

***Ю.Н. МАРЧУК***

**АВТОМАТИЧЕСКИЙ ПЕРЕВОД  
МНОГОЗНАЧНЫХ ТЕРМИНОВ**

В центре всех знаковых систем находится язык, роль которого в современных условиях возрастает. Существование, развитие и действие языка определяются принципами новой философии языка XX в., сформулированными Ю.В. Рождественским (8). Среди этих принципов главным можно назвать принцип наименования. Наименование, т.е. называние предмета, качества, обстоятельства и действия, является основной единицей языка, которая должна отражать сущность называемого явления максимально правильно и точно, в противном случае возможны всякого рода искажения, недоразумения, ошибки, неправильные действия и т.п. В работах Юрия Владимировича слову уделяется основное внимание, от исследования философских и чисто теоретических проблем номинации до конкретной работы по составлению тезауруса, являющегося основой обучения человека от его школьного возраста (тезаурус «Круг знания»).

Большое внимание Юрий Владимирович уделял современной информатической ситуации, в частности разным информационным системам, главным образом обосновывая их роль в новом этапе развития семиозиса в условиях массовой коммуникации. В рамках и с учетом этой ситуации сформулированы принципы новой философии языка. Кратко рассмотрим эти принципы в аспекте выбранной нами темы.

1. Язык (языковая деятельность) является распорядительной частью семиотической деятельности (там же). Она рядоположена системам прогностики, неприменных искусств (музыка, танец, изобразительное и пластическое искусство), прикладных искусств

(архитектура, костюм, дизайн), средств управления (меры и измерения, команды, средства ориентирования), игры, обряды, искусственные языки, в частности языки науки, техники, математики. Все эти средства семиотической деятельности составляют социальный интеллект. Общественная мысль течет от прогноза через образы искусств и техники к управлению разделением труда. Одновременно, в противотоке этому, обобщается и передается культура через средства обучения и воспитания. Неязыковая семиотика требует профессионального умения для раскрытия и создания тонких смыслов искусства, техники, науки и педагогики (в широком смысле слова). Языковая деятельность, будучи общим достоянием, обеспечивает назначение, истолкование и понимание всех неязыковых знаков и знаковых систем. Такова служебная и распорядительная роль языка в образовании общественного интеллекта.

2. Сам язык есть сочетание техники создания языковых знаков, составляющих индустрию языка: техника устной речи, материалы и орудия письма, книжная печать, все виды информационных технологий. Каждая технология создания языковых знаков имеет свои потенции для раскрытия смыслов, широты или узости общения между людьми (потенциальный обмен речевых коммуникаций), возможности зафиксировать и упорядочить наличную культуру личности, общества и его частей и организаций. Особенно важна языковая технология для фиксации культуры. Так, устная речь может формировать культуру лишь в памяти людей, письменная – хранить культурозначимые тексты и оперировать ими, создавая их смысловую организацию в виде каталогов и словарей, речь на компьютерах – моделировать все знаковые системы, кроме обрядов и танца (системы записи речи, игровые и обучающие системы, расчеты, моделирующие системы, дающие научный анализ и прогноз, электронные системы сочинения и исполнения музыки, компьютерная анимация, создание дизайна костюма, системы САПР, машинный перевод и, наконец, хранилища культурных сведений – системы информационного поиска). Компьютерные технологии ставят язык в положение, когда он максимально детально передает глубинные и профессиональные смыслы, свойственные неязыковым знаковым системам.

3. Отношения между языком и другими знаковыми системами и их частями есть отношения именованности. Слово как лексис

становится особенно ответственным, так как правильное именование, лежащее в основе лексических единиц, не только толкует назначение и применение всех вещей, но и определяет их понимание, воспитание людей и управление общественными процессами.

Обосновывая понятие и роль тезауруса терминов, Ю.В. Рождественский очень подробно останавливается на роли терминов в языковой ситуации (там же). Термины в современном развитом языке, говорит он, выполняют особую роль, состоящую в конечном формулировании содержания текста, составленного из слов, фразеологизмов и других терминов, как итог познавательного процесса. Итог познавательного процесса является не только личным достоянием автора текста, но и тем, что общественно признано и закреплено либо традицией, либо инстанциями, группами людей, специально выделенными для установления терминологии, как, например, для технических терминов.

