

**Опыт составления учебного глоссария для
компьютерной обучающей программы**
(на примере темы «ОРГАНИЗМ»)

Иван Зарянов

Университет г. Тампере

Институт современных языков и
переводоведения

Кафедра перевода русского языка

Дипломная работа на соискание учёной степени
магистра философии

Сентябрь 2009

Tampereen yliopisto

Kieli- ja käännöstieteiden laitos

Käännöstiede (Venäjä)

JUHANI, ZARIANOV:

Опыт составления учебного глоссария для компьютерной обучающей программы (на примере темы «ОРГАНИЗМ») (Verkkopohjaisen opetusohjelman sanaston laatiminen, esimerkkitapauksena aihealue ”Elimistö”)

Pro gradu – tutkielma s.(sis. suomenkielinen lyhennelmä s.), liitteet s.

Syyskuu 2009

Nykyään tietotekniikkaa käytetään monella ammattialalla, myös apuvälineenä opetuksessa. Tässä työssä on täydennetty Tampereen yliopistossa kehitetty kielenopetusohjelmaa, SysMLL:ää, varten koostettua elimistö-aiheeseen liittyvää sanastoa. Sanaston pohjalta ohjelmaan luotiin sanakirja-artikkeleita, joita on testattu verkko-ohjelman harjoitusten avulla.

Tutkimuksen tavoitteena on ollut laatia elimistö-aiheinen sanasto, jota voidaan tarpeen mukaan käyttää itseopiskelussa Tampereen yliopiston kieli- ja käännöstieteiden laitoksella.

Pro gradu -tutkielmassa perehdytään opetussovelluksen sanaston laatimisen eri vaiheisiin: karkea sanajako, taajuuksien määrittäminen ja sanaston muokkaaminen sen pohjalta sekä lopullisen sanaston täydentäminen.

Tutkielmassa tarkastellaan myös perinteisten painettujen sanakirjojen ja verkkosanakirjojen eroja. Empiirisen tutkimusaineiston avulla voidaan havaita, millaisia sanoja on lisättävä sanastoon tai poistettava siitä.

Tutkimuksen tuloksena tuotettiin teeman mukainen sanasto, joka vietiin SysMLL-ohjelman tietokantaan. Lopuksi sanasto testattiin ja siihen tehtiin tarvittavat korjaukset. Sanastoa on siis pystytty kehittämään, niin että se soveltuu tarkoitettuun käyttöön. Tutkimuksen pohjalta on helppo luoda ja kehittää muitakin aihealueita SysMLL-verkko-ohjelmassa.

Ключевые слова: *ассоциативная карта, дефиниция, лексикография, словарная статья, словник, терминология, учебный глоссарий, частотный словарь*

Keywords: ajatuskartta, leksikografia, määritelmä, sanakirja-artikkeli, sanasto, sanastotyö, taajuussanakirja, termityö, verkkosovellus

Содержание

Введение	1
1 Типология словарей.....	3
1.1 Активный и пассивный словари.....	4
1.2 Двухязычный словарь.....	4
1.3 Частотные словари.....	7
1.4 Словник	8
1.5 Оценка качества составления словника	9
2 Электронная лексикография	10
2.1 Электронные словари.....	11
2.2 Электронные словари лексической сочетаемости	12
2.3 Корпусы текстов.....	13
2.4 Компьютерное обучение иностранным языкам.....	17
2.5 Обучающие программы и тесты.....	18
2.6 Лексические тренажеры.....	19
2.7 Оценка качества функционирования компьютерных программ по обучению иностранному языку	20
3 Терминоведение.....	22
3.1 Ассоциативная карта или логико-понятийная схема.....	23
3.2 Виды дефиниций.....	24
3.3 Языковые примеры употребления термина в словарной статье	29
4 SysMLL–лексический тренажёр нового поколения	31
4.1 Лексикографические аспекты.....	33
4.2 SysMLL и двухязычный словарь – сходства и различия.....	34
4.3 Словник SysMLL.....	36
4.4 Терминологические аспекты в работе по составлению глоссария	37
5 Глоссарий по теме ”Организм” в обучающей программе SysMLL.....	39
5.1 Общая характеристика глоссария.....	41
5.2 Уровни сложности глоссария	43
5.3 Толкование	45
5.4 Структура статьи.....	46
5.5 Ввод данных	47
5.6 Примеры употребления.....	48
5.6 Упражнения в SysMLL	49

5.7 Интроспекция	54
Заклучение	56
Библиография	58
Источники исследовательского материала	59
Справочные материалы	59
Интернет-источники	60
Приложения	62
Приложение 1	62
Приложение 2	65
Приложение 3	68
Приложение 4	69
Lyhennelmä	i
Johdanto	i
1. Sanakirjat	ii
2. Sähköinen leksikografia	ii
2.1. Tekstikorpuksat kielentutkimuksessa ja kääntämisessä	iii
2.2. Tietokoneohjelmat vieraiden kielten opiskelussa	iii
3. Terminologia	iv
4. SysMLL	iv
5. Sanastotyö	v
6. Sanakirja-artikkelin rakentaminen	v
7. Harjoitukset	vi
Lopuksi	vi
Lähdeluettelo	vii

Введение

На сегодняшний день очень актуальна проблема использования компьютера в качестве тренажёра для активизации языкового и грамматического материала (см. Hart 2002).

Общеизвестно, что разным студентам требуется различное количество времени для усвоения материала. По нашему мнению, обучающие компьютерные программы могут стать отличными помощниками преподавателю и студенту, хотя заменить контактное обучение они, конечно, не смогут. Студент может заниматься в компьютерном классе после занятий или работать с домашнего компьютера, зайдя на веб-сайт университета. Настоящее исследование является продолжением бакалаврской работы автора по теме «Разработка учебного глоссария по теме «Организм» (2007).

Цель дипломной работы – составление глоссария по теме «Организм» для лексического компьютерного тренажёра SysMLL. Тема «Организм» включает в себя части тела людей и животных, а также те естественные процессы, которые происходят с организмом живого существа. Лексический тренажёр SysMLL имеет два рабочих языка: русский и финский.

Задачей дипломной работы является углублённое исследование лексикографической проблематики в составлении интерактивных глоссариев. В ходе исследования планируется:

- отобрать термины-кандидаты в глоссарий;
- опираясь на частотный фактор, получить информацию о том, какие словарные единицы необходимо включать в словник.
- дополнить, усовершенствовать и оценить состоятельность глоссария по теме «Организм» в обучающей программе SysMLL.

Первая глава дипломной работы («Типология словарей») посвящена рассмотрению видов словарей, которыми переводчик непосредственно пользуется в своей работе. Также рассматриваются частотные словари, технология составления которых подходит для составления словника.

Во второй главе («Терминоведение») приводятся два основных направления терминоведения – теоретическое и прикладное, разрабатывающие практические принципы создания, перевода и описания терминов. Основной задачей терминоведения является структуризация и рассмотрение лексических единиц. Структуризацию терминов можно осуществить при помощи ассоциативной карты, принципом которой является нахождение ключевых слов. К ключевым словам впоследствии подбираются относящиеся к ним термины, которые в свою очередь необходимо описать при помощи дефиниций.

В третьей главе работы («Электронная лексикография») рассказывается о новом направлении практической лексикографии, которое составило конкуренцию традиционной, бумажной. Также описываются универсальность и преимущества электронных словарей и компьютерных программ перед бумажными «собратьями». Также здесь рассматриваются виды компьютерных обучающих программ и их актуальность.

В четвёртой главе («SysMML – лексический тренажёр нового поколения») рассказывается об обучающей компьютерной программе для изучения иностранных языков SysMML, которая была разработана в Тамперском университете с целью расширения, активизации и тестирования словарного запаса студентов.

Пятая глава («Глоссарий по теме "Организм" в программе SysMML») рассказывает о создании глоссария в программе SysMML, уровнях сложности, устанавливаемых редактором для облегчения изучения языка, структуре статьи, о том, какие аспекты учитываются при вводе нового термина, а также о наличии упражнений, проверяющих знания пользователя.

1 Типология словарей

В справочнике студента (Opinto-opas) говорится, что в переводе принято подразделять рабочие языки переводчика на три группы: А, В и С. Языком А переводчик владеет в совершенстве и использует его как входной и выходной язык. Обычно этот язык является для переводчика родным. Язык В не является для переводчика родным, но он владеет им на высоком уровне и может при определённых условиях использовать его в качестве выходного языка, например в зависимости от уровня сложности текста и т.д. Язык С также является для переводчика неродным, и переводчик может его использовать в качестве входного языка (Opinto-opas 2007: 2).

Берков считает, что одним из основных понятий лексикографии является "входной язык", тот, с которого переводят, и "выходной язык", то есть тот, на который осуществляется перевод. Наиболее распространёнными являются одноязычные словари, но есть и двуязычные словари, а также с несколькими входными языками, впрочем, последние довольно неудобны в использовании и несовершенны в языковом плане. С несколькими выходными языками издаются, например, терминологические словари (Берков 1996: 7).

При работе со словарём даже пассивный пользователь должен обладать определёнными знаниями грамматики входного языка словаря. Каким же должен быть «образовательный ценз»? Ответ на этот вопрос можно найти у Беркова, который предлагает два критерия: во-первых, в грамматическом очерке при словаре следует выделить правила, которые могут понадобиться при работе со словарём и, во-вторых, в языках со сложной морфологией делать акцент на альтернации, к примеру, *katu* ед.ч.– *kadut* мн.ч. В случае с обучающей программой, акцент на эти критерии можно помещать в зону комментариев, но опять же только случаи чередования, заслуживающие внимания, т.к. задачей программы является расширение лексикона, а не обучение грамматике. Подводя итоги вышесказанному, можно утверждать, что пользователь программы должен обладать знаниями по основам грамматики.

1.1 Активный и пассивный словари

Активный словарь представляет собой лексический массив, т.е. лексику и фразеологию, которая употребляется в данной речевой сфере. Как считает Черняк, также существует понятие активного словарного запаса отдельных носителей языка, это множество лексических единиц, которое человек использует в трудовой деятельности и повседневной жизни. Здесь имеется в виду научная и профессиональная терминология, книжная или экспрессивно-окрашенная лексика и фразеология, тем не менее, всё вышеописанное входит в общий активный словарь языка. Последний претерпевает постоянные изменения (Черняк 2002: 301).

Так, к примеру, в словарном запасе врача будет присутствовать большое количество медицинских терминов и анатомической терминологии.

В пассивный словарный запас входят устаревшие слова (архаизмы и историзмы), а также те, которые в силу своей новизны еще не получили широкой известности (неологизмы). Слова могут переходить из пассивного словаря в активный и наоборот. Так, в свое время слово «космодром» было неологизмом, но сейчас это слово входит в активный лексический состав современного русского языка.

1.2 Двухязычный словарь

При создании двухязычного словаря проводится работа по сопоставительному исследованию обоих языков, посредством этого можно сблизить обе сопоставляемые системы и определить, научно их охарактеризовав, что их различает и тем самым создает трудности при переводе (Смирницкий 1992: 6).

Словарь должен представлять языковую норму. Берков разделяет это понятие на три составляющих: 1) представленность лексики у писателей-классиков, 2) распространённость рассматриваемого явления, 3) соответствие основным законам языка (т.е. двухязычный словарь должен отражать реальный язык образованного населения, который используется на радио и телевидении). Так, обращаясь к лингвистическому словарю, читатель ожидает, что найдет в нем лингвистические

термины как из активного словаря, так и из пассивного. Но, к примеру, обращаясь к техническому словарю, пользователь должен найти там названия технических агрегатов, деталей, свойств и т.д.

Далее перейдем к макроструктуре и микроструктуре словаря, где макроструктура – общая структура словаря, т.е. содержание и связь его частей, а микроструктура – это формат словарных статей и параметры его заполнения (Потапова 2002: 24-28).

Основным модулем двуязычного словаря является словарная статья, состоящая из леммы, зоны фонетической информации, зоны грамматической информации, зоны эквивалента и отсылочной зоны.

Лемма (лексема), одно из важнейших понятий лексикографии – это слово, составляющее единицу словника, оно является представителем всех словоформ, также лемма является в словаре реестровым словом. Выбор самой подходящей леммы для данной словарной статьи называется лемматизацией (Берков 1996: 14-15).

В зоне фонетической информации описываются особенности произношения и ударения. Например, в финско-русском словаре особенности транскрипции финского слова практически никогда не указываются, т.к. в финском языке орфография, как правило, однозначно отражает произношение. Исключение составляют заимствованные слова, например: кредо – credo [kreedoo] или чеддер - cheddar[juusto], cheddarinjuusto [šeddar, tšedar]. В англо-русском же словаре указание транскрипции для английского слова обязательно, т.к. английская орфография непоследовательная и произношение / написание слов представляет проблемы даже для носителей языка.

В зоне грамматической информации словарь даёт активному и пассивному пользователю информацию, позволяющую образовать любую форму этого слова при наличии у пользователя определённых знаний о морфологии данного языка (Берков 1996: 109).

Подводя итоги по грамматической паспортизации слова, выделим определённые требования: грамматическая информация должна быть максимально компактной и понятной широкому кругу пользователей, она должна быть систематизированной и

желательно, чтобы её можно было бы извлекать поэтапно (Берков 1996: 117). Но это уже область электронных носителей, о которой будет сказано ниже.

Эквивалентом называют слово или словосочетание выходного языка, которое автор предлагает в качестве перевода единицы входного языка словаря. Эквиваленты подразделяют на два вида: полные — представляют (почти) тот же объём информации, что и лемма входного языка, например, финское существительное *äiti* и русское *мать* (Коппалёва 2001: 866); и частичные по определению, сообщающие только часть информации, весь объём информации может быть передан суммой эквивалентов, к примеру, существительное *poika* имеет два русских эквивалента — *сын* и *мальчик*. Информация, передаваемая эквивалентом, также может быть подразделена на два вида: знаковая и семантическая. Существует также безэквивалентная лексика, для которой в выходном языке нет готового соответствия к слову входного языка (Берков 1996: 14-16), например, для русских слов *сельсовет*, *воскресник*, *гармошка*, *народоволец*, *бить челом*, *толстый журнал* нет однозначных эквивалентов в большинстве европейских языков.

Членение слова на значения (филиация) является одной из сложнейших задач лексикографии: даже в схожих по объёму и качеству словарях она может быть разной. К тому же, каждое слово в каждом конкретном языке, в зависимости от культуры народа, географического положения и др., имеет свою самостоятельную коннотацию вне зависимости от количества эквивалентов (Берков 1996: 15-17).

Конкретная лексическая единица одного языка может иметь несколько эквивалентов в другом (Коппалёва 2001: 156).

Например, в финском языке есть много слов, у которых при переводе на русский выделяется большое количество зачастую совершенно разных значений.

Например, у финского слова *hylätä* в русско-финском словаре Куусинена насчитывается 14 эквивалентов:

1. *бросать/бросить*

2. *оставлять/оставить*

3. покидать/покинуть
4. отказываться/отказаться
5. отвергать/отвергнуть
6. отрекаться/отречься
7. не зачитывать/не зачитать
8. не одобрять/не одобрить
9. (за)браковать
10. отклонять/отклонить
11. отменять/отменить
12. аннулировать
13. юр. кассировать
14. дисквалифицировать

При переводе или чтении какого-либо текста необходимо учитывать оттенки значений слов и стили речи, в которых они употребляются.

1.3 Частотные словари

Частотные словари предназначены для профессионального использования и в качестве примера имеют большую ценность для данного исследования. Частотные словари составляются для обеспечения изучения языка категориями лиц, для которых данный язык является неродным. Частотный словарь обеспечивает необходимым лингвистическим материалом процесс обучения в заданных областях использования языка, в частности в построении лексического минимума. В этих словарях каждая единица имеет указание на частоту её употребления в выборке. При составлении обычных словарей используется несколько словарей или корпус текстов. Словники частотных словарей в настоящее время составляются на основе корпусов текстов: берется набор текстов, представительный для языка в целом, для некоторой предметной области или данного автора, и из него извлекаются словоформы, леммы и части речи. По мнению Алексева при работе над частотным словарём следует отличать словоупотребление, то есть цепочку букв от пробела до пробела, от словоформы, т.е. представителя слова в тексте. Разница состоит в том, что «слово – это

одна из разных единиц текста, а словоупотребление – это единица самого текста». Исходя из этого, единицей описываемого словаря может (теоретически) быть как словоформа, так и слово, при описании словоформы важно учитывать её морфологические и семантические признаки (Алексеев 2001: 6-13,72, 86,99).

При составлении глоссария для SysMLL определялась пороговая частотность («проходной балл») для всех лексем, отбираемых для словника. На основе исследований Алексеева можно сказать, что он тоже считает полезным использование частотности слова как критерия при отборе лексикографического материала.

Наука, которая занимается исследованием частотных словарей, называется статистической лексикографией, которая в свою очередь является частью количественной лингвистики (Алексеев 2001: 8,87).

