



Изучение основ лексикографии с применением информационных технологий в обучении студентов-филологов

Аннотация. *Статья посвящена использованию информационных технологий в обучении студентов лингвистике и основам лексикографии. Автор описывает некоторые примеры применения программ и интернет-сервисов для создания словарей в учебных целях.*

Ключевые слова: *база данных, словарь, лингвистика, информационные технологии в обучении.*

Раздел: *(01) педагогика; история педагогики и образования; теория и методика обучения и воспитания (по предметным областям).*

Широкое проникновение информационных технологий во все области научной и образовательной деятельности находит отражение в появлении и развитии новых дисциплин, образованных на стыке гуманитарных и точных наук. Одним из таких направлений является применение информационных технологий в лингвистике. Одноименный курс входит в естественнонаучный цикл подготовки бакалавров по направлению «Лингвистика».

При изучении дисциплины «Информационные технологии в лингвистике» студентами по данному направлению предполагается формирование следующих профессиональных компетенций:

- иметь навыки работы с компьютером как средством получения, обработки и управления информацией (ПК-25);
- уметь работать с традиционными носителями информации, распределенными базами данных и знаний (ПК-26);
- обладать способностью работать с информацией в глобальных компьютерных сетях (ПК-27);
- уметь работать с электронными словарями и другими электронными ресурсами для решения лингвистических задач (ПК-28).

Кроме традиционно решаемых в курсе задач овладения стандартным программным обеспечением (офисным пакетом) заслуживает отдельного внимания последняя компетенция – умение работать со словарями для решения лингвистических задач. По нашему мнению, кроме традиционного использования словарей студент-филолог должен знать основы лексикографии – науки о создании словарей, понимать процесс проектирования словаря, правильно определять параметры словаря, анализировать структуру словарной статьи.

Все эти вопросы можно успешно решать с применением информационных технологий. Использование электронного обеспечения курса, размещенного в системе управления обучением Moodle вуза, позволяет аккумулировать все информационные ресурсы в одном месте, обеспечить студенту быстрый доступ к ним и возможность увеличить объем и разнообразие изучаемого материала.

При создании словаря нужно определить его параметры и место в классификации словарей. Разработкой принципов классификации словарей занимались



Л. В. Щерба [1], А. С. Герд [2], С. В. Гринёв [3], В. В. Морковкин [4] и многие другие учёные-лексикографы.

Словари могут быть классифицированы по целой группе параметров:

- 1) по количеству языков – одноязычный, двуязычный, многоязычный;
- 2) по форме основной единицы словаря. Это могут быть единицы меньшие, чем слово (словари корней; морфем, n-буквенных сочетаний), эквивалентные слову (словари словоформ и слов), большие, чем слово (словари словосочетаний, фраз, цитат);
- 3) по специфике отбираемой лексики – терминологические, диалектные, просторечные, словари научной и узкоспециальной терминологии и т. п.;
- 4) по способу описания единиц словаря – толковые, энциклопедические и т. п.;
- 5) по способу упорядочения материала – алфавитный, алфавитно-гнездовой, тематический (идеографический), частотный;
- 6) по учету исторического контекста – исторический, неисторический;
- 7) по назначению – учебный и т. п.

Для иллюстрирования этого вопроса удобно использовать, например, интернет-сервис Яндекс.Словари (<https://slovari.yandex.ru/>). Сервис аккумулирует в себе более 50 словарей в электронной форме, среди которых – толковый словарь Даля, энциклопедический словарь Брокгауза и Ефрона, ряд переводных англо-русских словарей Lingvo и многие другие. Использование электронных словарей позволяет студентам научиться определять параметры каждого из них и установить, например, что русско-английский экономический словарь является двуязычным, алфавитным, переводным, специальным, словарь Элочки Любоедки – одноязычным, неисторическим, алфавитным словарем литературного персонажа одного писателя, а словарь по методике обучения математике – одноязычным, энциклопедическим, алфавитным и учебным [5].

Очень полезно также познакомить студентов с визуальными словарями.