Структура терминологии такова, что обеспечивает осуществление деятельности в системе раздельного труда, начиная от идейного замысла (философские термины), через его уточнение и критику (научные), через конструирование объектов (технические), связь объектов деятельности между собой (измерительные), реализацию в труде (искусства), реализацию иерархии деятельности как целого (команды), определение итогов труда (номенклатура) и, наконец, прогноз потребности в новой деятельности (прогностические термины).

В каждом виде деятельности термины будут личным и коллективно признанным обобщением текста в данной качественной специализации круговорота деятельности.

Жесткость и мягкость терминации, объем применения терминов зависят от того, какой уровень качественной специализации содержится в терминах и текстах, вводящих эти термины для организации деятельности. Дефинирование терминов в каждом виде деятельности производится различно.

Дефинирование философских терминов происходит посредством контекста философского сочинения – определения через контекст. Дефинирование научного термина происходит на базе традиции использования, но с уточнением нового содержания, которое вносит в него ученый – родовидовое определение: дефинирование технического термина в окончательной форме дано терми-

ностандартом, построенным по тезаурусному признаку – дефинирование по месту в системе. Дефинирование измерительной терминологии дается соглашением и отнесением к эталону – остенсивное определение: дефинирование терминов трудовых искусств (например, слово «процесс» – в праве, «аускультация» – в медицине, нота «до» – в музыке, «сверление» – в столярном деле) дается показом, т.е. это функциональное определение, по мысли Ю.В. Рождественского. Дефинирование команды дается в инструкции – определение-описание: дефинирование номенклатурного названия – через перечисление признаков; дефинирование прогностических терминов дается через отношение «вопрос–ответ» – эвристическое определение.

Различные виды терминов требуют и разного логического типа определений, следовательно, разного состава слов и фразеологизмов, из которых, в итоге компрессии текста, получается термин – единственная лексическая единица, к которой сводится содержание всего текста. Это значит, что при терминировании фактически изображается структура предметной и языковой деятельности. Сама система терминирования и последовательность классов терминов в ней есть как бы инструкция для реализации предметной и языковой деятельности.

Слово «терминология» неоднозначно. Оно может означать совокупность терминов конкретной предметной области («терминология хирургии», например), систему терминов определенной научной школы – узкое значение слова (например, понятие «фонемы» в терминологии Петербургской научной школы), и, наконец, науку о терминах. В последнем значении слово «терминология», как правило, заменяется термином «терминоведение». Терминоведение уже достаточно давно стало неотъемлемой частью прикладной лингвистики. Можно ли говорить о том, что терминоведение имеет свои собственные законы? Ответ на этот вопрос получаем в недавно вышедшем энциклопедическом словаре «Общее терминоведение» (9).

Законы терминоведения определяются в этом словаре как необходимые и устойчивые явления и процессы, а также связи явлений и процессов, существующие в области терминологии как объекте терминоведения и в области терминоведения как в науке о терминологии. Различные ученые по-разному определяют эти за-

коны. В.А. Татаринов, в частности, констатировал наличие терминологической неоднозначности в формах амбисемии, полисемии и эврисемии как всеобщий закон развития терминологической лексики. По заключению В.А. Татаринова, теоретические законы терминоведения имеют гносеологический статус и фиксируют познавательные закономерности в научно-методических приемах изучения термина. К таковым в современном терминоведении могут быть отнесены следующие законы: осознание термина как дескриптивной или как прескриптивной субстанции, рассмотрение термина как однозначной или неоднозначной единицы, закон изучения современного состояния терминоведения как проекции истории терминоведения, закон самостоятельности (интегративности) комплексности терминоведения, закон методологической полиаспектности терминоведения (материализм, феноменология, позитивизм, неопозитивизм), закон полиструктурности терминоведения как научной дисциплины, закон системного изучения термина и терминоведения и т.д. (9).