Для дальнейшего исследования данной темы следует ввести понятие *узус*, под которым понимается массовая и регулярная воспроизводимость данной единицы языка, отработанная и закреплённая в общественном употреблении. Узус складывается в реальной языковой практике носителей языка. Узус входит в состав триады: система-норма-узус. Узус также может фиксироваться фразеологическими и толковыми словарями и даже переходить в языковую норму (Энциклопедия «Русский Язык», узус).

1.4 Словник

Основой любого словаря является словник, который является перечнем словарных единиц, подлежащих переводу или толкованию, слова в котором могут быть расположены в алфавитном порядке (алфавитный словник) или по частоте (частотный словник) (Большая советская энциклопедия, *словник*).

Для составления словника необходимо решить, по какой методике будут отбираться слова. Составление словника можно начать с подбора необходимой лексики, которую можно найти в специализированной литературе по заданной теме. Затем используют метод выборки по частотной встречаемости и отбирают определенное количество терминов из общего числа найденных.

1.5 Оценка качества составления словаря

Хороший словарь должен отвечать общепринятым стандартам качества. Набор слов, словосочетаний и выражений должен образовывать закрытое множество. Словарь должен иметь удобный для пользования метод систематизации материала. Максимум сведений должен сочетаться с лаконичной и ясной формой изложения. Учебный словарь предполагает наличие иллюстраций.

При составлении словаря нужно, прежде всего, представить себя на месте человека, изучающего иностранный язык, продумать наилучший способ оформления и представления материала.

Необходимо пользоваться авторитетными источниками, такими как толковые словари, т.к. очень важно правильно и ясно объяснить значение слова.

В любом языке есть как различные грамматические, лексические, орфоэпические и прочие правила, так и исключения из них. Поэтому, помимо толкования слова, следует указывать его грамматические формы, образующиеся не по правилам, или особенности ударения, что поможет изучающему язык. Также очень полезно показать слово в каком-либо разъясняющем его контексте. А если на пути овладения чужим языком возникает недопонимание, иллюстрация поможет объяснить графически значение лексической единицы.

Приняв во внимание все вышеперечисленные аспекты, можно получить качественный словарь или глоссарий.

2 Электронная лексикография

Компьютерная (электронная) лексикография – это одно из направлений современной прикладной лингвистики. Она является принципиально новым направлением практической лексикографии и имеет свои собственные подходы не только к отображению, но и к содержанию словаря. Для начала рассмотрим негативные стороны традиционной, «бумажной» лексикографии. Существует два вида словарей: маленькие, представляющие информацию в сжатом виде, и большие профессиональные издания, объём которых может достигать нескольких томов (например, Большой академический словарь русского языка в 17-ти томах), поиск информации в последних значительно затруднён и требует больших временных затрат, что в наш век стремительного прогресса совсем неприемлемо. С другой стороны, обширное и глубокое описание лексических значений не всегда соответствует текущей языковой ситуации, т.к. процесс создания словарей длится довольно долго и многие вещи претерпевают изменения (Селегей 1996).

Во многих электронных словарях используются возможности гипертекста, под этим термином мы понимаем соединение смысловой структуры со структурами содержания при помощи технических средств. Гипертекст даёт возможность переходить по ссылкам к разным частям информации. Нелинейный характер гипертекста даёт возможность хранения информации в виде разветвлённой структуры, тем самым меняя композицию (Ооі 1998: 29-34).

Основными отличительными чертами электронных словарей являются гибкость, множественность, динамичность, обратимость, а также возможность использования мультимедийных средств. Электронный словарь представляет собой универсальное и удобное средство представления различной информации и не требует от пользователя специальной лингвистической подготовки (Шемякина 2002: 99-103).

Компьютерные же словари интересны своими новыми возможностями показа содержания словарной статьи, в некоторых словарных оболочках пользователь может выбрать интересующие его поля. В электронном словаре могут использоваться различные лингвистические технологии, такие как морфологический и синтаксический

анализ, автоматический анализ текста и т.п. Благодаря электронной форме словаря упрощается работа по поиску словарного термина, и поиск слова зависит уже не от лексемы, а от конкретной языковой единицы, которую пользователь ищет в словаре.

В электронных словарях нет необходимости выстраивания словарных статей в алфавитном порядке, поскольку сортировка автоматизирована, кроме того возможно упорядочивание по разным полям. а не только по входу словарной статьи, например, можно сортировать по дате ввода словарной статьи, по авторам статей, по тематической группе и т.п.

Составление электронного глоссария построено на иных принципах лексикографии, отличных от тех, которые приняты при создании печатных словарей. В компьютерной лексикографии применяется первичная разметка словарной статьи, в результате чего полнотекстовый поиск может различать эквиваленты от комментариев и примеров употребления. Единицей описания становится не лексема, а единое лексическое значение, на этом основании технологии анализа могут устанавливать соответствие между исходным запросом и лексическими значениями, которые релевантны для этого запроса по синтаксическим и семантическим критериям. При этом необходима синхронизация словарных описаний и формальных моделей. Такой подход решает проблему многофункциональности бумажных словарей. Лексикограф делает словарь универсальным как для случайных пользователей, так и для пользователей, использующих словарь в качестве рабочего инструмента. В будущем лексикографы всё больше смогут использовать интегральные словарные описания, основанные на формальных описаниях (Селегей 1996).

2.1 Электронные словари

В электронном словаре можно размещать намного больше информации и простота поиска слов в словаре облегчает его использование. Электронный словарь является проекцией электронной картотеки, а его печатная версия должна представлять собой распечатку базы данных по заданным форматам. Преимуществом электронного словаря является гибкость данных, т.е. используя программные средства можно редактировать уже существующие и заносить в базу данных новые словарные статьи. Компьютерные

технологии позволяют проводить рутинные операции по сортировке словника, разметке текста, поиску и редактированию фрагментов текста.

Электронный словарь может состоять из следующих функциональных блоков:

- База данных слов (электронная картотека).
- Поисковая система (для поиска словарных статей по базе данных).
- Морфологический анализатор (лемматизирует словоформы в словарных статьях, может использоваться для составления индекса)

(Кузнецов 2001)

2.2 Электронные словари лексической сочетаемости

Одной из центральных проблем при переводе является адекватность выбора слов и выражений. Как считает Попов, правила словоупотребления в родном языке нередко усложняют перевод терминов на иностранный язык, а также имеются и различия в смыслах одного и того же слова в разных языках. В словарях сочетаемости описываются коллокации. Словари сочетаемости используются при переводе наряду с двуязычными словарями (Попов 2002: 85).

Электронный словарь содержит информацию о переводных эквивалентах и о допустимой сочетаемости. Попов описывает следующие принципы создания такого словаря:

- Словосочетания одного языка описываются независимо от словосочетаний другого языка, для создания словарной части привлекается только оригинальный языковой материал. Благодаря этому исключается интерференция правил сочетаемости разных языков.
- Одноязычные словарные части объединяются в двуязычный словарь сочетаемости. Словарная статья состоит из ряда возможных терминов, при этом допускается несоответствие количества коллокаций для переводных эквивалентов.

- Языковой материал размещается таким образом, что словарь может использоваться как одноязычный или многоязычный. Такой словарь является полностью обратимым, т.е. его можно легко изменять, добавляя в него новые языковые части.
- Коллокации разбиваются на две основных группы, например, «интенсификаторы» (*kova pakkanen*) и "деинтенсификаторы" (*hieno pakkanen*). Коллокации можно снабжать стилистическими пометами, помогающими пользователю выбрать наиболее точный вариант термина.

Существует много трудов в области лексической семантики. Один из наиболее известных – теория «Смысл ↔ Текст», разработанная И.А. Мельчуком. Благодаря этой теории возможно эксплицитное теоретическое описание русской синтагматики (Попов 2002: 86).

2.3 Корпусы текстов

Корпус текстов – это вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то отрывки текстов данной проблемной области.

В принципе, любой набор более чем одного текста может быть назван корпусом (от лат. *corpus* – “body”). Часто отдельные тексты используются для разного вида литературного и лингвистического анализов. Однако, понятие корпуса как основы для электронной лингвистики все же отличается от проверки единичных текстов.

В.В. Рыков определяет корпус текстов как некоторое собрание текстов, в основе которых лежит логический замысел, логическая идея, объединяющая эти тексты. Логическая идея воплощается в правилах организации текстов в корпус; алгоритме и программе анализа корпуса текстов; сопряжённой с этим идеологии и методологии. По мнению Рыкова, корпус текстов принадлежит к четвёртой фактуре речи (тексты на машинном носителе).

В зависимости от поставленной цели выделяют несколько типов корпусов текстов:

1. по форме хранения:

- ✓ в звуковой форме;
- ✓ письменные;
- ✓ смешанные

2. по языку представления текстов:

- ✓ одноязычные;
- ✓ многоязычные;

3. по жанровой принадлежности:

- ✓ литературные;
- ✓ диалектные;
- ✓ разговорные;
- ✓ публицистические;
- ✓ смешанные;

4. по способам доступа:

- ✓ свободно доступные;
- ✓ коммерческие;
- ✓ закрытые;

5. по назначению:

- ✓ исследовательские;
- ✓ иллюстративные;

6. по динамичности:

- ✓ динамические (мониторные);
- ✓ статические;

7. по наличию дополнительной информации:

- ✓ аннотированные (размеченные);
- ✓ неразмеченные.

В.В. Рыков приводит несколько иную классификацию корпусов текстов:

1. по степени организации и структурированности:

- ∅ электронный архив – это тексты на электронном носителе, но их форма представлена на машинном носителе не стандартизирована и не унифицирована;
- ∅ электронная библиотека – тексты здесь представлены однородным и стандартизированным образом;

Ø корпус текстов – форма стандартизирована и унифицирована, тексты предназначены для отражения части лингвистической реальности;

Ø субкорпус – это некоторая автономная часть корпуса

2. по хронологическому признаку:

Ø синхронический;

Ø мониторинг (отслеживает текущее состояние языка);

Ø диахронический

3. по индексации:

Ø простой;

Ø аннотированный.

4. по языку:

Ø одноязычный;

Ø двуязычный;

Ø многоязычный.

5. по способу применения и использования корпуса:

Ø исследовательский;

Ø иллюстративный;

Ø параллельный.

6. по способу существования корпуса:

Ø динамический;

Ø статический.

Первоначальные лингвистические исследования, проводившиеся с помощью корпусов, сводились к подсчету частот встречаемости различных языковых элементов. Чаще всего этими элементами были слова, в других случаях – графемы, морфемы, словосочетания.

По прошествии времени корпусы стали осознаваться как мощные информационные ресурсы, могущие быть использованными в рамках различных лингвистических направлений. Так, корпусы являются богатым источником данных для лексикографии. На их основе и с применением компьютера словари могут составляться и пересматриваться гораздо быстрее, чем раньше, таким образом фиксируя текущее состояние языка и не успевая устаревать за то время, которое проходит от момента начала работы над ними до момента выхода их из печати.

С исследованиями по лексикографии тесно связаны исследования в области семантики. Наблюдая окружения той или иной лингвистической единицы в корпусе, можно установить определенные семантические признаки, характеризующие данную единицу. Часто слово входит сразу в несколько семантических категорий, поэтому следует говорить только о степени принадлежности слова к той или иной категории. Степень же может быть выявлена путем подсчета частот его распределения по разным категориям. (Рыков В.В., Диалог 2002)

В настоящее время корпуса письменных и устных текстов успешно применяются при обучении иностранным языкам и в лингвопедагогике. На базе корпусов формируются списки активной лексики, частотные списки терминов для использования в курсах по терминологии и т.п. Разработчики академических словарей и учебных пособий опираются на массивы аутентичных текстов (*Corpora*). Корпусная лингвистика дает материал для различного рода исследований языка и его вариантов, и определяет основной метод анализа текстов на базе корпусов (*Corpus-Based Approach*). Корпусный подход, или метод лингвистического исследования, основанный на корпусах текстов, ориентирован на прикладное изучение языка, его функционирование в реальных средах и текстах, что важно для преподавания языка (Баранов 2001: 112).

Особым типом корпуса является так называемый параллельный корпус, в котором тексту на русском языке сопоставлен перевод этого текста на другой язык или, наоборот, тексту на иностранном языке сопоставлен его перевод на русский язык. Между единицами оригинального и переводного текста (обычно — между предложениями) с помощью специальной процедуры устанавливается соответствие; эта процедура называется выравниванием, а тексты, соответственно, выровненными.

Выровненный параллельный корпус представляет собой важный инструмент для научных исследований (в том числе и для исследований по теории и практике перевода); он может также использоваться при обучении русскому и иностранным языкам.

Один из самых больших корпусов русского языка – *Национальный корпус русского литературного языка*, объем которого больше 140 млн.слов. Но также существуют и

другие, например, *Корпус газетных текстов русского языка конца XX века*, *БОКР*, *Balkan Rusistics* и т.д.

Самый крупный корпус финского языка – *FTC, Finnish Text Collection*, с помощью которого можно проверять частотность слов при работе с глоссарием.

2.4 Компьютерное обучение иностранным языкам

Наука, исследующая применение компьютера в изучении иностранных языков, называется компьютерной лингводидактикой. Это междисциплинарная область знаний, тесно связанная с новыми информационными технологиями, прикладной лингвистикой, компьютерной психологией, теорией и практикой компьютерного обучения.

Лингводидактика имеет три направления исследований:

1. Разработка теоретических аспектов использования компьютера в обучении иностранному языку
2. Разработка и создание компьютерных программ
3. Интеграция компьютера в учебный процесс

(Бовтенко 2005: 8).

Задача лингводидактики – помочь педагогу-практику в применении компьютерных технологий для обучения иностранному языку. Это может быть осуществлено на базе опыта, приобретённого в результате практической работы по созданию электронных обучающих материалов. Неумалима и важность выбора инструментов, которые преподаватель сможет использовать в своей работе. Другим аспектом в этом вопросе является активное участие педагога в создании обучающих программ на всех стадиях разработки, а также возможность постоянного редактирования и обновления продукта самим преподавателем.

2.5 Обучающие программы и тесты

Большое количество обучающих программ и их труднодоступность затрудняет процесс исследования. В своей статье «Структура и содержание лексических баз данных для обучающей программы по иностранным языкам.», Михайлов(2006) разделил все обучающие программы по иностранным языкам на три категории:

- Пустые оболочки для создания рабочих глоссариев / упражнений
- Программы-тесты
- Обучающие игры

Рассмотрим эти программы подробнее. Оболочки для глоссариев создаются программистами, не имеющими лингвистического образования и представляют собой программу в которой, предусмотрен ввод пар «слово – эквивалент», существует возможность добавления графического изображения, но ни в одной из известных программ нельзя давать примеры употребления, также в этих программах не приводится толкование и грамматическая информация (напр. Hot Potatoes, KVocTrain, Open Book). Лучшими из этого типа программ являются утилиты для заучивания слов, входящие в пакет программ, обслуживающих электронные словари, например, в словарях Lingvo, для которых разработан лексический тренажер Lingvo Tutor.

Далее рассмотрим программы-тесты, они создаются преподавателями иностранных языков в сотрудничестве с программистами. Чаще всего такие программы представляют собой электронные реализации «бумажных» упражнений по лексике и грамматике. В этих программах есть упражнения и тесты, которые строятся по шаблонам например: «Вставить слово в правильной форме» или «Выбрать правильный ответ». Упражнения разбивают по темам или по уровню сложности. У этих программ есть свои недостатки, например, обучаемый не получает комментариев по своим ошибкам.

Последними Михайлов описывает игровые программы, которые представляют собой компьютерные реализации языковых игр: кроссвордов, «Виселицы», "Крестиков и ноликов" и т.п., а также сюжетные игры. Такие программы позволяют развивать

некоторые языковые умения и расширяют словарный запас. По своей структуре эти программы не способны помочь в обучении грамматике.

2.6 Лексические тренажеры

Основная цель программы – предоставить изучающим иностранный язык эффективное средство для расширения лексического запаса. Лексический тренажер можно представить как эффективную, объемную и расширяемую электронную картотеку – подобно словарной базе данных. Для каждого отдельного слова накапливается большой объем статистической информации, которая позволяет студенту быстро найти и выучить проблемные слова. Программа помогает обучаемому повторить плохо усвоенные слова, выучить не только значение слов, но и их произношение. Лексический тренажер может содержать в себе довольно объемный словарь. Кроме изучения слов, обучаемый может выполнять задания для закрепления полученных знаний.

Такая программа позволяет выучивать только словарные пары. Она полезна тем, что с её помощью студент разучивает правильное произношение слов. Эта программа хороша на раннем этапе изучения иностранных языков, также для «любителей», т.е. людей которым иностранный язык нужен, например, для кратковременной поездки на отдых за границу.

Примером лексического тренажера является программа Lex!, она обеспечивает изучение и запоминание иностранных слов во время работы за компьютером. Программа периодически активизируется и предлагает переводить слова, поддерживает транскрипцию, работу с произвольным количеством словарей, имеет полностью настраиваемые режимы работы. Возможно создание и удаление словарей, добавление и редактирование лексики, изменения временных параметров работы и уровней интенсивности обучения (Информация по программе Lex! и её описание).