Этот вид словарей содержит не только языковые единицы – слова и словосочетания, но и иллюстрации к каждому из них. Такие словари редко встречаются в печатной форме по причине трудности издания и дороговизны, зато в электронной форме выглядят очень привлекательно и вызывают большой интерес. Так, английский визуальный словарь Merriam Webster (<http://visual.merriam-webster.com/>) содержит большое количество английских слов с картинками и является толковым, одноязычным, тематическим по принципу расположения терминов. Система поиска облегчает нахождение слов по определённой теме и способствует повышению интереса к изучению английского языка.

Работу по определению параметров словарей студенты могут проводить как на занятиях при наличии доступа в Интернет, так и самостоятельно. Все необходимые ссылки с заданиями и указаниями по их выполнению размещаются в соответствующем курсе системы Moodle вуза.

Одним из учебных заданий для студентов лингвистического направления может быть задание на создание мини-словаря какого-либо текста или совокупности текстов. В случае с обучением лингвистов по профилю «Перевод и переводоведение» можно говорить как о русскоязычных, так и об иноязычных текстах. Создание словаря даёт возможность студенту проанализировать круг используемой лексики и её особенности, получить начальные навыки исследовательской работы.

При создании любого словаря важно придерживаться основных этапов проектирования, что облегчает работу и предотвращает обнаружение ошибок и несоответствий с замыслом на более поздних этапах. Студент-филолог должен обладать знаниями теоретических основ и некоторыми практическими навыками в этой обла-



сти. Процесс разработки зависит от характеристик словаря и также довольно сложен. Обычно создание словаря разделяется на ряд определённых этапов.

На первом этапе происходит проектирование словаря. Под проектированием словаря в лексикографии понимают определение его типа, назначения и основных характеристик. Результат этого этапа – проект словаря и методика его разработки. На этой основе принимаются предварительные решения о композиции словаря, определяется состав отражаемой в нём информации.

Среди характеристик словаря важным параметром являются его назначение или цели создания. Цель создания словаря тесно связана с его остальными качествами и может быть информационной, нормативной, учебной, систематизирующей или иной.

Например, словарь диалектов должен зафиксировать существующие слова и примеры их употребления, а значит, выполняет аккумулирующую функцию, а отнюдь не нормативную, предполагающую фиксацию норм, правил употребления. Словарь по методике обучения математике [6] объединяет, упорядочивает и приводит в систему основные термины методики математики, следовательно, выполняет функцию систематизации понятийного аппарата предметной области [7]

На втором этапе осуществляется отбор лексического материала и составление словника будущего словаря. В соответствии с целью создания должны быть отобраны словарные единицы, термины, подлежащие дальнейшему включению в словарь и описанию. Отобранные словарные единицы объединяют в перечень – так называемый словник. Например, для словаря по методике обучения математике [8] вначале был составлен словник [9], содержащий термины, подлежащие описанию в словаре.

На третьем этапе (основном) проводится непосредственный анализ и описание специальной лексики, включенной в словник.

На четвертом этапе происходит подготовка словаря к изданию – редактирование, уточнение и проверка взаимности ссылок и т. д. [10] Если словарь создаётся в электронной форме, происходит отладка программных средств создания, проверка гиперссылок, связывающих словарные статьи.

При составлении словника необходимо определить, по какому принципу необходимо принимать решение о включении слова в будущий словарь и каковы должны быть источники для выбора слов.

Если речь идёт о специализированном словаре (математическом, экономическом, биологическом), то источниками должны быть тексты определённой предметной области (математики, экономики, биологии соответственно), если же о словаре общей лексики, то тексты могут быть самыми разнообразными по тематике. В этом случае количество текстов должно быть довольно объёмным, а соотношение текстов разной тематики – взвешенным.

Для упрощения обработки можно предложить студентам в учебных целях обработать один или несколько текстов для составления небольшого словника по данным источникам.

При отборе терминов для словника согласно основам лексикографии могут применяться самые различные методы – частотный, экспертный, семантический или на основе систематизации предметной области, позиционный и т. п.

Эти же методы могут применяться при обработке текстов в процессе реферирования и аннотирования. Вопрос о процессе реферирования и аннотирования текстов и о принципах работы автоматизированных систем, выполняющих такие функции, также входит в перечень вопросов для изучения в рамках курса «Информационные технологии в лингвистике». При составлении реферата или аннотации необходимо также выявлять ключевые единицы обрабатываемого текста.