Таким образом, мы установили важность термина как единицы наименования в современной языковой коммуникации, с одной стороны, и с другой – общий состав законов терминоведения. Массовая коммуникация состоит из массовой информации, информатики и массовой рекламы, по Ю.В. Рождественскому. Во всех этих трех составляющих велика роль терминов, которые в реальных текстах активно взаимодействуют со словами, которые мы относим к общеупотребительной лексике. Границы между общеупотребительной лексикой и терминами чрезвычайно расплывчаты и в целом ряде случаев вообще не могут быть проведены. Так, например, «компьютер» – термин или общеупотребительное слово? Можно сказать, в соответствии с теоретическими положениями, приведенными выше, что ответ на этот вопрос зависит от типа текста, в котором встретилось это слово. В специальном тексте по вычислительной технике «компьютер», безусловно, должен считаться термином, в популярном изложении или в стихах, например, он не отличается от других общеупотребительных слов языка. Если предметное значение этого слова более или менее стабильно и устойчиво (вычислительная машина), то целый ряд слов-терминов многозначен, т.е. может означать совершенно разные предметы. Возьмем, например, слово «волна». Общеупотребительное значе-

ние этого слова перекрещивается со специальным (радиоволна). Таких многозначных слов-терминов и одновременно общеупотребительных слов в каждом естественном языке существует множество. Отсюда в рамках терминоведения возникает закон однозначности-многозначности терминов.

Рассмотрим эту проблему с точки зрения функционирования терминов. В рамках информатики термины организуются в тезаурус научно-технических терминов: см., например, тезаурус под редакцией Ю.И. Шемакина (Шемакин, 1972). В этом тезаурусе термины сгруппированы по предметным полям, к которым они относятся. Принцип тезауруса вообще чрезвычайно важен для пользования терминологией, для обучения и образования, в частности тот тезаурус, который обоснован и начат составлением под руководством Ю.В. Рождественского (тезаурус «Круг знания»). Каждый тезаурус дает как бы начальную и фундаментальную группировку терминов по их основным значениям. Однако в реальных текстах, в том числе и научно-технических, взаимодействие и взаимовлияние различных тематических областей может быть настолько сильным, что значения терминов претерпевают изменения по сравнению с теми, которые заложены в исходных тезаурусах. Именно поэтому в число основных законов терминоведения включен закон о неоднозначности (многозначности) терминов в ее различных видах (Татаринов, 2006).

Тезаурус – инструмент информатики. Современная информатика базируется на информационных системах, где существенной частью процесса создания и использования систем является автоматическая обработка текстов на естественных языках. Среди наиболее актуальных видов языковой коммуникации в рамках информатики, равно как и вообще в массовой коммуникации, нужно отметить перевод с одного естественного языка на другой (другие языки). Индустрия переводов растет с каждым годом по мере того, как все новые естественные языки выходят на арену мировой цивилизации. Оборот индустрии переводов в мире составляет миллиарды долларов, он не знает кризисов или спадов. Первое место среди переводимых материалов занимают научно-технические тексты. На них приходится примерно 75–80% переводимых в мире материалов. Языковые барьеры являются самыми значительными и трудно преодолеваемыми на пути мировой речевой коммуникации.

Другие барьеры, такие как географический, таможенный и пр., сравнительно легко преодолеваются, в то время как языковой барьер представляет серьезные трудности. Специалисты по информатике рассматривают три способа преодоления языковых барьеров: 1) универсальный язык (естественный или искусственный), 2) изучение иностранных языков основным составом коммуникантов, 3) перевод.