2.7 Оценка качества функционирования компьютерных программ по обучению иностранному языку

Общепринятыми характеристиками оценки компьютерных продуктов в обучении иностранным языкам являлась оценка их технических характеристик, уровень надёжности графического дизайна, надёжность программного решения и т.д. (Балыхина 2006: 32).

Однако, как выяснилось на практике, вышеперечисленные характеристики не совсем подходят для интересующих нас программ, т.к. эти факторы не являются основополагающими цели применения продукта. М.Г. Евдокимова считает, что решающими факторами оценки программных продуктов для обучения иностранных языков, должны отражать представления о природе языка, механизмов его усвоения, о стиле и характере взаимодействия преподавателя и студента.

Различны и способы использования компьютера при обучении, первый из которых подразумевает «тьюторскую форму» обучения, т.е. компьютер заменяет преподавателя, а второй способ подразумевает применения компьютера в качестве инструмента, например лексический тренажёр по изучению лексики или различные игровые варианты программ.

В идеальном варианте тьюторская программа должна представлять собой интеллектуальную обучающую систему, включающую в себя представление о языке как предмете обучения, индивидуальный подход к обучаемому в зависимости от уровня его языковой подготовки, а также методики обучения языкам и экспертные модули (Евдокимова 1999).

Н.В. Ялаева определяет следующие преимущества компьютерных программ:

- Обучающие компьютерные программы являются относительно новым явлением и представляют больший интерес, чем простые бумажные источники.
- Компьютерные программы – это эффективное средство самообучения и хорошая форма самостоятельной работы, многие программы рассчитаны на то, чтобы студент сам выбирал уровень сложности.

- Компьютер предполагает огромные возможности в варьировании заданий, он может воспроизводить и записывать звук, показывать видеоматериалы и самостоятельно генерировать задания.
- Использование компьютера при изучении языка способствует повышению общей компьютерной грамотности обучаемых.

3 Терминоведение

Терминология — специфический сектор лексики, система терминов — слов научного (философского, общественно-политического, научно-технического) языка, обладающих определенным, четко отграниченным значением; значение каждого термина координировано со значением всех остальных терминов той же сферы (Литературная энциклопедия).

В настоящее время в терминоведении выделяют два основных направления: теоретическое, исследующее развитие и употребление специальной лексики, и прикладное, разрабатывающее практические принципы создания, перевода и описания терминов. Терминоведение, в свою очередь, подразделяют на общее и отраслевое. Общее терминоведение изучает свойства, проблемы и процессы специальной лексики, а отраслевое занимается изучением специальной лексики в конкретных языках. Ещё можно обратить внимание на сопоставительное терминоведение, которое исследует общие свойства и особенности специальной лексики разных языков, например, русского и финского. Ономаσιологическое терминоведение занимается изучением структурных форм специальных лексем и выбором оптимальных форм наименований. Для нашего исследования особенно важным может оказаться функциональное терминоведение, изучающие современные функции термина в подготовке переводчиков и использовании терминов в компьютерных системах (Гринёв 1993: 15).

Отбор специализированной лексики для обычного словаря относится к наиболее сложным проблемам словарной работы. Трудность состоит в выборе слов для словника, является ли слово общеупотребительным или узкоспециальным?

Работа по поиску терминов начинается с упорядочения, а последнее в свою очередь – с систематизации понятий выбранной области по категориям и «построение классификационных схем понятий». В результате выявляются признаки понятий и на их основании создаются дефиниции. На следующем этапе работы производится анализ терминологии, первым этапом которого является семантический анализ, выявляющий отклонения значения терминов от содержания называемых ими понятий, а также выявляются термины, не имеющие наименований. Вторым этапом терминологического

анализа является этимологический анализ, на основании которого определяются наиболее эффективные способы и модели образования терминов по заданной теме, при помощи этого анализа можно выявить неудачные формы терминов и определить способы их замены. На следующем этапе проводится нормализация терминов, включающая в себя унификацию, в процессе которой происходит упорядочение терминов, и оптимизацию, которая выбирает оптимальную форму термина. Цель унификации – это закрепление соответствия между понятием и представляющим его термином (Ахманова 1968: 158).

Подводя итоги, сошлёмся на Э.Я. Якубайтиса, который считает, что тексты определений терминов должны быть отделены от описания особенностей обозначаемых ими понятий. При определении терминов следует как можно меньше использовать другие термины.

При выборе терминов немаловажна последовательность, наличие системности. В словарь должны включаться термины приблизительно одинаковой важности и частотности, а также парные термины. Также важно учитывать дескриптивность и нормативность термина (Кудашев И.С., Кудашева И.О. Диалог, 2008).

То есть, если в тему «Организм» включен термин *малый круг кровообращения*, то надо включить и *большой круг кровообращения*, т.к. эти термины парные. А, например, при описании полости рта надо включить все основные термины, к этой области относящиеся: *зубы, язык, нёбо* и т.д.

3.1 Ассоциативная карта или логико-понятийная схема

Сбор терминов на заданную тему можно производить, читая текст и выписывая слова, а можно систематизировать материал при помощи ассоциативной карты (диаграмма связей, Приложение 2). Такой принцип изображения системного мышления был разработан профессором Джозефом Новаком (Новак 1995: 30-57) еще в 60-е годы. Принципом такой схемы является нахождение ключевых слов, к которым впоследствии подбираются относящиеся к ним термины, а далее к найденным терминам - их подтермины. Схема может иметь древовидную или другие формы в зависимости от

вида информации. Процесс продолжается до того момента, пока тема не будет исчерпана.

В основе этой техники лежит принцип «радиального мышления», относящейся к ассоциативным мыслительным процессам, отправной точкой которых является центральный объект. Такие диаграммы позволяют структурировать и классифицировать информацию.

Например, к *ССС (Сердечно-сосудистой системе)* относятся *сердце* и *сосуды*, сосуды в свою очередь делятся на *артерии* и *вены* и т.д.

Получаемая графическая структура позволяет создать сеть из понятий и выделить основные и второстепенные термины, а также определить недостающие понятия, которые при простом поиске терминов могут быть упущены из виду.

Логико-понятийные схемы показывают логические типы отношений между специальными понятиями в одной терминсистеме. Такими отношениями могут быть родовые, синонимические и однородные. Понятийные отношения в рамках этого метода могут описываться на сигматическом и прагматическом уровнях. На семантическом уровне отношения воспроизводятся путём перечисления терминов-синонимов, однородных терминов, родовых пар терминов. На прагматическом уровне представляется графически понятийная иерархия, отражая при этом определённую совокупность понятий, связанных разными типами отношений (Табанакова 2009).

3.2 Виды дефиниций

Дефиниция (= определение, толкование) раскрывает содержание понятия, позволяет отличать предмет, отражаемый понятием, от сходных с ним предметов, устанавливать значение того или иного термина. Раскрыть содержание понятия — значит перечислить его существенные признаки, то есть признаки, необходимые и достаточные для того, чтобы отличить данный предмет от сходных с ним предметов. Определение должно быть написано ясным, понятным языком. Оно имеет нейтральное значение и может использоваться во всех стилях речи.

Далее приводятся виды определений и примеры, часть которых придумана, другие взяты из словарей. Классификация видов определения взята из учебника по лексикографии Хаарала (Haarala, 1981).

Содержательное определение

Содержательное определение поясняет смысл термина, указывает ближайшее понятие и отличительные признаки по сравнению с определениями других понятий.

(1)

ноготь

роговая пластинка на концах пальцев человека.

Групповое определение

Групповое определение поясняет смысл термина, указывая на все ближайшие к нему понятия, расположенные в иерархическом порядке.

(2)

тело человека

бедро, кисть, плечо, ребро

(3)

костно-мышечная система

кости тела, хрящи, суставы и связки, соединяющие их, а также мышцы тела

Контекстуальное определение

Контекстуальное определение позволяет понять незнакомое слово через контекст. К типичным контекстуальным определениям относят дополнительную часть статьи, например, примеры употребления в толковом словаре.

(4)

железа, *ы, мн. железы, желёз, железам, ж.* Орган, вырабатывающий и выделяющий гормоны или другие вещества, обеспечивающие жизнедеятельность организма. *Железы внутренней секреции. Грудная ж. Слюнная ж.* (Ожегов, 2009)

Нулевое определение

Нулевое определение – это такое определение, которое описывает лишь главные черты понятия, а само понятие не описывается.

(5)

hengityselimistö

elimet, jotka osallistuvat kaasujen vaihtoon ympäristön ja verenkierron välillä (Nienstedt, Kellosalo, 2007)

Кольцевые определения

Кольцевое определение имеет место, когда понятие объясняется с помощью собственного термина (внутреннее кольцо), или когда несколько понятий взаимосвязаны друг с другом (внешнее кольцо).

Однако такое построение толкований не приветствуется, поскольку в конечном итоге пользователя словаря заставляют "ходить по кругу".

Внутреннее кольцо

(6)

легкие

Как известно, жизненно важными органами человека являются легкие, где происходит газообмен, и кислород из воздушных мешочков, носящих название альвеол, передается в кровеносные сосуды. (НКРЯ)

Внешнее кольцо

(7)

защита от повреждений

препятствующая нанесению вреда защита

(8)

препятствующая нанесению вреда защита

Деятельность, главная задача которой – предотвращение несчастных случаев.

(9)

защита

Деятельность, направленная на уменьшение ущерба в результате несчастного случая.

Несовершенное определение

Несовершенное определение – это такое определение, в котором отсутствуют важные элементы либо частично, либо полностью. Несовершенные определения используются в основном в словарях общей лексики.

(10)

печень

один из внутренних органов человека

В этом определении не сообщается ни об одном свойстве определяемого понятия.

Отрицательное определение

Отрицательное определение отображает лишь те черты и значения, которые не относятся к понятию. Отрицательное определение применяется лишь тогда, когда отсутствие каких-либо свойств является важным при объяснении значения слова.

(11)

Сердечная недостаточность

Сердечная недостаточность – это не остановка сердца.

Дополняющее (редундантное) определение

Дополняющее определение – это такое определение, которое объясняет не только смысл понятия, но и сообщает об его основных свойствах.

(12)

желчь

густое вещество желтоватого цвета, вырабатываемое печенью и поступающее в двенадцатиперстную кишку (НКРЯ)

(13)

желчный пузырь

орган позвоночных животных и человека, в котором накапливается жёлчь (НКРЯ)

В данной дипломной работе использовались разные типы определений, особенности которых пояснялись ранее. Правила составления словника таковы, что невозможно использование кольцевого и контекстуального определений, где слово объясняется в контексте или же с помощью собственного термина. Повторение неуместно.

Отрицательное определение не встречается в данном глоссарии, так как оно на практике вообще используется довольно редко.

Чаще всего в нашем материале используется содержательное определение, так как основной способ пояснения смысла термина – указание на близкое по значению с ним понятие.

Примеры содержательных определений:

(14)

ruumis – Ihmisen tai eläimen koko elimistö

(15)

melo – Физическая сторона человека, антоним – душа

(16)

maksa – Vatsaontelossa sijaitseva suuri (ruoansulatus)rauhanen.

(17)

mozжечок – Отдел головного мозга позвоночных животных и человека, расположенный в задней части черепной коробки и участвующий преимущественно в регуляции равновесия тела и координации движений.

(18)

kynsi – sormen tai varpaan pään sarveismuodostuma

Примеры нулевых определений:

(19)

keuhkoputki – Parillinen hengityselin, kuuluvat alahengitysteihin ja ne johtavat hengitettävä ilma henkitorvesta keuhkoihin.

(20)

sisäänhengitys – hengitysvaihe jonka aikana keuhkot täyttyvät ilmalla

(21)

selkäranka – elimistön sisäinen luista ja rustoista koostuva taipuva tukiranka

(22)

sappirakko – maksaan kiinnittynyt ruoansulatusnestettä sappea erittävä elin

Примеры дополнительных определений:

(23)

вена – Кровеносный сосуд, по которому кровь возвращается из разных частей организма в сердце.

(24)

венозный – Прилагательное от 'вена'.

Однако при занесении данных в программу поиск определений к некоторым словам был затруднён. Так, сложности возникли с такими анатомическими терминами, как *вертебральный* и *дорсальный*. Впрочем, впоследствии выяснилось, что эти термины являются узкоспециальными и в учебном глоссарии не нужны.

3.3 Языковые примеры употребления термина в словарной статье

Почти во всех словарях, в зависимости от типа, даются примеры употребления. Например, в толковых словарях и тезаурусах пример даётся наряду с дефиницией для лучшего понимания значения слова. В переводных словарях, в зависимости от того, на кого рассчитан словарь, примеры даются для иллюстрирования употребления слова.

Для описания языковых явлений Хаарала выделяет три вида языковых примеров: описательный, количественный и контекстуальный.

Задача описательного примера - дать пояснение термина, описать его общие свойства. Например, *печень – это крупная железа человека и животных, вырабатывающая желчь и участвующая в процессе пищеварения.*

Количественный пример описывает составляющие, структуру термина. *Печень состоит из 2-х частей - левой и правой, разделенных между собой связкой (фальциформной), и имеет 4 доли: левую, правую, квадратную и хвостовую, которые кровоснабжаются левой и правой ветвями портальной вены и печеночной артерией.*

Контекстуальный пример поясняет термин через контекст. *Рак печени встречается в 10 раз чаще в развивающихся странах - Азии и Африке.*

Каждый термин в словнике SysMLL по теме «Организм» истолковывался и во многих случаях иллюстрировался. Также приводились примеры использования слова в контексте. Если у человека, изучающего иностранный язык, возникают затруднения при понимании лексических единиц, то употребление слов и разъяснение их в контексте проясняет значение термина.

При поиске примеров использовался Национальный корпус русского языка (НКРЯ), который, на момент написания данной работы включал в себя собрание русских текстов в электронной форме общим объемом более 140 тыс. словоупотреблений.

Корпус предназначен для всех, кто интересуется самыми разными вопросами, связанными с русским языком: профессиональных лингвистов, преподавателей языка, школьников и студентов, а также для лиц, для которых русский язык не является родным.

Из корпуса текстов можно получить огромное количество примеров употребления терминов, что дает возможность подобрать наилучший вариант, поясняющий термин в контексте.

4 SysMML–лексический тренажёр нового поколения

SysMML (Multilingual System for Language Learning) – обучающая программа для изучения иностранных языков. Данная программа разработана в Тамперском университете. В её создании принимали участие: Институт современных языков и переводоведения, Центр обучения иностранным языкам, Отделение вычислительной техники и Виртуальный университет. Программа создана для эффективного обучения студентов отделения переводоведения иностранным языкам. Целью создания системы является расширение, активизация и тестирование словарного запаса студентов.

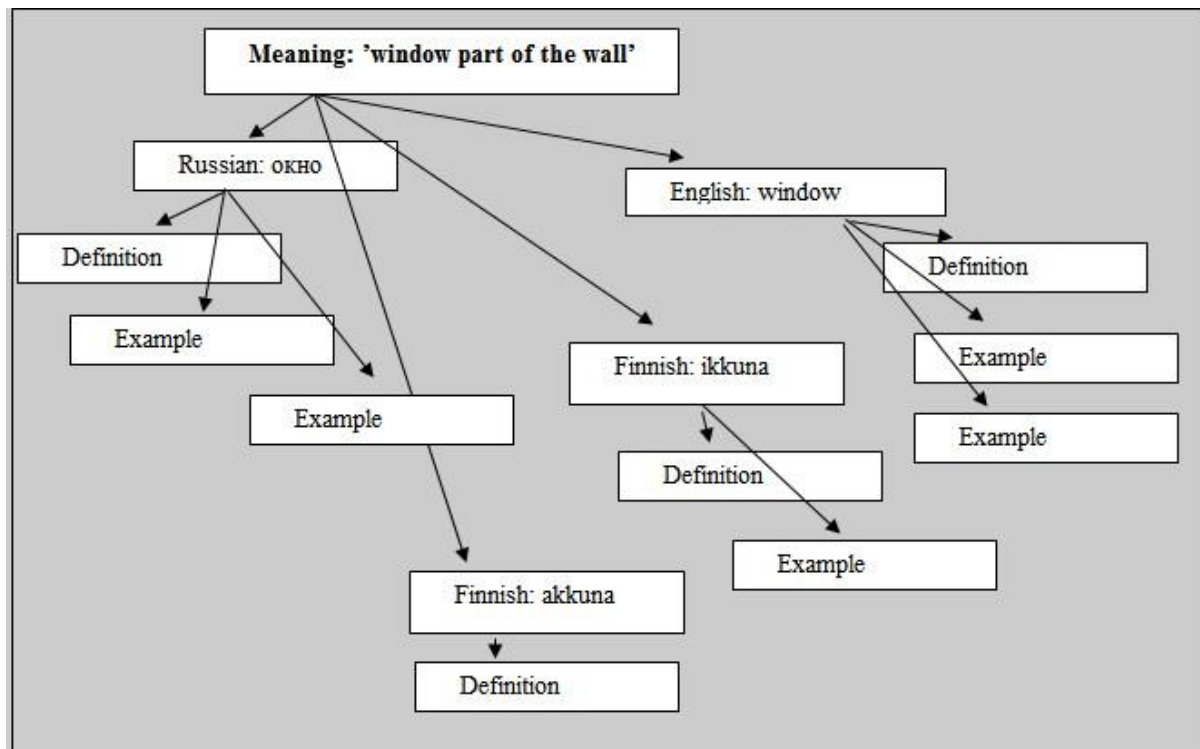
Базовым является русско-финский модуль, впоследствии будет добавлен и английский, также программу можно обогатить и другими языками. База данных и программная оболочка размещены на одном из серверов Тамперского университета и доступны через Интернет по адресу: <https://www11.uta.fi/laitokset/kielikeskus/sysmml/> (для входа в систему необходим логин и пароль, которые можно получить у администратора базы данных). В настоящее время идет тестирование программы на небольших группах пользователей. Описываемая в настоящей работе обучающая программа является лексическим тренажером нового поколения, усиленным в плане лингвистических данных и реализующая «индивидуальный подход» к пользователю.