Частотный метод основан на подсчете частоты встречаемости слов в тексте. Наиболее важными, ключевыми словами считаются такие знаменательные слова текста, которые с учетом всех синонимов встречаются в тексте наибольшее число раз.

Этот метод состоит в разбиении текста на отдельные слова и подсчете частоты употребления каждого из них. Полученный список слов с частотами необходимо обработать. Из этого списка исключаются предлоги, союзы, местоимения и т. п. Кроме того, исключаются также незначащие слова – те слова, которые не несут смысловой нагрузки или являются общеупотребительными, встречающимися в текстах любой тематики. Поскольку одно и то же слово может встречаться в тексте в разных формах, то следует произвести объединение таких словоформ в единое словоупотребление/слово. Этот процесс, с одной стороны, довольно кропотлив, трудно формализуем, а с другой – позволяет получить более четкую картину системы. Оставшиеся слова ранжируются по частоте встречаемости. Слова, занимающие первые 10, 20, 50 мест (в зависимости от объема анализируемого текста), являются важнейшими и должны быть включены в создаваемый словарь [11]. Такие слова наилучшим образом характеризуют смысл текста и могут быть использованы для создания аннотации и реферата.

Частотный метод обработки текста может быть реализован разными программными средствами: от программ стандартного офисного пакета MS Office до специализированных программ и сервисов.

В текстовом редакторе MS Word имеется возможность разбиения текста на отдельные словоформы с использованием средства «Найти и заменить», если в качестве разделителя слов принять знак пробела и удалить все знаки препинания. Полученный список слов можно упорядочить по алфавиту, а затем обработать в табличном процессоре Excel, объединяя разные формы одного слова в единое словоупотребление и суммируя их частоту. Исключив незначащие слова, можно получить список ключевых слов текста, упорядоченный по частоте.

Возможно применение специальных программ и онлайн-сервисов для частотной обработки текстов, однако наиболее наглядным нам представляется использование онлайн-сервиса для создания облака слов Wordle (<http://www.wordle.net/>).

Сервис позволяет создать для произвольного текста небольшого объема графическую иллюстрацию – облако слов, где величина слова соответствует частоте его встречаемости в тексте. Чем чаще встречается слово, тем крупнее оно будет изображено. Сервис не позволяет объединять словоформы в единое слово, однако очень наглядно представляет картину значимости слов текста. При более подробной работе можно определить количество отображаемых слов, цвета и шрифты. Сервис Wordle визуализирует получаемые статистические результаты, что позволяет студентам понять и освоить часто вызывающий затруднения частотный метод обработки.

Экспертный метод отбора слов состоит в использовании совместной работы коллектива экспертов в изучаемой области, широком обсуждении для решения о включении слов в словарь. Такой метод для учебной работы малопригоден.

Семантический метод, пожалуй, наиболее сложен и состоит в выделении важнейших смысловых элементов текста и установлении связей между ними. Создаваемую семантическую сеть связей между словами рекомендуется визуализировать графически с помощью специальных программных средств. При создании семантических сетей можно использовать также методы математического моделирования [12].

В качестве иллюстрации семантического метода можно познакомить учащихся с сервисом Snappywords.com. На данном сайте представлены результаты проекта, разрабатываемого в Принстонском университете (США). Сайт представляет графический



интерфейс для выполнения запросов к лексической базе данных WordNet. Snappywords также можно назвать визуальным словарем, в котором можно рассматривать каждое слово в семантической сети лексических отношений между связанными словами.

При знакомстве с разными типами словарей следует обращать внимание студентов на сходство и различие в структурах словарных статей в словарях разных типов.

В толковом словаре статья может содержать слово и его морфологические, стилистические и иные характеристики, возможно, примеры употреблений. В энциклопедическом – слово (понятие) и его описание. В переводном словаре – слово и варианты его переводов, примеры употреблений, морфологические и стилистические характеристики и т. п. В частотном словаре – слово и его частота и так далее.

Нетрудно заметить, что во всех словарных статьях присутствует заголовочная единица – слово или словосочетание, а также характеристики, набор которых зависит от типа словаря. Таким образом, словарь можно трактовать как информационную систему или базу данных, где все статьи укладываются в таблицу конкретной структуры. Набор столбцов будет зависеть от типа словаря и будет соответствовать характеристикам слова (заголовочной единицы), а каждая строка такой таблицы будет содержать данные из словарной статьи (см. таблицу).