Кратко рассмотрим возможности и перспективы этих способов. Искусственный универсальный язык с незапамятных времен предлагался учеными как способ общения всех людей на Земле независимо от их родного языка. Ньютон, Лейбниц, Декарт и другие ученые выдвигали свои проекты универсального искусственного языка. Наибольшее распространение получил язык эсперанто (см., например, Линс, 1999). Несмотря на то, что во многих странах есть кружки эсперантистов, широкого распространения он не получил и вряд ли может рассматриваться как эффективное средство преодоления языковых барьеров. Другие искусственные универсальные языки имеют еще меньшее распространение. Несмотря на то, что некоторые энтузиасты продолжают разработку универсальных языков (см., например, Колегов, 2003), перспективы их сколь угодно широкого применения достаточно маловероятны. Более актуальна проблема использования какого-либо из естественных языков для всемирного общения. В настоящее время широкое распространение приобрел английский язык, который понимают во многих странах мира, не сильно сопротивляющихся глобализации. Если исходить из решения вопроса о выборе естественного языка для всеобщего общения из числа живущих на Земле людей, говорящих на своем языке как на родном, то следует выбрать китайский язык, так как на нем говорит большинство живущих людей в мире. Английский язык как родной занимает не слишком высокое место в ранге языков по численности говорящих на них. Специалисты Европейского экономического общества проводили специальное исследование, одним из выводов которого являлся постулат «Чем более развита данная нация, чем дальше она продвинулась на пути прогресса, тем большую цену она готова платить за право говорить на своем языке». Поэтому мы видим, как распространяет свой язык, например, «Альянс Франсез». В связи с этим можно утверждать, что какой-либо естественный язык в будущем вряд ли

станет всемирным языком, на котором люди Земли смогут общаться по любым темам и при всех обстоятельствах.

Вторым способом преодоления языковых барьеров является обучение иностранным языкам. Хорошо известно, что практически во всех странах мира в средней школе происходит изучение иностранного языка, одного или нескольких. Опять-таки по данным международных исследований видно, что практически везде результаты такого изучения достаточно скудны, т.е. хорошего знания иностранного языка школьник не получает, за исключением редких случаев, когда дело происходит в странах на международных языковых перекрестках – страны Европы, Люксембург и т.п., где ребенок может изучать два и более языка уже в семье. Иностранные языки преподаются также и в высших учебных заведениях, но и здесь специалисты вряд ли свободно владеют иностранным языком (языками) в той степени, в какой необходимо для преодоления языковых барьеров.

Остается третий способ – перевод. Как сказано выше, индустрия переводов растет с каждым годом по мере того как в мировую коммуникацию входят новые естественные языки и на них появляются новые информационные материалы. По состоянию на сегодняшний день, основная масса переводов осуществляется человеком (*human translation*) – русский эквивалент «человеческий перевод» не совсем стилистически приемлем. Человеческому переводу противостоит перевод машинный, или автоматический, осуществляемый компьютером. Он в десятки раз быстрее человеческого и дешевле, однако современный машинный перевод без редактирования характеризуется достаточно низким качеством и в целом ряде случаев не удовлетворяет требованиям информационного обслуживания. Несмотря на это, потребности в переводе столь велики, что в сети Интернет и в других системах массовой коммуникации на коммерческой основе действуют многие системы машинного перевода (см., например, обзор П.Н. Хроменкова [Хроменков, 2005]).

В машинном переводе, разработки которого начались уже достаточно давно, а именно в конце 50-х годов XX в., можно отметить две основные тенденции. Одна из них основана на предположении, что все естественные языки имеют одну логическую основу, построены по принципам логики, и можно создать некий



единый логический язык, машинный перевод на который и с которого можно будет затем осуществлять для каждого отдельного естественного языка (Марчук, 1983, 1985). Нетрудно заметить, что при этом воспроизводится идея некоторого всемирного универсального языка. Однако по полной аналогии с развитием и становлением единого искусственного универсального языка, который так и не стал действенным средством всемирного общения, язык «смысла» для машинного перевода за много лет потраченных на его разработку усилий не дал никаких практических результатов. Все дело в том, что понятие «смысл» в естественном языке весьма широко и зависит от контекста общения – контекста в самом широком значении этого слова, как в экстралингвистическом, так и в чисто лингвистическом. В.А. Звегинцев еще в 1976 году убедительно показал, что набор предложений «Охотник выскочил из-за куста и ударом ноги убил волка», «Волк был убит ударом ноги выскочившего из-за куста охотника», «Из-за куста выскочил охотник и ударом ноги убил волка», «Ударом ноги выскочившего из-за куста охотника волк был убит» и т.п. никак нельзя считать имеющим один и тот же смысл, хотя экстралингвистическая ситуация вроде бы при этом остается одной и той же во всех вариациях. Каждая вариация – перифраза – имеет свой собственный смысл, реализуемый в конкретном лингвистическом контексте (Звегинцев, 1976). Логический анализ может быть положен в основу лингвистического анализа текста лишь в очень ограниченных предметных областях – например в математике, хотя и в таких областях новое открывается, как правило, в отступлениях от уже существующего, признанного и подчиняющегося логическим законам. Лингвистика имеет свои собственные закономерности и особенности, связанные в первую очередь с тем, что мы называем контекстом в самом широком смысле.