Как показывает практика, хорошо структурированная словарная база данных значительно повышает потенциал компьютерных программ по изучению иностранных языков. Преимуществом SysMML является то, что лексический материал легко пополнять новыми записями, корректирование и уточнение также не составляет проблем. Программа предназначена для обучения переводчиков, но также подходит и для других студентов, углубленно изучающих иностранный язык.

Основной единицей хранения в базе данных SysMML является лексико-семантический вариант, а не лексема, что позволяет гибко решать проблему многозначности, омонимии и отсутствия точного эквивалента в другом языке. Значения слов описываются с помощью наборов английских ключевых слов, например ‘director film production person’ = рус. *режиссер, режиссер-постановщик*, англ. *director*, фин. *elokuvahjaaja*. Данный подход в некоторой степени напоминает подход применяемый

в словарных проектах WordNet, FrameNet и др., но в сильно упрощенном виде (о WordNet см., например, <http://wordnet.princeton.edu/>).

таблица 1, **Значение слова**



(Михайлов 2006)

Группы записей с одинаковым значением объединяются в гнезда, как в только что приведенном примере. Таким образом, тренажер может получать из базы данных группы переводных эквивалентов и (квази) синонимические ряды.

В базе данных на настоящий момент около 4015 записей, большую часть которых составляют русские и финские слова и выражения. Пакет веб-приложений генерирует различные упражнения: просмотр межъязыковых соответствий, подбор переводных эквивалентов, подбор слов и выражений по толкованию.

Для каждого пользователя ресурса создается свой профиль, позволяющий задать языковую пару, уровни сложности и темы для работы. Программа запоминает ошибки пользователя и постоянно заставляет его делать «работу над ошибками», повторяя те задания, с которыми пользователь не справился.

По мнению Михайлова, опираясь на личностно-ориентированный подход, обучающая компьютерная программа предполагает самостоятельность студентов, опору на их знания в других областях, учёт социокультурных особенностей, целенаправленное формирование умений и навыков, перераспределение ролей студента и преподавателя, присвоение последнему функций помощника и консультанта.

Процесс работы с SysMLL должен основываться на ряде принципов

- Систематизация языковых единиц
- Использование внутренней мотивации, а также личного вклада студента
- Учёт языкового уровня студента
- Учёт влияния родного языка и особенностей изучаемого на процесс обучения

4.1 Лексикографические аспекты

Многие терминологические словари составляются с несколькими выходными языками издаются (Берков 1996: 7). Вышеуказанное свойство применимо и в случае с программой SysMLL, в которой списки терминов четко подразделены на темы и в дальнейшем составители программы предусмотрели добавление к финскому и русскому языкам других языков, в данном случае английский эквивалент будет очень полезен для составления таких глоссариев. При составлении глоссария важно учитывать потребности пользователя программы.

В программе SysMLL предусмотрены три уровня сложности: элементарный (elementary), базовый (intermediate) и специальный (advanced). При работе с программой пользователь может самостоятельно выбирать наиболее подходящий для себя уровень сложности. Освоив лексику более лёгкого уровня, пользователь может перейти на следующий уровень. Дефиниция в большинстве случаев представлена на уровне языка А. На данном этапе программа SysMLL рассчитана на две категории пользователей: носителей финского и носителей русского языков. Поэтому при составлении дефиниций важно учитывать этот фактор.

В программе SysMLL была поставлена задача описать наиболее используемую лексику по заданной теме, поэтому данный вид классификации попадает в обе из вышеописанных тем. Одним из центральных понятий в лексикографии является словник, который является совокупностью слов (лемм) входного языка. Лемма – это представитель всех форм слова. Слова в словнике располагаются в алфавитном порядке (Берков 1996:12-14).

4.2 SysMLL и двуязычный словарь – сходства и различия

Основным модулем двуязычного словаря является словарная статья, как уже описывалось в главе 1.1, она имеет определённые зоны, в которых происходит описание слова. В бумажной версии словаря написанная статья не подлежит изменениям, изменения возможны лишь в последующем издании словаря. Словарная статья SysMLL имеет также определённые рамки, но, несмотря на это, отличается в лучшую сторону от словарной статьи двуязычного словаря. База данных SysMLL постоянно дорабатывается и дополняется. Пользователи могут сами находить ошибки в словарных статьях и исправлять их совместно с администратором. Таким образом, словник SysMLL не является написанным «на века», а живёт в ногу со временем.

Далее рассмотрим зоны словаря. Основой словарной статьи является лемма. Особенности её отбора посвящена следующая глава. Зона фонетической информации в словаре закрыта. Тренажёр предназначен для студентов, обладающих по крайней мере минимальными грамматическими знаниями по рабочим языкам. Двуязычный же словарь рассчитан на более широкую аудиторию пользователей и должен обладать фонетической информацией. Однако существуют словари, которые не имеют фонетической зоны, например, финско-русские словари .

Зона грамматической информации в двуязычном словаре даёт сведения по образованию любого элемента парадигмы слова, другими словами даётся парадигматическая характеристика слова (Берков 1996: 109). Такой словарь является справочником, в лучшем случае – инструментом для работы. В задачу SysMLL входит обучение лексике, а не грамматическим основам языка. Однако информация о нестандартных случаях помещается в зоне комментариев.

Зона эквивалента в двуязычных словарях содержит все значения описываемого слова. В словарной базе данных, как уже говорилось, основной единицей является не лексема, а лексико-семантический вариант. Записи словах разных языков связываются между собой через зону значения. Если имеется несколько слов, имеющих одно и то же значение (синонимия) или если в одном из языков значение оказывается более широким, чем в другом, то получается группа словарных статей, связанных между собой. Филяция в словнике программы была проведена с учётом заданной темы глоссария, т.е. значения, выходящие за рамки темы, не заносились в базу данных. Например, в "Новом словаре русского языка" у слова *лопатка* есть четыре значения: «кость», «плоский вертикальный выступ на стене здания», «молодой стручок гороха» и уменьш. к сущ. *лопата*. В тему «Организм» подходит только одно значение из этого ряда: 'кость'. Такой вид филиации называется семантизацией, т.е. выбирается наиболее подходящее значение для заданной темы.

При работе с программой аспект эквивалента представлен в двух ракурсах. Перед работой с программой пользователь может просмотреть слова и эквиваленты на обоих языках, далее при выполнении упражнений к описываемому слову не даётся эквивалента, для того, чтобы обучаемый сам догадался, о чём идёт речь, в процессе обучения это достаточно важный момент, т.к. при подаче «готового» ответа – процесс запоминания малоэффективен, когда же обучаемый начнёт сам думать над проблемой, то обучение станет намного эффективнее. Эквивалент слова в упражнениях программы заменяет дефиниция, прочитав которую студент должен понять, о каком слове идёт речь. Кроме того, в базе данных имеются примеры употребления слова, так что обучаемый всегда может знать, в каких контекстах данное слово употребляется. В двуязычных словарях пример использования слова не всегда представлен ввиду ограничений объема в бумажных версиях, а в электронных – из-за некоторых особенностей интерфейса.

Последним объектом сравнения двуязычного словаря и словарной статьи лексического тренажёра может стать отсылочная зона. В двуязычном словаре она указывает пользователю ту зону словаря, в которой он может найти более обширную информацию по интересующей его теме. Например, при поиске слова *наряжаться* в Русско-финском словаре под ред. Куусинена указано «см. *нарядиться*», т.к. *нарядиться*

это совершенный вид глагола *наряжаться*, семантически отличается от первого глагола не очень сильно и может даваться в виде отсылочной статьи. В лексическом тренажёре эта зона словарной статьи не может присутствовать, поскольку для составления упражнений нужны полные словарные статьи. Поэтому для таких слов толкование дублируется и указывается грамматический признак либо толкование модифицируется так, чтобы было понятно, о каком элементе пары идет речь.

Подводя итоги вышесказанному, нельзя сказать, хуже или лучше лексический тренажёр, чем двуязычный словарь, ясно лишь одно, что у этих двух продуктов разное назначение.

4.3 Словник SysMLL

При составлении первичного словника для SysMLL допустимо влияние фактора, описанного выше: при определении частотности слов многие слова не достигли требуемого уровня частотности, и в результате этого были исключены из словника. Однако их важность в повседневной устной речи могла быть намного выше, чем в письменных источниках. Берков в этом случае советует руководствоваться функциональной ролью лексемы в данном языковом коллективе, а также ролью эквивалента у говорящих на выходном языке. Он напоминает о нижнем пороге словника, который характеризуется одним уровнем частотности, т.е. при изменении уровня частотности меняется и объём словника.

При составлении словника для SysMLL использовалась специализированная литература по зоологии, биологии и анатомии человека, см. «Человек. Учебник 8 кл. общеобразовательных учреждений» (2006), «Биология: Животные. Учебник для 7 кл. общеобразовательных учреждений» (2006) и др.

Работа над словником началась с составления ассоциативной карты (англ. *mind map*), вся тема «Организм» была поделена на подтемы разного уровня. За основу было выбрано две основных ветви «человек» и «животный мир», последние объединяет все специальные названия частей тела живых организмов и другую лексику.

Рассмотрим этот вопрос подробнее, тема «человек» состоит из двух основных подгрупп: общеупотребительные, например, *рука, нога* и специальные, например, *пищевод, предсердие* и т.д. Как отмечено выше, схема была разделена на две основных темы, «человек» и «животный мир». Первая была поделена на системы органов человека (анатомическая терминология): пищеварительная, сердечно-сосудистая(ССС), нервная и т.д. Далее выявлялись органы, составляющие систему, например в СССР входят сердце и сосуды, после этого, проводилась классификация органов, например *сосуды: вены, артерии, артериолы*, и при необходимости включались части органов, например *сердце: желудочек, предсердие*. (Приложение 3)

4.4 Терминологические аспекты в работе по составлению глоссария

Для составления глоссария по теме «Организм» в качестве теоретической базы используется терминография, наука, изучающая составление словарей специальной лексики. Для этого необходимо определить понятия, с которыми мы будем работать в дальнейшем.

Термин – это понятие, принимаемое для наименования общих понятий. **Номен** – это единичное понятие, наименование конкретной вещи или явления (БСЭ, *номен, термин*). Отличие термина от номена состоит в том, что термин – это общее понятие, а номен – конкретное. Рассмотрим эту особенность на конкретных примерах. В глоссарии "Организм" слово *кость* является термином, а слово *адреналин* – номеном. Семантическими характеристиками термина являются его однозначность, полнзначность, т.е. его значение отражает минимальное количество признаков, облегчающих понимание обозначаемого им понятия и отсутствие синонимов. К функциональным признакам термина относят его внедрённость в профессиональный лексикон, общепринятость и употребительность, термин должен быть достаточно современным (Суперанская и др. 1989: 29).

При работе с терминами применяется лингвистическая модель терминологической системы по заданной теме. Все термины состоят в «родовидовых» отношениях, они должны быть логичны и системны. В терминосистеме знаковые элементы должны быть

материальны, например, переданы графически. Термины необходимо объединять в группы, различные по обозначаемым ими понятиям и формальным признакам (Гринёв 1993: 30).

В глоссарий желательнее заносить по большей части термины, а не номены, которые являются словами более специальными и вопрос об их разучивании менее актуален. Поиск конкретных терминов необходим во избежание использования переводчиками предтерминов, которые не отвечают требованиям, предъявляемым к терминам. В качестве предтерминов применяются: описательные обороты, которые не являются краткими и не описывают точно сущность явления; сочинительные словосочетания; причастные и деепричастные обороты. Предтермины обычно используются для описания новых понятий, они отличаются своей недолговечностью, неустойчивостью формы, отсутствием стилистической нейтральности и не соответствуют требованиям краткости и общепринятости (Гринёв 1993: 25).

Переводчики могут использовать предтермины при переводе реалий, лексем, обозначающих явления, которые отсутствуют в выходном языке. В глоссарий темы Организм предтермины по определению включены быть не могут.

Заключительным этапом в составлении глоссария является кодификация полученной терминосистемы, т.е. занесение записей в базу данных SysMLL. В случае невозможности описания термина необходима его «стандартизация». В случае с глоссарием по теме «Организм», где на данном этапе работы присутствуют два языка, необходима «гармонизация» терминов, поиск точных соответствий в обоих языках. На этом этапе возможно выявление разницы в количественном соотношении терминов между языками. Это связано с тем, что на разных языках одно и то же явление не всегда имеет однозначное название и свой отдельный термин (Гринёв 1993: 30).

5 Глоссарий по теме "Организм" в обучающей программе SysMLL

Глоссарий в SysMLL – это комплексное введение в одну из тем программы. Каждая тема описывает определённую область знаний, например, тема «Автомобиль», описывает лексику, связанную со строением и функционированием автомобиля. Глоссарий состоит из терминов, смысл которых раскрывается через описание, представленное в словарных статьях программы. Каждый термин глоссария приводится в своей начальной форме (например, именительный падеж единственного числа у существительных, инфинитив у глаголов и т.п.) (Лингвистический энциклопедический словарь 2008, *глоссарий*).

Глоссарий по теме организм составлялся по двум критериям – частотному и употребляемости. Материал для работы был определён с учётом тематики и уровня глоссария, который должен давать представления о базовой терминологии по теме "Организм". В российской системе образования такую информацию представляют учебники по биологии, анатомии человека и зоологии (см. список литературы Драгомилов и др. 2006).

Данная тема довольно широка для исследования и в ней существует два основных раздела – специальные термины и слова, употребляемые в повседневной речи. Для примера можно привести слова *живот* и *желудок*, в разговорной речи значение у этих слов примерно одинаковое, но в специализированной лексике *желудок* – это внутренний орган человека, *живот* же как отдельное понятие отсутствует, специализированная лексика может использовать его в других контекстах, например, *острый живот*. Этот термин означает состояние, когда диагноз ещё не поставлен, но ясно, что имеются какие-то отклонения в брюшной полости. В карте памяти можно проследить различные типы классификации терминов. Классификация терминов возможна только при наличии связей или последовательности между терминами. Для темы «человек» классификация составлялась на основании Атласа анатомии человека с дополнением терминов из учебника по анатомии человека (см. приложение 1).

В работе по составлению глоссария все основные органы животных организмов, как то *сердце, лёгкие, почки* и др. описывались как органы человека, т.к. классов и видов животных организмов огромное количество. При описании пищевода земляного червя и собаки возникнут большие различия значения терминов, поэтому для облегчения понимания термина взята анатомия человека, который является по теории Дарвина самым развитым животным организмом.

Для «животного мира» такая классификация не совсем возможна из-за большого разнообразия фауны и соответственного количества повторений, а также многозначности, т.е. у разных видов животных одно и то же слово может означать несколько разные вещи. За идею по составлению словника о животном мире также взята частотность употреблений, и что самое важное, специфическое понятие, которое не встречается у других видов. Конкретизируем эту стратегию на примерах, для сравнения возьмём два «похожих» представителя класса млекопитающих: *кенгуру австралийского* и *зайца-русака*, у обоих из них есть *хвост, лапы, уши, глаза* и т.д.

Слова, отбираемые в глоссарий: *хвост, лапа* – каждое слово дано только один раз, т.е. мы не проводим различия между *хвостом* зайца и кенгуру. Слова *уши и глаза* уже включены в раздел "человек" и совпадают с общим строением, а главное с функциональным предназначением соответствующих органов осязания у человека. Наконец, имеется термин, неактуальный ни для зайца, ни для человека, а только для кенгуру и прочих сумчатых – *сумка*. Этот термин проверяется на частотность в НКРЯ в субкорпусе со снятой омонимией, частотность составляет 90, по этому критерию слово должно подходить для включения в словник, но слово является многозначным, и 99% его значений сводятся к бытовому изделию «сумка», а искомое значение «детёныш в сумке» встречается лишь несколько раз, таким образом, слово *сумка* в значении "орган" в словник включено не будет.

