципы организации словаря указаны в статье [1] (параграф 2) не столько как недостатки, а как планы на будущее. Странно, что планы на будущее автора онлайн-словаря рецензент трактует как недостатки словаря, что говорит о предвзятости рецензента.

Так ведь ни в вышеупомянутых пяти онлайн-словарях российских «коллег»-плагиаторов это также не реализовано (то, что в скобках у рецензента «смешение синонимических...»). Замечу, что таких слов в общей массе слов словаря-справочника относительно немного и работа может быть выполнена за год-два.

О какой закрытости проекта глаголет рецензент — не понятно, если открытую версию словаря версию 6.0 (без зашифрованных синонимических связей), размещенную в Интернет ещё в 2010 годом, «лексикографы» заимствовали и поместили в свои примитивные оболочки, см. [2].

В связи с вышесказанным совершенно непонятно утверждение рецензента о «невозможность интеграции словарной базы с внешним ПО». На сайте trishin.ru в «Новостях» и в Приложении №3 статьи [2] говорится о том, что мы ищем партнёров по внедрению словарной базы в различные внешние ПО, и что проблем с внедрением не может быть, так как «Словарь-справочник удобен и необходим при реализации интеллектуального поиска информации в различных корпоративных системах (**структура его словарной базы на языке Fox Pro, в которой кроме слов имеются их основы и коды одного из 19 эталонных слов для указания всех склонений или спряжений слова, проектировалась именно для таких задач**)». Да и в самой рецензируемой статье лёгкость интеграции словаря с другими ПО отмечена!!!

### **Замечания рецензента 2 к статье [1]**

1) «**В виду отсутствия (???-В.Т.)** поддержки каких-либо видов связей, кроме синонимии, другие виды связей используются ошибочно».

### *Комментарий*

Если кроме синонимии других видов связи в словаре нет, что справедливо, то как же эти «другие виды связей используются ошибочно».

Рецензент хоть вдумывается в то, что пишет?

2) «Обобщающие термины преподносятся как синонимы»

### *Комментарий*

Авторы везде в статье стараются вместо понятия «синоним» (неоднозначно трактуемом разными школами лингвистов) использовать более широкое «обобщающее» понятие «синонимическая связь», и обсуждению этого вопроса посвящен целый параграф 2.

3) «Однако легко нашлись **распространённые (???-В.Т.)** разговорные варианты слов, отсутствующие в словаре».

### *Комментарий*

Жаль, что автор не приводит примеры этих самых «распространённых разговорных», которых он «легко нашёл». Кстати, разговорные варианты имеются обязательно в письменных текстах (а создание обобщенного, объединенного словаря русского языка является одной из целей его автора).

Следует сказать, что на сайте [trishin.net](http://trishin.net) продемонстрирована **неполная словарная база** (и там ясно об этом сказано) на 526 тыс. слов и словосочетаний и 1954 тыс. синонимических связей, а в полной базе на текущий момент имеется 587 тыс. слов и словосочетаний, и 2154 тыс. синонимических связей. Так что в дополнительных 200 тыс. синонимических связях рецензент, возможно, и найдёт свои «легко найденные», но не указанные им варианты.

4) «Преобразования, проводимые при переложении на другую программную платформу, элементарны»

### *Комментарий*

Речь в статье в разделах о технической реализации словаря в Интернет идёт не столько о преобразованиях, а о нашей программной в Интернете реализации огромной словарной базы (включающей целый ряд ноу хау) и о достигнутом малом времени отклика на запросы. Если рецензент знает подобные реализованные в Интернет проекты, то пусть их укажет, а не пишет ахиною об «элементарности преобразований».

5) «Нет анализа подобных работ, в том числе, в списке литературы»

По поводу анализа «аналогов» и списка литературы см. ответ первому рецензенту.

### *Итог*

Сопредседатели конференции Л. Калиниченко и Я. Манолопулос на основании «рецензий» предложили сократить текст статьи до 2-страничного сообщения, в котором требовалось учесть замечания рецензентов. Ввиду абсурдности претензий анонимных рецензентов, авторы статьи отказались её сокращать.

Рецензии являются частью компании замалчивания словаря (попробуйте, к примеру, найти этот онлайн-словарь в Яндексе без упоминания фамилии автора!) и только подтверждают вывод о нравах в науке и недобросовестной научной конкуренции, упомянутых в самой рецензируемой статье [1].

## Литература

1. Ю. В. Куянов и В. Н. Тришин «Эволюция автономного словаря ASIS® к интерактивному общедоступному веб-словарю», 2016, см. [www.trishin.net](http://www.trishin.net).
2. Тришин В.Н. Мощь русского языка по данным синонимического словаря-справочника системы ASIS®. // Журнал «Вестник УМО. Экономика, Статистика и Информатика», № 6, 2013, стр. 7–13, <http://www.audit-it.ru/articles/soft/a119/597862.html>
3. Куянов Ю.В., Тришин В.Н. Количественный анализ Большого словаря-справочника синонимов русского языка. // Журнал «Научное обозрение: гуманитарные исследования», № 9, 2015, стр. 105–111, <http://www.audit-it.ru/articles/soft/a118/838916.html>.
4. Тришин В.Н. Две рецензии на статью «Эволюция автономного словаря ASIS® к интерактивному общедоступному веб-словарю» и комментарий к ним [www.trishin.net/doc/critics-11.pdf](http://www.trishin.net/doc/critics-11.pdf)

24 сентября 2016