Неслучайно в прикладной лингвистике последнего времени получило развитие понятие «дискурс», которому посвящаются как теоретические, так и практические исследования и разработки. По нашему мнению, этот термин на самом деле означает именно контекст высказывания, устного или письменного, и в границах этого высказывания и следует искать «смысл» или «содержание» текста. А.И. Новиков убедительно показал разницу между этими двумя понятиями (Новиков, 2002). «Смысл» не всегда соответствует со-

держанию и часто вообще расходится с ним. Так, художественный текст, описывающий разговор мужчины и женщины в поезде около окна вагона, когда они обмениваются репликами о том, что видят из окна, имеет в качестве содержания описание природы и событий, увиденных из окна, но «смысл» их разговора не в этом, а в выяснении отношений. Такие данные были получены в результате убедительных экспериментов с информантами. «Смысл» может совпасть с содержанием только в научном тексте, например по химии, что и было отмечено в эксперименте.

Все это означает, что механическое извлечение смысла из текста и его перевод на компьютере по чисто формальным правилам может быть осуществлен только с помощью некоторых достаточно четких моделей, учитывающих контекст слова или словосочетания в весьма широком значении этого термина. Вследствие этого в настоящее время системы машинного перевода, действующие в практическом плане, строятся так или иначе по моделям, которые мы называем моделями переводных соответствий.

Это второе направление в современном машинном переводе. Оно ориентировано главным образом на создание алгоритмов передачи переводных соответствий в рамках конкретных языковых пар. Модель перевода на уровне переводных соответствий описана мною в книге «Методы моделирования перевода» (М., 1985). Существующие системы машинного перевода практической ориентации так или иначе воспроизводят ее, иногда включая разного рода модификации, не меняющие сущности. Переводные соответствия в рамках данной языковой пары и определенного направления перевода делятся на эквивалентные, вариантные и трансформационные. Это, кстати, соответствует идеям традиционного (человеческого) перевода. Другим компонентом модели является совокупность словаря, грамматики и семантики данной предметной области. Взаимодействие двух компонентов модели дает возможность построить эффективную систему алгоритмов. Существенной частью системы является контекстологический словарь, с помощью которого разрешается лексическая многозначность (неоднозначность) слова, благодаря использованию алгоритмов, запрашивающих контекст о наличии (отсутствии) тех или иных детерминант – признаков, определяющих выбор конкретного значения (перевода) слова. Классы детерминант могут быть самые различные – от морфологических

признаков и лексической сочетаемости слов до некоторых синтаксических функций, которые в системе определяются на этапах анализа, предшествующих работе контекстологического словаря. Последний представляет собой набор алгоритмов перевода многозначных слов, разработанный на основе некоторого исходного корпуса текстов и их переводов на другой язык.

Исходный корпус текстов в данном случае представляется важной составляющей и некоторым фундаментом процесса. Универсальный контекстологический словарь, который был бы пригоден для многих типов текстов, является утопией. В настоящее время получила развитие теория подязыков, в рамках которой употребление слов, в частности терминов, определяется достаточно четко. (См., например, работу по машинному переводу с китайского языка на русский В.М. Зелко, 1991). Но нельзя сказать, что понятие подязыка, как и родственные понятия (стиль речи, жанр и т.п.), имеют однозначную трактовку как в теории, так и в практике. В реальных текстах также достаточно часто взаимодействуют разные стили, подязыки, тематические области и т.п. В результате автоматический (машинный) перевод слов, в частности терминов, достаточно часто ведет к ошибкам и непониманию текста.