Частотность русских терминов проверялась в глобальной поисковой системе Google.ru, а финских в Google.fi. Порог частотности был определён по частотности употребления слов, первоначально отобранных в глоссарий. Для примера возьмём 5 слов-претендентов: *альвеола* (62300), *барабанная перепонка* (40600), *выдыхать* (66000), *ресница* (34300), *ласта* (55500). Среднее арифметическое суммы вышеупомянутых слов

составляет 51740, т.е. это есть порог частотности. Однако в глоссарий были включены несколько слов, порог частотности которых ниже 51740. Это такие слова, как *барабанная перепонка* (40600), *голосовая связка* (37600), *зигота* (38600) и *ресница* (34300). Они являются важными терминами по теме и используются в прессе.

В процессе работы составлялось два списка – русский и финский. Сразу отметим, что слова не переводились с русского, а был создан самостоятельный словник терминов на финском языке. После составления глоссария на обоих языках, выяснилось, что большинство терминов на русском языке имеет соответствия в финском.

5.1 Общая характеристика глоссария

В составленном автором глоссарии общее количество терминов составляет 282, всего в глоссарии 342 термина по теме «Организм», из которых на финском языке – 178, а на русском – 159. Небольшое различие в количествах между финским и русским списками связано с тем, что в русском языке терминов больше. Например, в русском языке *jalka* – *нога* (человека), *jalka* – *ножка* (насекомые), то есть одному финскому термину *jalka* соответствует два русских – *нога*, *ножка*.

В глоссарии создано 3 уровня сложности: элементарный, базовый и специальный. Далее представлена таблица количества терминов по уровням сложности.

таблица 2, Уровни сложности глоссария

	Количество русских терминов	Количество финских терминов	Всего
Элементарный уровень	44	37	81
Базовый уровень	36	34	70
Специальный уровень	73	72	145

Оказалось, что самым многочисленным является специальный уровень, в который вошло 145 слов. В элементарный уровень вошло 81, а в базовый – 70. Следует учитывать, что специальная лексика является менее употребляемой и более многочисленной в отличие от общеупотребительных, часто используемых слов элементарного уровня.

По частям речи деление русских терминов представлено ниже:

- существительные 140
- прилагательные 1
- глаголы 4

В финском языке кроме существительных присутствует всего лишь один глагол, что объясняется уровнем частотности при составлении глоссария.

В глоссарии по теме «Организм» значительно преобладают существительные; стоит отметить, что составляя список терминов в словник, основное внимание обращалось на порог частотности слов, существительные (объекты) используются наиболее часто, таким образом, прилагательные (свойства) и глаголы (процессы) оказались «в стороне», что, безусловно, следует учитывать в будущем при отборе слов-кандидатов.

Среди существительных встречаются *pluralia tantum*, то есть термины, не имеющие формы единственного числа. Это термины, обозначающие парные предметы, например: *жабры* (рус.) или *aivot* (финск.):

- русский язык
 - ед. и мн. ч. 137 терминов
 - только мн.ч. 5 терминов
- финский язык
 - ед. и мн. ч. 139 терминов
 - только мн.ч. 3 терминов

В результате анализа выяснилось, что часть терминов употребляется только во множественном числе, как на финском, так и на русском, но есть и исключения, например *жабры* мн.ч. – *kidukset* мн.ч., *лицо* – *kasvot* мн.ч.

5.2 Уровни сложности глоссария

В глоссарии по теме «Организм» в программе SysMLL слова распределялись по уровням сложности: «элементарный», «базовый» и «специальный». Распределение происходило на основе частотного фактора, который проверялся в поисковике Google.

Изначально распределение слов по уровням сложности было произведено по субъективному фактору, что привело к искажению материала и неверной оценке лексических единиц. Затем весь материал был еще раз основательно изучен и подобраны критерии, на основе которых все слова подразделяются на три уровня. Главная задача, стоящая перед нами, – достичь объективности, убрав личностный фактор при анализе и обработке глоссария.

Распределение слов происходило следующим образом: русские термины с частотностью употребления от 151 000 000 до 3 000 000 раз в поисковике Google.ru попадали в элементарный уровень глоссария по теме «Организм» в программе SysMLL, от 3 000 000 до 1 000 000 – в базовый и от 1 000 000 до 1 – в специальный.

Рассмотрим на конкретных примерах:

<i>Адреналин</i> (3 880 000)	}	Элементарный уровень (151 000 000–3 000 000)
<i>Беременность</i> (10 800 000)		
<i>Волос</i> (33 000 000)		
<i>Крыло</i> (3 830 000)		
<i>Нога</i> (19 600 000)		
<i>Рот</i> (22 700 000)		
<i>Сердце</i> (50 800 000)		

Висок (2 210 000)

Бровь (2 820 000)

Висок (2 210 000)

Мышца (1 110 000)

Морда (1 940 000)

Паутина (1 540 000)

Ребро (2 980 000)

Базовый уровень (3 000 000–1 000 000)

Альвеола (62300)

Антиген (268000)

Борода (320000)

Вымя (164000)

Гортань (644000)

Жабры (204000)

Зрачок (905000)

Специальный уровень (1 000 000–1)

Финские термины с частотностью употребления от 102 000 000 до 500 000 раз в поисковике Google.fi были отнесены к элементарному уровню, от 500 000 до 100 000 раз – к базовому, а от 100 000 до 1 раза – к специальному.

Рассмотрим на примерах:

Anatomia (12 900 000)

Harja (931000)

Häntä (2 230 000)

Iho (1 950 000)

Jalka (760 000)

Элементарный уровень (102 000 000 – 700000)

Räpylä (112000)

Nokka (407000)

Sormi (284000)

Kotelo (481000)

Fysiologia (135000)

Базовый уровень (400000 – 100000)

Entomologia (1780)

Häntäluu (10600)

Imusolmuke (8350)

Jalkaterä (12200)

Keuhkoputki (1990)

Специальный уровень (100000 - 1)

Чем чаще используется слово, тем легче оно для понимания. Если слово употребляется редко, значит оно не очень нужно начинающему. Лексические единицы специального уровня разучиваются в самый последний момент, так как вызывают наибольшие затруднения. Создание разноуровневой системы поможет при изучении иностранных языков, т.к упорядочивает изложенный материал и дает возможность осваивать новые лексические единицы поэтапно и качественно.

5.3 Толкование

В теоретической части работы представлены следующие виды определений: содержательное, групповое, контекстуальное, несовершенное, отрицательное, дополняющее и кольцевое. Существует намного больше видов дефиниций, см. главу 3.2, но в данной работе рассмотрены определения, наиболее подходящие для нашего вида исследования.

В SysMLL роль дефиниции описательная, т.е. прочитав её на родном или иностранном языках, пользователь должен догадаться, о каком слове идёт речь. Такие виды определений используются в загадках и кроссвордах. Они дают пользователю

представление о признаках слова, его качествах, местоположении (например, внутренние органы) и т.д.

5.4 Структура статьи

В программе SysMLL разработчики создают словники на разные темы. Например, можно изучать термины из таких разделов, как «Автомобиль», «Жизнь и земля», «В космосе», «В магазине», «История», «Математика», «Организм» и т.д.

Есть возможность вставки в базу данных графических изображений, позволяющих наглядно объяснять значение словарной единицы. Также делаются необходимые комментарии и выбирается уровень сложности. Немаловажен выбор примеров, поиск которых проводился в НКРЯ. Веб-интерфейс для ввода данных можно увидеть на скриншоте 1.

The screenshot shows a web-based form for editing an article. The form is organized into several sections:

- Topic*:** A dropdown menu set to "Организм" with an "Edit Topics" link.
- Meaning*:** A dropdown menu set to "nostril noun" with "SAVE MEANING" and "DELETE MEANING" buttons.
- Keywords:** A text input field containing "nostril noun".
- Image:** A button labeled "Image editor" and a status indicator "1 image in data".
- Word*:** A text input field containing "ноздря" with a "Find" button.
- Language*:** A dropdown menu set to "Russian".
- Definition*:** A large text area containing the text "Одно из двух наружных отверстий носа".
- Comments:** A text input field containing "ноздри мн. ч." and a placeholder "Type here a definition, using the word language.".
- Difficulty level*:** A dropdown menu set to "Intermediate".
- Buttons:** "Update", "Save as new", "Delete word", and "Add new word".
- Usage example:** A text area containing "Все уже думали о застолье, нервно ловили ноздрями запахи, долетавшие из квартир." and a dropdown menu set to "366" with a "DELETE USAGE E" button.

скриншот 1, **Окно редактора статей**

Окно редактора статей показывает, как проходит процесс создания глоссария. Автор статьи выбирает нужную тему, заносит значение слова, затем добавляет изображение (для этого в интерфейсе существует отдельное диалоговое окно), вводит слово, выбирая при этом входной язык, и дает определение к лексической единице. Также выбирается уровень сложности термина и приводится пример, разъясняющий употребление данного слова в контексте.

5.5 Ввод данных

Следующим этапом работы после составления глоссария было создание словарных статей. Структура статьи рассмотрена выше. Создание словарной статьи начинается с выбора темы (topic), в данной работе рассмотрена тема «Организм». Далее необходимо подобрать ключевое слово (keyword), которое будет описывать значение русского и финских терминов. В этой графе часто указывается и название части речи слова, а также возможное указание на вышестоящий термин, например, крестец – sacrum part of spine noun (крестец, часть позвоночника).

Следующий пункт статьи – изображение (image). Изображение должно отображать суть термина, чтобы, посмотрев на него, пользователь мог бы догадаться, о чём идёт речь. Проиллюстрируем это с помощью скриншота 2.



скриншот 2, Пояснение слова через изображение

Первое изображение ресницы отмечено стрелкой, т.к. на изображении глаза можно увидеть несколько объектов, например веко, глаз и т.д. Второй пример этой же иллюстрации показывает веко, поэтому никаких дополнительных указателей не потребовалось. Некоторые изображения приходилось менять или дорабатывать. Например, в случае с ногтем, посмотрев на рисунок, где был нарисован палец, было сложно догадаться, что речь идёт именно о ногте, поэтому последний пришлось отметить стрелкой.

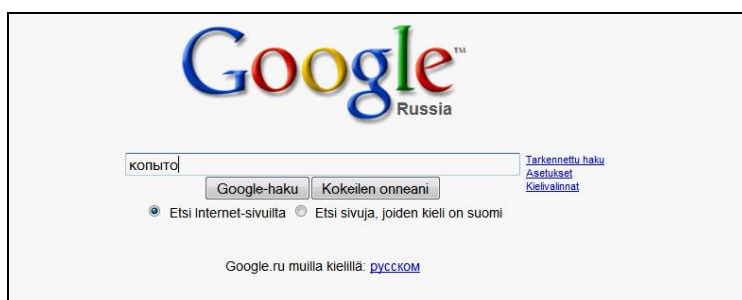
На следующей строке вводится описываемое слово (word), после чего предстоит выбор языка. В программе есть два основных языка русский и финский. При вводе важно не ошибиться с выбором языка, в противном случае программа не будет выводить на экран нужный термин. Так произошло, например, с термином *зигота*: автором по ошибке был выбран английский язык, после чего слово в упражнениях по русскому или финским языкам отсутствовало.

Зона комментариев создана для указания особенностей термина, которые могут вызвать сложности, например, нестандартным словоизменением. Например слово *лоб*: мн. *лбы*, род.п. *лба*, д.п. *лбу*, в.п. *лбы*, т.п. *лбом*, *лбами*, п.п. *лбе*, *лбах*, , в данном примере показано, что при склонении слова *лоб* появляется беглая гласная.

Уровни сложности определялись по частотности слов, о чем рассказывалось выше в отдельном разделе.

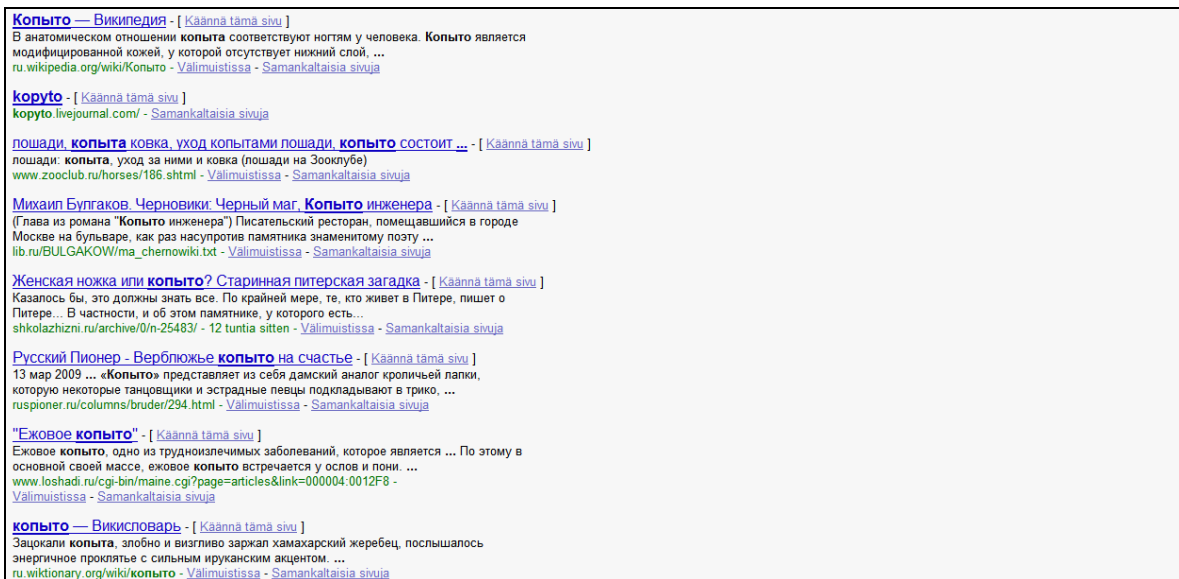
5.6 Примеры употребления

Последний этап создания словарной статьи – это ввод примера употребления термина. Пример должен быть понятный, четко разъясняющий смысл лексической единицы. Примеры употребления терминов были найдены с помощью глобальной поисковой системы Google.



скриншот 3, Поиск термина в системе Google

В графе поиска вводился термин, и затем из предложенных примеров отбирался подходящий.



скриншот 4, **Выбор примера употребления слова в поисковой системе Google**

5.6 Упражнения в SysMLL




В программе SysMLL созданы тренировочные упражнения, с помощью которых пользователь может оценить объем своего словарного запаса и разучивать новые слова.

На данный момент пользователь программы SysMLL может пользоваться следующими упражнениями:

1. Просмотр слов
2. Слова и эквиваленты
3. Слова и эквиваленты 2
4. Слова и толкования
5. Слова и толкования 2
6. Картинки и слова

Просмотр слов – это подготовительное упражнение, которое представляет вниманию пользователя выбранную им для изучения лексику (пара языков, темы, уровни

сложности). Упражнение дает возможность ознакомиться с парами слов и выражений, почитать толкования и сравнить толкования одного и того же (или похожего) понятия на разных языках. Для того, чтобы активизировать словарный запас, нужно наблюдать слова в употреблении, поэтому после изучения толкований необходимо прочитать примеры употребления слов. А графические изображения помогают избежать ошибки в понимании. Интерфейс упражнения показан на скриншоте 5.

<p>WORD: untuva</p> <p>DEFINITION: pienuotoonin pehmeä höyhen</p> <p>EXAMPLES: Pojan leuassa kasvoi jo vaaleaa untuvaa.</p>	
 <p>WORD: kantapää</p> <p>DEFINITION: Kantaluu ja sitä ympäröivien kudosten muodostama rakenne. Se on jalkaterän takaosa.</p> <p>EXAMPLES: Sitten vain otetaan kiinni kantapäästä ja viedää vastustajan kantapää mahdollisimman lähelle hänen pakaroit ja vaannetaan kasilla kantapäätä. (Japanilaisesta hallinnasta kolmikuristus jalolla.)</p>	
 <p>WORD: kytkiluu</p> <p>DEFINITION: Luu, joka suojelee sisäelimiä. Nämä luut muodostavat eräänlaisen häkin sisäelinten ympärille. Ihmisellä näyttää luuta on 12 paria.</p> <p>EXAMPLES: Jos kytkiluu murtuu kipsiä ei laiteta.</p>	
<p>WORD: связки</p> <p>DEFINITION: Плотные ткани, соединяющие кости скелета или отдельные органы. Располагаются преимущественно в области суставов, укрепляют их, ограничивают или направляют движения в суставах. (обычно -- во множественном числе)</p> <p>COMMENTS: ед. ч. связка</p> <p>EXAMPLES: На прошлой неделе Надя вывихнула ногу, в результате порвались связки.</p>	<p>WORD: nivelside</p> <p>DEFINITION: toisinsa niveltäviä luita yhdistävä kudosside</p> <p>EXAMPLES: Kun nilkka nujahtaa, kova kipu aiheutuu yleensä nivelsiteiden venähdyksestä.</p>
 <p>WORD: скелет</p> <p>DEFINITION: Совокупность костей, представляющая собой твердую основу тела животных.</p> <p>EXAMPLES: Тут сохранился необычный экспонат – скелет вымершей морской коровы.</p>	<p>WORD: luuranko</p> <p>DEFINITION: Selkärangan sisäinen tukiranka, johon kuuluvat vähintään pääkello ja selkäranka sekä lähes aina kytkiluu.</p> <p>COMMENTS: huom. luusto-kaikki luut</p> <p>EXAMPLES: Noin 65 miljoonaa vuotta vanha Triceratops-dinosauruksen luuranko tulee huutokaupttavaksi Pariisissa kesäviikolla. 7,5-metrinen luurangon arvo on arvioiden mukaan 500 000 euroa.HS.14.4.2008</p>

скриншот 5, Просмотр слов I

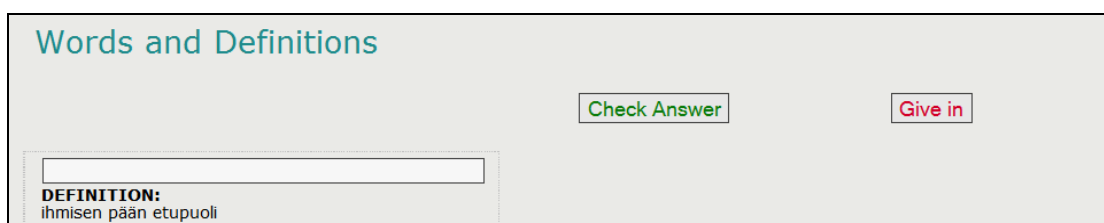
В упражнении выводятся русские и финские термины, которые являются эквивалентами, толкования, комментарии к тем лексическим единицам, при употреблении которых возникают трудности, а также примеры употребления. Также могут быть слова без соответствий, в этом случае будет только слово на входном языке.