Структура словарной статьи

Слово (заголовочная единица)	Характеристики по типам словаря			
	Переводной	Толковый	Энциклопедический	Частотный
	Перевод, морфологические и стилистические характеристики, примеры употребления	Морфологические и стилистические характеристики, значение, примеры употребления	Описание, история вопроса, ссылки на связанные статьи и источники	Абсолютная и/или относительная частота
...

Учебным заданием для студентов является анализ структуры статей в словарях разных типов, выделение характеристик и проектирование структуры подобной таблицы. В качестве примера базы данных используется также «Словарь смоленских говоров», разрабатываемый в Смоленском государственном университете. Студенты рассматривают примеры статей из данного словаря, анализируют структуру статьи, принципы заполнения электронной базы данных [13].

Для апробации электронного словаря в учебном процессе на основе системы дистанционного обучения Moodle Смоленского государственного университета создана её демонстрационная версия «Смоленские говоры». Демоверсия представляет собой учебный курс, включающий описание словаря, методические рекомендации для студентов по работе с информационной системой и саму базу данных.

Система Moodle даёт возможность создания и использования однотабличных баз данных, что облегчает для студентов понимание структуры словаря и способствует формированию профессиональных компетенций в области лексикографии и информационных технологий.

Под руководством преподавателя студенты наполняют базу данных «Смоленские говоры» словами и примерами употребления в соответствии с методическими указаниями. Введённые данные проверяются преподавателем и после утверждения становятся доступными для просмотра участниками курса [14].

Электронный словарь «Смоленские говоры», реализованный в виде базы данных, является ещё одним ярким примером применения информационных технологий



в лингвистике. Он предполагает активное взаимодействие студентов и преподавателей и служит, с одной стороны, формированию ценной с научной точки зрения лингвистической базы данных, с другой стороны, способствует развитию и интеграции межкафедральной научно-исследовательской работы.

Таким образом, рассмотренные примеры применения информационных технологий в лингвистике способствуют овладению студентами лингвистического направления основами лексикографии и формированию необходимых профессиональных компетенций.

Ссылки на источники

1. Щерба Л. В. Опыт общей теории лексикографии // Языковая система и речевая деятельность. – Л.: Наука, 1974. – С. 265.
2. Герд А. С. Основы научно-технической лексикографии (Как работать над терминологическим словарем). – Л.: Изд-во Ленингр. ун-та, 1986.
3. Гринёв С. В. Основы лексикографического описания терминосистем: автореф. дис. ... д-ра филол. наук. – М., 1990. – 43 с.
4. Морковкин В. В. Основы теории учебной лексикографии: дис. ... д-ра филол. наук. – М., 1990. – 75 с.
5. Самарина А. Е. О некоторых параметрах электронного методического словаря по обучению математике в школе // Успехи современного естествознания. – 2004. – № 6. – С. 100–102.
6. Сенькина Г. Е., Тимофеева Н. М., Самарина А. Е. и др. Словарь по теории и методике обучения математике. – Смоленск: СмолГУ, 2008.
7. Самарина А. Е. Методический словарь по обучению математике в школе как средство систематизации предметной области // Успехи современного естествознания. – 2004. – № 6. – С. 98–99.
8. Самарина А. Е. Краткий словарь-справочник по теории и методике обучения математике: алгебра. – Смоленск: СГПУ, 2005. – 108 с.
9. Самарина А. Е. Словник к методическому словарю по обучению математике. – Смоленск: СГПУ, 2003.
10. Гринёв С. В. Указ. соч.
11. Самарина А. Е. Применение частотного анализа текстов в процессе конструирования методического словаря по математике // Системы компьютерной математики и их приложения: сб. ст. – Смоленск: СГПУ, 2005. – С. 214–220.
12. Киселева О. М., Тимофеева Н. М., Быков А. А. Формализация элементов образовательного процесса на основе математических методов // Современные проблемы науки и образования. – 2013. – № 1. – С. 224.
13. Самарина А. Е., Лунькова Е. С. Электронная база данных «Словарь смоленских говоров» и её применение в обучении студентов филологических специальностей // Инфокоммуникационные технологии в региональном развитии: сб. тр. Межрегион. науч.-практ. конф. – Смоленск, 2013.
14. Самарина А. Е. База данных в системе Moodle и возможности её применения в обучении // Концепт. – 2013. – № 08 (август). – ART 13163. – URL: <http://e-koncept.ru/2013/13163.htm>. – Гос. рег. Эл № ФС 77-49965. – ISSN 2304-120X.