Рассмотрим некоторые примеры. Перевод с английского на русский язык, сделанный системой PARS, называется «Гуртовщики мыши». В английском исходном тексте используется термин Drivers. Слово «мышь» остается переводом английского mouse, поскольку как в биологическом, так и в компьютерном словаре этот перевод не меняется. Однако «гуртовщики» ясно показывают, что drivers перевелось либо по общеупотребительному словарю, либо по какому-нибудь сельскохозяйственному. Можно также предположить, что в системе использовался единый словарь без разделения по предметным областям, и в этом словаре указанным английским словам – между прочим терминам – были приданы именно такие значения.

Возьмем первый абзац этого текста в русском переводе.

«Майкрософт компания получает много откликов после появления Окон 95. Мы выявили, что много пользователей встретили проблему мыши. В этом документе Служба Технического Упора Майкрософт компании сводит вместе всю полезную информацию о

возможных проблемах с мышами и гуртовщиками мыши и забота-стрелание».

Первое впечатление от данного отрывка состоит в том, что с точки зрения морфологии и синтаксиса перевод качественный, так как почти нет ошибок в падежах, согласовании, порядке слов. Многие термины переведены правильно (пользователи, мышь, компания и т.п.). Однако есть и существенные ошибки, которые не совсем понятны, если в переводе использовался специальный словарь программного обеспечения. Так, «Окна 95» представляется излишним, если учесть распространение термина Windows практически независимо от тематики текстов. «Служба техничного упора» идет как перевод Technical Support System (или Service).

Какого рода контексты нужны для того, чтобы избавить перевод по крайней мере от основных ошибок?

Прежде всего отметим, что должна быть правильно определена принадлежность переводимого текста к определенному предметному полю. При четком определении текста как принадлежащего к области компьютерного обеспечения не появился бы «гуртовщик» или «Окна 95». Для подобной идентификации текста нужны достаточно мощные формальные средства, если мы хотим, чтобы такую идентификацию проводил не редактор, а система машинного перевода. Представляется, что здесь могли бы помочь статистические данные о количестве и распределении однозначных терминов, принадлежащих определенному тематическому полю, — например строительных в текстах по строительству, медицинских в текстах по медицине и пр. Нетрудно определить, что при этом выбор исходного массива текстов имеет решающее значение.

Здесь следует подчеркнуть именно однозначность терминов, т.е. четкую принадлежность их к конкретной предметной области. Многозначные термины, имеющие разное значение в различных предметных областях, не могут служить показателями существа предметной области. Так, термин «реакция» может иметь совершенно разные значения в разных предметных областях. Такой же термин, как «кариес», вряд ли можно встретить где-либо еще, кроме как в медицинском тексте по стоматологии. Всякая метафоричность при этом исключается, так как машинный перевод по определению не может применяться для перевода художественных текстов.

Трудности на этом пути – взаимодействие двух и более тематических полей в пределах одного текста. Говоря о медицине и строительстве, например, нетрудно представить себе текст, где говорится о типичных болезнях строительных рабочих. Какой тематический словарь при этом следует выбрать? Думается, что в этом случае поможет установление некоторых порогов статистического распределения однозначных терминов. Так, можно предположить, что в такого рода смешанном тексте больше будет медицинских терминов, чем специфических однозначных терминов строительства. Подобные данные опять-таки можно будет получить на базе исходного массива текстов.

Рассмотрим теперь принцип организации и действия автоматического контекстологического словаря для разрешения лексической многозначности слов, в том числе и терминов. Подробно идея, методы составления и действия автоматического контекстологического словаря описаны в (6). Несмотря на то что эта информация была опубликована достаточно давно, в настоящее время действующие системы машинного перевода, как об этом можно судить по результатам, редко используют контекстологические словари в переводе многозначных слов. Это можно объяснить, во-первых, трудоемкостью составления такого словаря и, во-вторых, отсутствием более или менее продвинутой теории такого словаря. Мною были высказаны только основные, начальные принципы, которые не потеряли своего значения и в настоящее время, но должны быть подкреплены новыми теоретическими соображениями и практическими рекомендациями с учетом далеко продвинувшейся вперед новой информационной технологии. Эта технология, безусловно, дает в руки исследователю и разработчику новые мощные средства решения проблемы, надо только теоретически их осмыслить и предложить новые механизмы осуществления теоретических положений на практике.