«Слова и эквиваленты» – это упражнение, в котором пользователь должен дать словарные эквиваленты для слов на экране. В помощь пользователю на экран выводятся толкования на двух языках и картинки.



скриншот 6, Слова и эквиваленты

«Слова и толкования» – это упражнение, в котором пользователь должен вспомнить слово по его толкованию.



скриншот 7, Слова и толкования

Если термин введен правильно, на экране появляется соответствующее сообщение .



скриншот 8, «Совершенно верно!»

Если ответ неверный, то появляется надпись «Wrong!».



скриншот 9, «Неправильно!»

«Картинки и слова» - это упражнение, в котором пользователю предъявляется картинка, а он должен вспомнить соответствующее слово и ввести его в окошко для ответа. Если правильный ответ неизвестен, можно узнать его, нажав кнопку «Give in».



скриншот 10, Картинки и слова



скриншот 11, Картинки и слова II. Правильный ответ

Система ведет учет работы пользователя, считает правильные и неверные ответы, фиксирует время, которое пользователь находился в системе. Все это может использоваться как для самоконтроля, так и для оценки работы студента и его прогресса в изучении лексики.

Login time	Duration	Activity	Topics	Levels	Languages	Correct answers	No of questions
07.06.09 11:49	00:01:30	Words and definitions: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	1	1
07.06.09 11:32	00:00:00	Image and words: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	0	1
31.05.09 17:04	00:00:00	Image and words: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	0	1
31.05.09 16:43	00:00:26	Words and definitions: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	2	2
31.05.09 16:12	00:28:05	Image and words: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	0	7
29.05.09 15:49	00:24:29	Image and words: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	9	12
29.05.09 14:36	01:13:10	Words and definitions: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	12	21
29.05.09 13:38	00:08:50	Words and definitions: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	9	14
29.05.09 13:00	00:38:25	Words and equivalents: exercises	Организм	Easy, Medium, Hard	Finnish, Russian	56	59

скриншот 12, Рабочий журнал

Информация об ошибках пользователя сохраняется в базе данных, что позволяет системе повторно задавать вопросы, на которые студент не смог ответить. Кроме того, пользователь может получить список своих ошибок и работать с этой лексикой целенаправленно.

Word	Correct answers	Wrong Answers	Total
silmäluomi	3	10	13
lainhuudatus	1	9	10
aortta	5	9	14
silmäkulma	1	8	9
nivelside	4	7	11
vartalo	6	7	13
jalkaterä	3	7	10
käsivarsi	3	7	10
entsyymi	0	6	6
veren paine	1	5	6
kohta	0	4	4
irtaimet esineet	0	4	4
etuosake	1	4	5
tytär yhteisö	0	4	4
talonpoikainen	1	4	5
yhtiömuodon muuttaminen	0	4	4
talonpoikaisnainen	0	4	4
eukko	0	4	4
aviopuoliso	0	4	4
lonkka	3	4	7
selkäydin	0	4	4
sääri	1	4	5
kilpi	0	4	4
kärsä	1	4	5
kiinteistörekisteri	0	3	3
näyttelijä	0	3	3
suppea esituskinta	0	3	3

скриншот 13, Отчет об ошибках

Таким образом, у пользователя есть возможность улучшать свои языковые знания.

5.7 Интроспекция

В ходе проверки упражнений в программе SysMLL были замечены некоторые недочеты, которые затем подверглись корректировке.

Трудность возникла при выполнении упражнения "Слова и толкования 2", где появилось большое количество клеток для ввода ответа. Запрашивался термин *sakset (ravun kahteen pihtimäiseen osaan päättyvät etujalat)*. Но этого ответа было недостаточно. Разгадка заключалась в том, что данный термин был введен в глоссарий в составной форме *sakset (ravun)*, однако не каждый пользователь может наглядно запомнить начальную форму введенного термина. Но без уточнения (*ravun*) ответ не засчитывался. Недочет был исправлен по средствам удаления уточнения *ravun*.

Также была найдена ошибка в слове *aortta*: в глоссарии данная лексическая единица занесена с опечаткой (*aorttta*), таким образом, правильный вариант не принимался. Данный недочет исправлен администратором.

Автор глоссария протестировал упражнения по теме «Организм», были получены следующие результаты. Общее время тестирования – 3 часа 50 минут. Остальное время было затрачено на создание словарных статей и корректировку внесенного материала. Далее рассмотрим подробнее процесс тестирования на примере «элементарного уровня».

На первом этапе слова изучались на русском языке, который является для автора родным. Работа с программой начинается с раздела Просмотр слов, время работы в этом разделе не фиксируется. На одной странице изображается по 5 слов с дефиницей, комментариями, примером использования слова, а также картинкой, если таковая имеется. Всего по теме 68 страниц в этом разделе. Следующим идёт раздел Слова и эквиваленты. За 20 минут, проработалось 75 вопросов, из которых был дан 71 правильный ответ и 4 неверных ответа. После допущенных ошибок программа повторила слово *мех* 7 раз, из которых 3 раза по 3 слова подряд и 1 раз 2 раза подряд. В слове *рука* ошибок допущено во время данного сеанса не было, но программа

повторила его 4 раза. Таким образом, из-за частого повторения одного и того же слова у пользователя пропадает желание работать с программой, этот недочёт в создании программы нуждается в корректировании.

На втором этапе изучались финские слова. В Просмотре слов было изучено 64 термина. Правильных ответов было дано 69, а неправильных 3. Слово *karva* повторилось 17 раз, из которых 4 раза подряд. Слово *sarvi* 2 раза, слово *varvas* 8 раз, *hiki* 4 раза. Примечателен и тот факт, что не одно из вышеупомянутых повторений не входит в список типичных ошибок.

Далее работа продолжалась в разделе «Слова и толкования». Здесь были выявлены недочёты в определениях, т.е. прочитав определение пользователю сложно догадаться о каком слове идёт речь. К примеру финское слово *suoli* было описано как «*guoansulatuskanavan osa*», т.е. по-русски это будет звучать, как «орган системы пищеварения», по картинке тоже трудно догадаться, т.к. на ней изображена вся система пищеварения. Решено было изменить определение на более понятное добавив слово самый длинный орган системы пищеварения: «*pisin guoansulatuskanavan osa*». Таким же образом были исправлены и другие словарные статьи глоссария. Повторения слов были такие же, как и в предыдущем разделе, например, в слове «*kieli*» не было допущено ни одной ошибки, но повторялось оно 3 раза.

В целом, изучив внимательно термины, дефиниции к ним, пользователь может без труда выполнить все задания в программе SysMLL, тем самым пополнив свой словарный запас.

Заключение

Задачей данной работы было составление глоссария по теме «Организм» в лексическом тренажёре SysMLL, разработанном в Тамперском университете для изучения иностранных языков и в качестве вспомогательного инструмента в процессе обучения переводчиков. Основной целью работы было создание словарных статей на основе разработанного словника.

Работа началась с изучения теоретических аспектов. Создание словарных статей основано на лексикографии, науке, занимающейся составлением словарей. На сегодняшний день большая часть лексикографической теории основана на создании традиционных, так называемых бумажных словарей. Некоторые аспекты, такие как общие типы словарей, а также активный и пассивный словари, было необходимо рассмотреть в данном исследовании для определения концепции работы.

Анализ строения частотных словарей положил основу для создания словника, отбор слов в который производился на основе частотного фактора. Предварительная выборка слов производилась из учебников по биологии и анатомии человека, а затем частотность использования слов проверялась в НКРЯ (Национальном корпусе русского языка) для словника на русском языке и поисковой системе Google для словника на финском языке.

После составления первоначального словника был определен порог частотности слов, который устанавливал ограничения при вводе лексической единицы в глоссарий. Термины, соответствующие установленным рамкам, заносились в глоссарий. Ключевые слова, определение лексической единицы, уровень сложности, примеры, комментарии, иллюстрации – основные пункты, на которые необходимо было обратить внимание при обработке словника.

В ходе работы выяснилось, что в глоссарии по теме «Организм» практически у всех русских терминов есть финские эквиваленты и наоборот. Известно, что существует безэквивалентная лексика, что объясняется различиями в культуре и менталитете народов. Однако в биологической и анатомической терминологии, областях,

изучающих строение живых организмов и анатомию человека, как правило, можно найти соответствия терминов на обоих изучаемых языках. Например, *нога* – *jalka*, *печень* – *maksa*, а *гемоглобин* – *hemoglobiini*. Бывают разговорные варианты лексических единиц, типа *печенка*, *бородка*, от терминов *печень* и *борода*, однако, они не имеют отдельного финского эквивалента (*печень*, *печенка* – *maksa*).

Практической частью работы было самотестирование по теме «Организм», каждый уровень сложности программы отрабатывался в отдельности. В результате тестирования, были выявлены следующие недочёты глоссария:

- неточность дефиниции, т.е. прочитав дефиницию пользователь не мог с точностью определить искомое слово
- некорректный пример использования, слово, показанное в примере, охарактеризовано недостаточно понятно
- в графе "комментарии" были не всегда указаны особенности склонения, форм слова и др. необходимая информация
- ключевое слово подобрано некорректно, без учёта совместимости терминов на обоих языках

Выявленные недочёты в словарных статьях подверглись корректировке и дополнению совместно с администратором программы.

Данная работа является первым в своём роде исследованием программы SysMLL и является отправной точкой для новых исследований в этой области. Пользуясь методиками, описанными в данной работе, в дальнейшем необходимо расширить остальные темы в программе до уровня полноценных глоссариев. Немаловажно исследование эффективности работы программы на респондентах. Это даст возможность выявления возможных ошибок в составлении словарных статей, а также позволит выявить лакуны в базе данных программы. Необходимо продолжить оценку релевантности определения уровней сложности на основании частотности. Возможно получится выявить и другой принцип разделения терминов, опираясь на знания студентов. Благодаря вышесказанному тема «Организм» и вся программа SysMLL будет доведена до совершенства и её можно будет применять в учебном процессе.

Библиография

1. Haarala R. Sanastotyön opas HKI: KOTUS, 1981.
2. Hart, Andrew. Media education in 12 European countries. A comparative study of teaching media in mother tongue education in secondary schools. Swiss Federal Institute of Technology, 2002.
3. Novak, Joseph D. & Gowin, D. Bob 1995. Opi oppimaan. Helsinki: Gaudeamus.
4. Ooi Vincent B.Y. Computer Corpus Lexicography, Edinburgh: Edinburgh University Press 1998.
5. Алексеев П.М. Методика квантитативной типологии текста: Учебное пособие. Л: ЛГПИ, 1983.
6. Алексеев П.М. Частотные словари, СПб: СПбГУ, 2001.
7. Балыхина Т.М., Гарцов А.Д. Информатизация обучения языкам: становление компьютерной лингводидактики, Higher Education Today М:2006/4.
8. Баранов А.Н. Введение в прикладную лингвистику, М: Эдиториал УРСС, 2001.
9. Берков В.П. Двухязычная лексикография, СПб: СПбГУ, 1996.
10. Бовтенко М.А. Компьютерная лингводидактика: Учеб. пособие. М: Просвещение, 2005.
11. Гринев С. В. Введение в терминоведение, М.: Московский Лицей, 1993.
12. Кудашев И.С., Кудашева И.О. Полезные дополнения к традиционной практике составления переводных терминологических словарей. // Конференция Диалог 21, 2008. <http://www.dialog-21.ru/dialog2008/materials/html/42.htm>.
13. Писаренко В.И. Использование компьютерных технологий в обучении иностранным языкам, Открытое образование //М: МЭСИ, МАОО 2004/1.
14. Попов А.А. Применение новых информационных технологий для представления многоязычного словаря лексической сочетаемости М.2002.
15. Потапова Р.К. Новые информационные технологии и лингвистика. – М.: МГЛУ, 2002.
16. Суперанская А.В., Подольская Н.В., Васильева Н.В. Общая терминология и вопросы теории. М., 1989.

17. Табанакова В.Д. Логико-понятийный анализ терминологии и его прагматика. // Интегративный анализ компонентов текста при обучении общению. – Тюмень: Изд-во ТюмГУ, 1998, с. 5-22. <http://frgf.utmn.ru/last/No1/tabanoqv.html>.
18. Черняк В.Д., Русский язык и культура речи. М.: Высшая школа, 2002.
19. Якубайтис Э.Я. Проблемы обучения терминологии, Открытое образование М.: МЭСИ, МАОО 2004/1.
20. Ялаева Н.В. Компьютерное тестирование как средство интенсификации обучения английскому языку в юридическом вузе. Диссертация на соискание учёной степени кандидата педагогических наук. Екатеринбург: УГПУ, 2004.

Источники исследовательского материала

1. Leinonen M., Nyberg T. Biologia: Ihminen НКІ: WSOY, 2007.
2. Valste J.A., Holopainen S. Biologia: Elämä НКІ: WSOY, 2002.
3. Драгомилов А. Г., Маш Р.Д. Биология: Человек. Учебник для 8 кл. общеобразовательных учреждений. М.: Вентана-Граф, 2006.
4. Константинов В.М., Бабенко В.Г., Кучменко В.С. Биология: Животные. Учебник для 7 кл. общеобразовательных учреждений. М.: Вентана-Граф, 2006.
5. Самусев Р. Атлас анатомии человека М: Диля, 2005.

Справочные материалы

1. Nienstedt W., Kellosalo J. Lääketieteen termit. Duodecim 2007
2. Ахманова О. С. Словарь лингвистических терминов, М.: Советская Энциклопедия, 1968.
3. Большая Российская энциклопедия. М: Советская энциклопедия, 1978. <http://slovari.yandex.ru/dict/bse>
4. Ефремова Т. Ф. Новый словарь русского языка, М.: Русский язык, 2000.
5. Краулов Ю. Энциклопедия «Русский язык», М: БРЭ, 2003
6. Коппалёва Ю.Э. Новый финско-русский словарь, СПб: МГВ, 2001.
7. Кузнецов С.А. Современный толковый словарь русского языка. СПб: Норит, 2001.
8. Куусинен М.Э. и др. Новый большой русско-финский словарь; М: Живой язык, 2002.

9. Лингвистический энциклопедический словарь, М: Директмедиа, 2008.
10. Литературная энциклопедия, <http://slovari.yandex.ru/dict/litenc>
11. НКРЯ: Национальный корпус русского языка <http://ruscorpora.ru/>
12. Ожегов С.И. Толковый словарь русского языка, М: Оникс, 2009
13. Смирницкий, А.И. (ред.). Большой русско-английский словарь. . М.: Русский язык, 2002. <http://slovari.yandex.ru/dict>
14. Смирницкий А.И. (ред.). Русско-английский словарь. М.: Русский язык, 1992
15. Шемякина А.В. Компьютерный переводной словарь Мультилекс 3.5, М., 2002. Русский Язык. Энциклопедия. М: Дрофа, 1997.

Интернет-источники

1. АBBYY Lingvo: <http://www.lingvo.ru/multilingual/>
2. Рыков В.В., Компьютерная лексикография. Диалог, международная конференция 2002 год <http://www.dialog-21.ru>
3. Hot Potatoes: <http://web.uvic.ca/hrd/halfbaked/>.
4. KVocTrain: <http://edu.kde.org/kvoctrain/>.
5. Open Book: <http://www.vinidiktov.ru/openbook.htm>.
6. Vtrain: <http://www.paul-raedle.de/vtrain/home-ru.htm>.
7. Евдокимова М.Г. http://sputnik.mto.ru/biblioteka/VIO/VIO_01/Present/ITO/1998-99/d/evdokimova.html
8. Информация по программе Lex! и её описание. <http://www.yvisoft.net/lex/faq.php>
9. Корпусная лингвистика <http://corpora.iling.spb.ru/theory.htm>.
10. Мир словарей, Психологическая энциклопедия
http://mirslovari.com/content_psy
11. Научно-исследовательский вычислительный центр МГУ: Когнитивные аспекты лексикографии <http://www.infolex.ru/>
12. Русский языкъ http://www.teneta.ru/rus/rj_ogl.htm
13. Конторович Е.А., Шеляхина Н.В. Компьютерные технологии как необходимая составляющая при совершенствовании обучения иностранным языкам специалистов в техническом вузе.

http://www.sgu.ru/faculties/physical/departments/it-physics/international2007/docs/Kontorovich_E.A._Shelyahina_N.V.doc.