Anna Samarina,

Candidate of Pedagogic Sciences, Associate Professor at the chair of Information Technologies, Smolensk State University, Smolensk

a.e.samarina@gmail.com

Learning basics of grammar with the use of information technology in training of students-philologists

Abstract. The paper is devoted to the use of information technologies in teaching linguistic and basics of lexicography for students of philological specialties. The author describes some examples of applying software and internet-services for creation of dictionaries for studying purposes.

Key words: database, dictionary, linguistics, information technologies in education.

References

1. Shherba, L. V. (1974) "Opyt obshhej teorii leksikografii", *Jazykovaja sistema i rechevaja dejatel'nost'*, Nauka, Leningrad, p. 265 (in Russian).
2. Gerd, A. S. (1986) *Osnovy nauchno-tehnicheskoy leksikografii (Kak rabotat' nad terminologicheskim slovarem)*, Izd-vo Leningr. un-ta, Leningrad (in Russian).

ISSN 2304-120X



9 772304 120159

03



3. Grin'ov, S. V. (1990) *Osnovy leksikograficheskogo opisanija terminosi-stem: avtoref. dis. ... d-ra filol. nauk*, Moscow, 43 p. (in Russian).
4. Morkovkin, V. V. (1990) *Osnovy teorii uchebnoj leksikografii: dis. ... d-ra fi-lol. nauk*, Moscow, 75 p. (in Russian).
5. Samarina, A. E. (2004) "O nekotoryh parametrah jelektronnogo metodicheskogo slo-varja po obucheniju matematike v shkole", *Uspehi sovremennogo estestvoznanija*, № 6, pp. 100–102 (in Russian).
6. Sen'kina, G. E., Timofeeva, N. M. & Samarina, A. E. et al. (2008) *Slovar' po teorii i metodike obuchenija matematike*, SmolGU, Smolensk (in Russian).
7. Samarina, A. E. (2004) "Metodicheskij slovar' po obucheniju matematike v shkole kak sredstvo sistema-tizacii predmetnoj oblasti", *Uspehi sovremennogo estestvoznanija*, № 6, pp. 98–99 (in Russian).
8. Samarina, A. E. (2005) *Kratkij slovar'-spravochnik po teorii i metodike obuchenija matematike: algebra*, SGPU, Smolensk, 108 p. (in Russian).
9. Samarina, A. E. (2003) *Slovník k metodicheskomu slovarju po obucheniju matematike*, SGPU, Smolensk (in Russian).
10. Grin'ov, S. V. (1990) Op. cit.
11. Samarina, A. E. (2005) "Primenenie chastotnogo analiza tekstov v processe kon-struirovaniya metodicheskogo slovarja po matematike", *Sistemy komp'juternoj matematiki i ih prilozhenija: sb. st.*, SGPU, Smolensk, pp. 214–220 (in Russian).
12. Kiseleva, O. M., Timofeeva, N. M. & Bykov, A. A. (2013) "Formalizacija jelementov obrazovatel'nogo processa na osnove matematicheskikh metodov", *Sovremennye problemy nauki i obrazovanija*, 2013, № 1, p. 224 (in Russian).
13. Samarina, A. E. & Lun'kova, E. S. (2013) "Электронная база данных 'Slovar' smolenskikh govorov' i ejo primenenie v obuchenii studentov filologicheskikh special'nostej", *Infokommunikacionnye tehnologii v regional'nom razvitii: sb. tr. Mezhreg. nauch.-prakt. konf.*, Smolensk (in Russian).
15. Samarina, A. E. (2013) "Baza dannyh v sisteme Moodle i vozmozhnosti ejo primeneniya v obuchenii", *Koncept*, № 08 (avgust), ART 13163. Available at: <http://e-koncept.ru/2013/13163.htm>, Gos. reg. Jel № FS 77-49965, ISSN 2304-120X (in Russian).

Рекомендовано к публикации:

Горевым П. М., кандидатом педагогических наук, главным редактором журнала «Концепт»