Принцип организации контекстологического словаря – поиск в тексте детерминант, т.е. контекстных признаков, каждый из которых, или в некоторой совокупности, определяют конкретное лексическое значение данного слова в данном тексте. При этом понятие «значение», которое тоже довольно расплывчато и по-разному трактуется в разных теориях, приравнивается переводу данного многозначного слова в исходном конкордансе. Конкорданс, т.е.

особым образом организованный словарь, в котором слово подается во всех контекстах его употребления в исходном массиве, является необходимым материалом для составления контекстологического словаря.

Детерминантами могут быть морфологические, синтаксические и лексические признаки, определяющие значение слова.

Действует автоматический контекстологический словарь в определенной последовательности этапов алгоритма обработки (перевода) текста, после того как определена необходимая морфологическая, синтаксическая и другая информация, на которую может опереться поиск детерминант. Однозначные слова, т.е. слова, которые в исходном конкордансе имеют всегда один и тот же перевод, в контекстологический словарь не попадают. Кроме того, до работы контекстологического словаря действует словарь идиом, неразложимых словосочетаний, компоненты которого однозначно определяют свое значение в рамках соответствующей идиомы.

Подводя итог предыдущим рассуждениям, определяем следующую последовательность действий для составления алгоритмов автоматической обработки (перевода) терминов в системе машинного перевода.

1. Отбирается исходный массив текстов, подлежащих переводу.

2. По возможности точно определяется предметная область, к которой принадлежат переводимые тексты. При этом максимально используются объективные критерии, такие как статистические характеристики распределения однозначных и многозначных терминов.

3. На основе выбранного массива составляется исходный словарь-конкорданс, в котором выбранное слово приводится в контексте его употребления. При наличии параллельных текстов приводятся также переводы многозначных и однозначных слов-терминов.

4. Классифицируются детерминанты определения значений (переводов) многозначных слов-терминов.

5. Составляется алгоритм поиска детерминант – алгоритм перевода данного многозначного слова-термина.

6. Производится проверка действия и эффективности алгоритма перевода по текстам, не вошедшим в исходный конкорданс, но принадлежащим выбранной предметной области.

7. По результатам проверки и опытных переводов пополняется список детерминант и корректируется работа алгоритма перевода данного многозначного термина.

### Список литературы

1. *Звегинцев В.А.* Предложение и его отношение к языку и речи. – М., 1976. – 306 с.
2. *Зелко В.М.* Проблемы разработки лингвистического обеспечения системы китайско-русского информационного машинного перевода: Дис. канд. филол. наук. – М., 1991. – 165 с.
3. *Колегов А.В.* Международный язык-посредник эльюнди. – Тирасполь, 2003. – 496 с.
4. *Линс У.* Опасный язык. Книга о преследованиях эсперанто. – М., 1999. – 575 с.
5. *Марчук Ю.Н.* Методы моделирования перевода. – М., 1985. – 201 с.
6. *Марчук Ю.Н.* Проблемы машинного перевода. – М., 1983. – 232 с.
7. *Новиков А.И.* Доминантность и транспозиция в процессе осмысления текста // Проблемы прикладной лингвистики 2001. – М., 2002 – С. 155–180.
8. *Рождественский Ю.В.* Философия языка. Культуроведение и дидактика. Современные проблемы науки о языке. – М., 1996. – 239 с.
9. *Татаринов В.А.* Общее терминоведение. Энциклопедический словарь. – М., 2006. – 526 с.
10. *Хроменков П.Н.* Современные системы машинного перевода. – М., 2005. – 159 с.
11. *Шемакин Ю.И.* Тезаурус научно-технических терминов. – М., 1972. – 672 с.