14. Tampereen yliopiston humanistisen tiedekunnan opinto-opas, TRE, 2008 (http://www.uta.fi/opiskelu2/o-opas/hum/hum-opas_137-212.pdf).
15. Михайлов М.Н. Структура и содержание лексических баз данных для обучающей программы по иностранным языкам. // Конференция Диалог 2006. http://www.dialog-21.ru/dialog2006/materials/html/Mikhailov.htm#_ftn1.
16. Свободная энциклопедия Википедия <http://ru.wikipedia.org/>
17. Селегей В. Электронные словари и компьютерная лексикография, 1996 http://www.lingvoda.ru/transforum/articles/selegey_a1.asp

Приложения

Приложение 1

Список слов для глоссария на русском языке:

1. адреналин (3 880 000)¹
2. альвеола (62300)
3. анатомия (2 600 000)
4. антиген (268000)
5. антитело (812000)
6. аорта (376000)
7. аппендикс (99700)
8. артериальное кровяное давление (190000)
9. артерия (805000)
10. барабанная перепонка (40600)
11. бедро (3 610 000)
12. беременная (3 230 000)
13. беременность (10 800 000)
14. борода (320000)
15. бородка (318 000)
16. бровь (2 820 000)
17. бронхи (512000)
18. брюшко разг. (386000)
19. вдох (2 010 000)
20. вдыхать (353000)
21. вздыхать (90000)
22. веко (1 240 000)
23. вена (7 130 000)
24. венозный (266000)
25. висок (2 210 000)
26. волос (33 000 000)
27. выдыхать (66000)
28. выдох (1 670 000)
29. вымя (164000)
30. гемоглобин (709000)
31. голень (907000)
32. головной мозг (683000)
33. голосовая связка (37600)
34. гортань (644000)
35. грива (756000)
36. грудь (1 980 000)
37. губа (2 080 000)
38. десна (934000)
39. дыхание (11 500 000)
40. дыхательная система (308000)
41. дышать (1 860 000)
42. жабры (204000)
43. жало (732000)

¹ частотность использования слова в глобальной поисковой системе Google

44. железа (8 810 000)
45. желудок (5 770 000)
46. желудочек (450000)
47. желчь (511000)
48. желчный пузырь (288000)
49. живот (19 100 000)
50. запястье (923000)
51. зародыш (422000)
52. зигота (38600)
53. зоология (1 520 000)
54. зрачок (905000)
55. зуб (11 400 000)
56. ихтиология (57100)
57. кисть руки (1 230 000)
58. кишечник (1 790 000)
59. кишка (684000)
60. клешня (113000)
61. клюв (511000)
62. ключица (368000)
63. коготь (232000)
64. кожа (19 200 000)
65. конечность (1 440 000)
66. копчик (133000)
67. копыто (1 160 000)
68. костно-мышечная система (96500)
69. кость (6 660 000)
70. крестец (178000)
71. кровь (39 700 000)
72. кровообращение (1 360 000)
73. крыло (3 830 000)
74. куколка (568000)
75. лапа (3 920 000)
76. лапа (55500)
77. лёгкие (7 380 000)
78. лимфа (286000)
79. лимфатическая система (138000)
80. лимфатический узел (70700)
81. личинка (897000)
82. лопатка (1 520 000)
83. матка (2 140 000)
84. мех (6 620 000)
85. младенец (3 120 000)
86. мозжечок (197000)
87. морда (1 940 000)
88. мышца (1 110 000)
89. нёбо (31 300 000)
90. нерв (1 850 000)
91. нервная система (2 050 000)
92. новорожденный (2 410 000)
93. нога (19 600 000)
94. ноготок (61900)

95. ноготь (3 410 000)	120. рот (22 700 000)
96. ноздря (250000)	121. связки (1 110 000)
97. обмен веществ (1 400 000)	122. селезенка (280000)
98. орнитология (111 000)	123. сердце (50 800 000)
99. палеонтология (402000)	124. сердечно-сосудистая система (966000)
100. палец (15 900 000)	125. скелет (2 330 000)
101. панцирь (647000)	126. сосуд (1 930 000)
102. паутина (1 540 000)	127. спина (13 500 000)
103. перо (6 070 000)	128. спинной мозг (208000)
104. печёнка (306 000)	129. стопа (1 230 000)
105. печень (6 500 000)	130. сустав (2 150 000)
106. пищеварительная система(305000)	131. тело (32 200 000)
107. плечо (13 200 000)	132. туловище (867000)
108. позвоночник (4 300 000)	133. усы (1 670 000)
109. почка (1 610 000)	134. ус (4 550 000)
110. пот (3 940 000)	135. ухо (15 000 000)
111. потовая железа (64300)	136. фермент (878000)
112. предсердие (67300)	137. физиология (3 290 000)
113. присоска (207000)	138. хвост (7 270 000)
114. проток (550000)	139. хобот (495000)
115. пух (4 230 000)	140. череп (4 930 000)
116. пятка (1 730 000)	141. челюсть (2 780 000)
117. ребро (2 980 000)	142. чешуя (622000)
118. ресница (34300)	143. шерсть (3 100 000)
119. рог (10 900 000)	144. шея (5 630 000)

145. шкура (1 490 000)
146. щека (2 450 000)
147. щиколотка (116000)
148. щупальца (573000)
149. эндокринная система (191000)
150. энтомология (177000)

1. adrenaliini (67200)
2. aikuinen (357000)
3. aineenvaihdunta (46500)
4. aivot (559000)
5. anatomia (12 900 000)
6. antigeeni (47000)
7. eläintiede (89300)
8. entomologia (1780)
9. entsyymi (22500)
10. eteinen (127000)
11. fysiologia (135000)
12. hammas (378000)
13. harja (931000)
14. hemoglobiini (49200)
15. hengitys(264000)
16. hengittää (412000)
17. hengityselimistö (4040)
18. hermo (723000)

151. язык (151 000 000)
152. яйцеклетка (559000)
153. яйцо (16 500 000)

Приложение 2.

Список слов для глоссария на финском языке:

19. hermosto (83200)
20. hiki (2 940 000)
21. hikirauhanen (807)
22. hius (161000)
23. huokonen (3080)
24. huuli (533000)
25. häntä (2 230 000)
26. häntäluu (10600)
27. ien (12 600 000)
28. iho (1 950 000)
29. iktyologia (357)
30. imuneste (3410)
31. imusolmuke (8350)
32. imusuonisto (2240)
33. jalka (760000)
34. jalkaterä (12200)
35. jänne (78400)
36. kantapäätä (59400)

37. karva (567000)
38. kaula (787000)
39. kavio (134000)
40. keuhkoputki (1990)
41. keuhkot (103000)
42. keuhkorakkula (1910)
43. kidukset (4980)
44. kieli (7 700 000)
45. kilpi (509000)
46. kitalaki (8400)
47. kohtu (2 320 000)
48. korva (2 700 000)
49. kotelo (481000)
50. kulmakarva (32700)
51. kuono (150000)
52. kurkunpää (12500)
53. kylkiluu (13300)
54. kynsi (51200)
55. käpälä (430000)
56. kärsä (841000)
57. käsi (29 100 000)
58. lapaluu (4580)
59. laskimo (16200)
60. leuka (208000)
61. lihas (264000)
62. lintutiede (2730)
63. lonkka (108000)
64. luu (102 000 000)
65. luuranko (78100)
66. mahalaukku (47500)
67. maksa (7 7800 000)
68. muna (14 200 000)
69. munasolu (20100)
70. munuainen (21200)
71. nilkka (65500)
72. niska (7 520 000)
73. nivel (114000)
74. nivelside (3070)
75. nokka (407000)
76. ohimo (8190)
77. olkapää (67900)
78. paleontologia (9570)
79. parta (5 380 000)
80. perna (4 750 000)
81. pikkuaivot (5560)
82. pistin (318000)
83. poski (126000)
84. pää (1 910 000)
85. pääkallo (181000)
86. päävaltimo (200)
87. raaja (361000)
88. ranne (715000)

89. raskaus (427000)	115.suoli (2 000 000)
90. rauhanen (59800)	116.suolisto (36600)
91. rinta (547000)	117.suomu (51900)
92. ripsi (73100)	118.suoni (3 820 000)
93. ristiluu (2090)	119.suu (10 600 000)
94. ruoansulatuselimistö (3160)	120.sydän (2 250 000)
95. ruumis (1 050 000)	121.sydänkammio (973)
96. räpylä (112000)	122.sääri (2 310 000)
97. sakset (2140)	123.tiehyt (7100)
98. sappi (1 040 000)	124.tiineys (8760)
99. sappirakko (9620)	125.toukka (277000)
100.sarvi (152000)	126.tsygootti (22300)
101.seitti (28700)	127.tuntosarvi (3340)
102.selkä (556000)	128.turkki (1 590 000)
103.selkäranka (45000)	129.tärykalvo (4650)
104.selkäydin (16100)	130.uloshengitys (34800)
105.sierain (11800)	131.umpilisäke (3410)
106.siipi (170000)	132.untuva (97200)
107.sikiö (132000)	133.utare (128000)
108.silmä (526000)	134.valtimo (211000)
109.silmäluomi (3040)	135.vartalo (95700)
110.silmäterä (32900)	136.varvas (409000)
111.sisäänhengitys (8740)	137.vasta-aine (192000)
112.solisluu (13200)	138.vatsa (373000)
113.sormi (284000)	139.vauva (1 080 000)
114.sulka (231000)	140.verenkierto (78700)

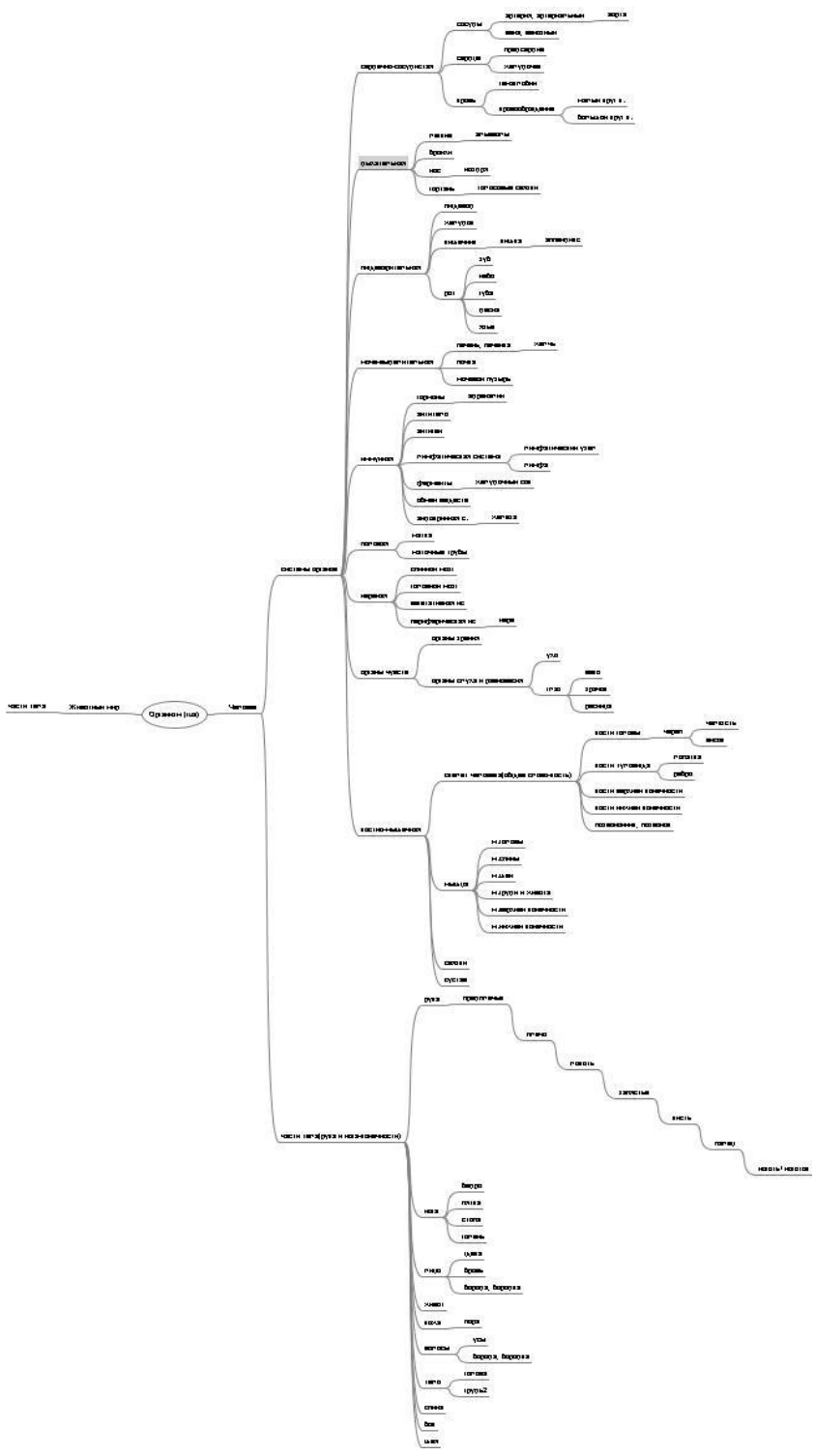
141.verenkiertoelimistö (8220)

143.äänijänne (235)

142.verisuoni (14800)

Приложение 3

Ассоциативная карта



Приложение 4

Сводка ошибок: самотестирование на введённом материале

Слово	Правильные ответы	Неверные ответы	Всего
silmäluomi	3	10	13
aortta	5	9	14
silmäkulma	1	8	9
nivelside	4	7	11
vartalo	6	7	13
jalkaterä	3	7	10
käsivarsi	3	7	10
veren paine	1	5	6
entsyymi	0	6	6
kohtu	1	2	3
lonkka	3	4	7
selkäydin	0	4	4
sääri	1	4	5
kilpi	0	4	4
kärsä	1	4	5
verisuoni	1	3	4
kulmakarva	3	3	6
suoli	0	3	3
kitalaki	1	3	4

ilme	2	3	5
lintutiede	1	3	4
hengittää	1	3	4
sierain	4	3	7
seitti	0	2	2
imuneste	3	2	5
keuhkorakkula	0	2	2
kotelo	1	2	3
ruoansulatuselimistö	2	2	4
tärykalvo	5	2	7
nilkka	4	2	6
eläintiede	0	2	2
antigeeni	3	2	5
mahalaukku	1	2	3
anatomia	1	2	3
sakset	0	2	2
альвеола	1	2	3
связки	0	2	2
гортань	1	2	3
кишка	0	2	2

huuli	0	1	1
leuka	0	1	1
otsa	1	1	2
ranne	2	1	3
eteinen	1	1	2
sappirakko	1	1	2
käsi	3	1	4
hikirauhanen	1	1	2
munuainen	0	1	1
utare	1	1	2
kavio	0	1	1
aikuinen	0	1	1
lapaluu	0	1	1
kasvot	1	1	2
желчный пузырь	1	1	2
keuhkoputki	1	1	2
kantapää	1	1	2
nivel	0	1	1
корпус	0	1	1
тело	0	1	1

туловище	0	1	1
физиология	1	1	2
нёбо	1	1	2
дыхательная система	2	1	3

Lyhennelmä

Verkkopohjaisen opetusohjelman sanaston laatiminen, esimerkkitapauksena aihealue ”Elimistö”

Pro gradu – tutkielma

Tampereen yliopisto 2009

Juhani Zarianov

Johdanto

Nykyään tietotekniikkaa käytetään monella ammattialueella, myös apuvälineenä opetuksessa. Tämän tutkielman teemana on tietokonepohjaisten kieltenopetusohjelmien sanastokomponentin kehittäminen. Tutkielmatyössä on täydennetty Tampereen yliopiston opetusohjelmaa, SysMLL:ää, varten koostettua elimistö-aiheista sanastoa. Sanaston pohjalta ohjelmaan luotiin sanastoartikkeleita, joita on testattu verkko-ohjelman harjoituksilla.

Tutkimuksen tavoitteena on ollut elimistö-aiheisen sanaston koostaminen ja sen muokkaaminen käytettäväksi tarpeen mukaan itseopiskelussa Tampereen yliopiston kieli- ja käännöstieteiden laitoksella.

Pro gradu -tutkielmassa on perehdytty tietokoneohjelmien sanaston laatimisen eri vaiheisiin: karkea sanajako, sanaston muokkaaminen taajuuksia määrittämällä sekä lopullisen sanaston täydentäminen.

Työssä tarkastellaan ja vertaillaan verkkosanakirjojen ja perinteisten painettujen sanakirjojen eroavaisuuksia. Empiirisen tutkimusaineiston avulla on pyritty selvittämään, millaisia sanoja on lisättävä sanastoon tai poistettava siitä.

Tutkimuksen tuloksena on tuotettu aiheenmukainen sanasto, jota kuvaavat sanakirja-artikkelit on syötetty SysMLL-ohjelmaan. Elimistö-aihe on lopuksi testattu harjoitusten avulla. Niiden pohjalta sanastoa on pystytty kehittämään niin, että se soveltuu käytettäväksi tarkoitukseensa.

Tehdyn tutkimustyön pohjalta voidaan myös vaivattomasti koostaa ja kehittää muita aihealueita SysMLL-verkko-ohjelmassa.

1. Sanakirjat

Käännöstieteessä puhutaan lähdetekstistä ja kohdetekstistä. Lähdetekstillä tarkoitetaan tekstiä, joka jotakin tarkoitusta ja kohderyhmää varten muotoillaan (käännetään) jollekin toiselle kielelle ja kohdeteksti on teksti, joka tämän prosessin (kääntämisen) tuloksena syntyy eli käännös. Sanakirjat voidaan kielten lukumäärän mukaan luokitella yksikielisiksi, kaksikielisiksi tai monikielisiksi. Yksikielinen sanakirja selittää hakusanan tarkoittamat käsitteet samalla, hakusanojen kielellä. Kaksikielinen sanakirja pyrkii löytämään merkitykseltään tai käyttötavoiltaan toisiaan vastaavia sanoja ja ilmaisuja kahden kielen välillä. Monikielinen sanakirja on useimmiten jonkin erityisalan sanasto, esimerkiksi lääketieteellinen sanakirja. (Berkov 1996: 7)

Taajuus- eli frekvenssisanakirja edustaa sanakirjoissa omaa lajiaan. Siinä sanat on järjestetty niiden taajuuden mukaan eli frekvenssisanakirjan avulla voidaan tarkistaa sanojen käytettävyyttä. Frekvenssisanakirjoja laaditaan jonkin määritellyn tekstikorpuksen pohjalta. Tekstikorpus on tekstien kokoelma. Korpushaulla tarkistetaan, kuinka monessa tekstissä ja kuinka monta kertaa kukin sanaehdokas esiintyy, toisin sanoen sanan jakauma ja lukumäärä eri teksteissä ja sen kokonaistaajuus. Tekstikorpusta voidaan hyödyntää sanaston laatimisessa: kun karkea erottelu on tehty, tarkistetaan, miten sanoja käytetään kielessä.

2. Sähköinen leksikografia

Nykyisin lähes kaikkien toimintojen olennaisena osana on tietotekniikan käyttö. Tietokoneesta on tullut välttämätön työväline. Sähköiset sanakirjat ovat yleensä kätevämpiä käyttää kuin perinteiset paperimuotoiset sanakirjat. Lisäksi verkkosanakirjojen käyttö on vaivatonta niiden joustavuuden ansiosta. Sanakirjan tietokantaa voidaan jatkuvasti muokata ja muutokset näkyvät heti eikä seuraavassa painoksessa kuten painetuissa sanakirjoissa. Nykyajan sanakirjaohjelma pystyy automaattisesti muokkaamaan yhden kaavan perusteella koko sanakirjan asua. Sähköisten sanakirjojen laatimisessa painottuu tiimityötapaa. Moniammatillinen ja -tieteinen yhteistyö tuottaakin yleensä kattavan tuloksen.

Sähköisessä sanakirjassa ei ole tarvetta laittaa sanakirja-artikkeleita aakkosjärjestykseen, vaan sana haetaan tietokannasta automaattisesti. Sanat järjestyvät automaattisesti, ja haku on mahdollista rajata esimerkiksi jonkin tietyn tekijän tai aiheen mukaan.

Sähköisen sanaston laatiminen eroaa perinteisestä leksikografisesta menetelmästä. Sähköisessä leksikografiassa sanakirja-artikkelissa käytetään ns. "ensimerkintää", jonka seurauksena kokotekstihaku voi erottaa määritelmän kommentteista ja esimerkeistä.

Kuvausyksikkönä toimii leksikaalinen tarkoite lekseemin sijasta, minkä perusteella tietokone pystyy yhdistämään asiakkaan kyselyn ja oman tietokannan leksikaalisen tarkoitteen, jotka ovat relevantteja syntaktisten ja semanttisten kriteerien mukaisesti. (Selegei 1998)

2.1. Tekstikorpuksat kielentutkimuksessa ja kääntämisessä

Korpuslingvistiikka tutkii sähköiseen muotoon laadittuja kirjallisen ja puhutun kielen aineistoja. Näin korpuksia voidaan käsitellä tietotekniikan avulla. Viime vuosikymmeninä sähköisiä korpuksia on alettu käyttää apuna kielentutkimuksessa. Korpusten käyttö on mahdollistanut aivan uudenlaisen näkökulman kielen- ja kääntämisentutkimukseen. (Eskola 2002: 22–25)

Korpuksat ovat yksi- tai monikielisiä ja ne sisältävät alkuperäisiä tekstejä tai käännöksiä. Tutkimuskohteista riippuen voidaan soveltaa eri tavalla koostettuja korpuksia. Käännöstutkimuksessa voidaan käyttää käännöskorpusta, joka koostuu lähdeteksteistä ja niiden käännöksistä.

Korpuksia voidaan soveltaa kääntämistaidon koulutuksessa. Käyttämällä korpusta opiskelijat parantavat kielitaitoaan sekä saavat monipuolisia ratkaisuja käännösongelmiin. Korpuksista selviää kielen käyttöön liittyviä seikkoja. Tästä johtuen korpuksesta voidaan etsiä sanojen käyttöesimerkkejä. (Eskola 2002: 31–32)

2.2. Tietokoneohjelmat vieraiden kielten opiskelussa

Vieraiden kielten opetuksessa on käytössä useita erilaisia tietokoneohjelmia. Niiden tutkiminen on kuitenkin vaikeaa, koska suurin osa on suojattu käyttöoikeudella ja ohjelmiin tutustuminen edellyttää niiden ostamista. Lehtori Mihail Mihailov on artikkelissaan jakanut kaikki opetussovellukset seuraaviin ryhmiin:

- § Tyhjät mallit työsanastoja tai harjoituksia varten
- § Kielitaitotestit
- § Kielenopetuspelit

3. Terminologia

Terminologia on tiede, joka tutkii erikoiskielen käsitteitä, termejä.

Tässä luvussa on tutkittu menetelmiä, joiden avulla voidaan etsiä sopivia määritelmiä sanojen kuvaukseen sekä sopivia käytäntöjä sanaston laatimiseen.

Ajatuskartta (mind map) on muistiinpanotekniikka, jonka avulla voidaan myös rakentaa tietyn aiheen sanasto. Työn aluksi on etsittävä avainsanoja, joihin liitetään niihin kuuluvia termejä. Rakenteeltaan ajatuskartta voi olla monenlainen, esim. puumainen, jolloin yhdestä ”rungosta” haarautuu ”oksia”.

4. SysMLL

SysMLL on verkko-ohjelmisto sanaston opiskelua varten. Se on kehitetty Tampereen yliopiston käännöstieteen yksikössä venäjän oppiaineessa. Mukana työssä olivat myös yliopiston Kielikeskus ja Tietokonekeskus. Nykyisessä tietokannassa on venäjän ja suomen kielen sanastoja. Fyysisesti SysMLL sijaitsee Tampereen yliopiston web-palvelimella.

SysMLL-ohjelma eroaa muista tarjolla olevista ohjelmista siten, ettei se ole kokoelma valmiita harjoituksia vaan se tuottaa harjoitustehtäviä tiettyjen kaavojen mukaisesti. Sen ansiosta harjoittelua voi jatkaa ohjelmalla rajattomasti oman tarpeen mukaan. Jokaisella ohjelman käyttäjällä on oma käyttäjätunnus. Opiskelija pystyy itse säätämään asetuksia omien tavoitteidensa mukaisesti. Esimerkiksi Lex!-sanasto-ohjelmassa käytetään aina samanlaisia harjoituksia, joiden rakenteen opiskelija oppii tuntemaan pian ja sen jälkeen opetuksesta ei ole vastaavaa hyötyä. SysMLL-ohjelma pystyy myös säätämään harjoituksia opiskelijan äidinkielen mukaan.

Työskentelyn alussa opiskelija valitsee parhaiten hallitsemansa kielen ja opiskeltavan kielen. Sen jälkeen hän valitsee opiskeltavan aiheen. Tällä hetkellä ohjelmassa on 56 aihetta, mm. autoilu, rakentaminen, laki, tiede, elokuvat, talous ja monia muita. Tässä tutkielmatyössä on

koostettu elimistö-aiheista sanastoa. Aihepiiri sisältää 161 venäjänkielistä termiä ja 179 suomenkielistä termiä.

5. Sanastotyö

Elimistö-aiheinen sanasto on perussanasto, joka kääntäjän on osattava kyseisestä aihepiiristä. SysMLL:ssä Elimistöön aiheeseen kuuluvat ihmisten ja eläinten jäseniä sekä elimiä sekä aiheeseen liittyvä sanasto. Esimerkiksi ihmisen *jälka* on sana kun norsun tai hevosen *jalka*, sanastossa kuvataan vaan ihmisen *jalan*, koska opiskelija voi yhdistää saman *jalan* kaikkiin eläimiin. Sanastotyössä on huomioitu käytetyimmät sanat, jotka voivat esiintyä tavanomaisissa käännösteksteissä. Erikoistermit eivät kuulu mukaan. Sanan merkitys avautuu määritelmän avulla. Ensiksi on luotu venäjänkielinen sanasto ja sen jälkeen vastaavasti suomenkielinen. Mainittakoon, että termiä ei käännetty venäjältä vaan luotiin erikseen ihan oma suomenkielinen sanasto ja sen jälkeen etsittiin termien vastaavuuksia.

Valinta on tapahtunut sanan käytettävyyden ja taajuuden mukaan. Venäjän kielen sanojen taajuuksia on tarkistettu Venäjän kielen kansalliskorpuksesta (www.ruscorpora.ru) ja suomen kielen sanojen taajuus on tarkistettu "Google"-internethakupalvelimella.

Sanaston laatimisen jälkeen sanat jaoteltiin kolmeen eri tasoon: perustaso (elementary), keskitaso (intermediate) ja vaikea taso (advanced). Jaottelu on tehty sanan taajuuden mukaan: vähäisen taajuuden sanat luokiteltiin vaikeatasoisiksi, keskitaajuudella esiintyvät sanat luokiteltiin keskitasolle. Laajan taajuuden omaavat sanat luokiteltiin perustasoon kuuluviksi.

6. Sanakirja-artikkelin rakentaminen

Sanakirja-artikkeli tässä työssä on pohja, josta ohjelma itse generoi erilaisia harjoituksia. Ensimmäisenä on valittava avainsana, joka on englanninkielinen ja viittaa toisen kielen ekvivalenttiin. Esimerkiksi venäjänkielisen sanan *нос* ja vastaavasti suomen kielen *nenän* yhdistää englanninkielinen *nose*, johon lisätään merkityksen tarkentavaa tietoa. Lopulta avainsana on tämän näköinen: *nose part of face*.

Seuraavaksi artikkelissa on kuva, jota katsomalla käyttäjä voi arvata, mitä käsitettä sanakirja-artikkelissa tarkoitetaan. Sitten valitaan artikkelin kieli. Sen jälkeen on määritelmä, joka on

samalla kielellä kuin sanakirja-artikkelin sana. Määritelmässä ei saa esiintyä sanakirja-artikkelin sanaa .

Määritelmän jälkeen on kommenttiosuus, jossa esitellään sanojen poikkeavia taivutusmuotoja tai muita erityistietoja.

Seuraavana on vaikeustason valinta, mikä määrittyy taajuuden mukaan. Lopuksi on sanan käyttöesimerkki, jotta opiskelija voi nähdä millaisissa konteksteissa sanaa käytetään.

7. Harjoitukset

SysML-ohjelman ytimenä ovat harjoitukset. Niitä on neljää lajia: sanojen katselu (browse words), sanat ja ekvivalentit (words and equivalents), sanat ja määritelmät (words and definitions), kuvat ja sanat (words and pictures). Harjoitusten avulla henkilö voi opiskella vierasta kieltä omassa aikataulussaan vaikkapa kotona.

Ennen harjoitusten tekemistä on valittava aihe, opiskeltava kieli ja vaikeustaso.

Harjoitukset kannattaa aloittaa katselemalla sanoja, minkä jälkeen siirrytään haasteellisempaan vaiheeseen: sanat ja ekvivalentit tai sanat ja määritelmät. Ohjelma luo itse luettelon virheistään, joten käyttäjä pystyy harjoittamaan niitä sanoja, joissa hän on tehnyt virhettä.

Elimistö-aiheinen sanasto on ollut tutkielman tekijän omassa koekäytössä. Sen tuloksena sanastosta on löydetty lukuisia virheitä, mm. epäselviä määritelmiä tai esimerkkejä, vääriä vaikeustasoja, sanojen toistumista jne. Myös itse järjestelmästä on löydetty korjattavia kohtia. Harjoituksissa sama sana on esiintynyt jopa kolme kertaa peräkkäin, vaikka jo aluksi opiskelija oli antanut oikean vastauksen. Kuvatun kaltaisia virheitä korjataan ohjelman tekijöiden toimesta.

Lopuksi

Tämä on ensimmäinen pro gradu -tutkielma, joka käsittelee SysML-ohjelmaa. Tutkielman luonne ei ole pelkästään teoreettinen vaan suurelta osin käytännöllinen. Tuloksena on luotu elimistö-aiheinen sanasto. Sanastotyö on käsittänyt kaksi vaihetta: ensin valikoitiin sanat ja

sen jälkeen tarkistettiin niiden taajuudet. Sanastoon sisältyy 161 venäjänkielistä termiä ja 179 suomenkielistä termiä.

Sanaston perusteella on luotu sanakirja-artikkeleita, joita ohjelma käyttää harjoitusten luomisessa. Artikkeleiden koostamisessa innovatiivisuutta ja omaa luovaa ajattelua on tarvittu ennen kaikkea vaikeustasojen määrittelemisessä ja määritelmälajien tutkimisessa.

Toisena saavutuksena on oman koekäytön kautta löydetty ohjelman toiminnassa ongelmakohtia, jotka on korjattu. Tämän jälkeen verkko-ohjelmaa SysMLL voidaan hyödyntää opetuksessa Tampereen yliopiston kieli- ja käännöstieteiden laitoksella.

Jatkossa uusien sanastojen tekijät voivat käyttää mallina tämän tutkielman sanakirja-artikkelin koostamisstrategiaa sekä sanaston rakentamismenetelmää. Ohjelman tekijät voivat edelleen hioa ohjelman rakennetta, jotta käyttäjän on vaivatonta työskennellä ohjelman kanssa.

Tulevaisuudessa tarvitaan verkko-ohjelmia, joiden sisältö on tarkasti hiottu ja jotka toimivat moitteettomasti.

Lähdeluettelo

- Berkov, Valerij P. 1996. *Dvujazyčnaja leksikografija*. St-Peterburg: Sankt-Peterburgskij gosudarstvennyj universitet.
- Eskola, Sari 2002. *Syntetisoivat rakenteen käännössuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä*. Joensuu: Joensuun yliopisto.
- Haarala, Risto 1981. *Sanastotyön opas*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Mihailov, Mihail 2006. Struktura i soderžanie leksičeskih baz dannyh dlja obučajuščej programmy po inostarannym jazykam [= Kielenopetusohjelmiston sanastotietokantojen rakenne ja sisältö]. *Kompjuternaja lingvistika i intellektual'nye tehnologii. Trudy meždunarodnoj konferencii "Dialog" 2006 (Bekasovo, 31 maja – 4 ijulja 2006 g.)* [Computational Linguistics and Intellectual Technologies, International Conference "Dialogue 2006" Proceedings, (Bekasovo, 31 May – 4 June, 2006)] (http://www.dialog-21.ru/dialog2006/materials/html/Mikhailov.htm#_ftn1)
- Novak, Joseph D. & Gowin, D. Bob 1995. *Opi oppimaan*. Helsinki: Gaudeamus.
- Ooi, Vincent B.Y. 1998. *Computer Corpus Lexicography*, Edinburgh: Edinburgh University Press.

- Selegei, V. 1996. Elektronnye slovari i kompjuternaja leksokografia (= Sähköiset sanakirjat ja ohjelmistoleksikografia). http://www.lingvoda.ru/transforum/articles/selegey_a1.asp
- SysMLL 2006 = *System for Multiple Language Learning*. School of Modern Languages and Translation Studies, University of Tampere - (<https://www11.uta.fi/laitokset/kielikeskus/sysmll/>)
- YVI Software. [Lex!-ohjelman kuvaus ja käyttötieto sivulla:] <http://www.yvisoft.net/lex/faq.php>