

*Русанова И.
(Одесса)*

К ПРОБЛЕМЕ СОВЕРШЕНСТВОВАНИЯ ПРОГРАММ МАШИННОГО ПЕРЕВОДА

Аннотация. В статье предлагается использование логико-системного языка ЯТО в качестве компонента метаязыка машинного перевода. Необходимость такого компонента подтверждается анализом ошибок машинного перевода.

Ключевые слова: машинный перевод, системно-логический, концепт, структура, статистический, синтагма, ошибки.

XXI век характеризуется расширением межязыковых и межкультурных связей. Растёт информатизация общества. Появилась потребность охвата максимального объёма информации в минимальных условиях.

История машинного перевода (МП) насчитывает около шестидесяти лет. За это время сменилось несколько поколений МП, выросли технические возможности компьютеров по производительности и объёму памяти. Появилась возможность решать проблемы смыслового содержания переводимого текста.

В настоящее время принято делить системы перевода на системы типа Transfer и системы типа Interlingua. Алгоритмы типа Transfer строятся на трёх процессах: анализ исходного текста в единицах и структурах исходного языка, преобразование этих структур и единиц на структуры языка перевода и синтез выходного текста. Системы типа Interlingua предполагают наличие метаязыка. Алгоритм перевода состоит из анализа исходного текста в терминах метаязыка и синтеза выходного текста соответствующего языка из метаструктур и метаединиц. Однако до сих пор не определён метаязык такого анализа. Этим объясняется актуальность выбранной темы, так как машинный перевод является одной из наиболее используемых компьютерных услуг.

Целью статьи является анализ работы машинных переводческих программ на основе изучения отрывков английских художественных текстов и определение путей устранения ошибок машинного перевода.

Сегодня в различных областях знания широко используются системные средства познания. Интерпретацию текста можно представить как системное представление объекта, которое осуществляется по принципу «концепт—структура—субстрат» или «структура—системообразующее

свойство—субстрат» [1]. Современные МП программы зачастую руководствуются системным представлением типа «субстрат — структура—концепт», то есть перевод начинается со слов. Такой анализ с точки зрения системного подхода является наименее релевантным. Неправильно выбранный вариант перевода отдельных слов влечёт за собой массу смысловых ошибок, как структурных, так и концептуальных.

Схему системного представления «концепт—структура—субстрат» следует использовать при анализе исходного текста. В свою очередь концепт текста обладает своей структурой. Сюда входит функциональный стиль, жанр, предметная область, тематика текста [2]. Концепт текста необходимо определить в первую очередь, что нетрудно сделать, используя статистические методы. Нужно выбрать наиболее частотные ключевые слова, дать им перевод и следить за неизменностью данного значения в процессе перевода. Далее следует переходить к анализу структуры предложений.

Системный язык тернарного описания (ЯТО) предлагает нам логико-системные модели [3; 4]. Однако вряд ли они используются разработчиками МП программ. В лучшем случае они оперируют традиционными синтаксическими моделями членов предложения. Первые нетрудно увязать с системно-логическими моделями. Единицы перевода определяются на базе структурных моделей. Это должны быть не просто слова, а функционально значимые единицы, то есть, кроме семантического значения, эти слова или группы слов должны выполнять функционально значимую роль предложения, например, должны быть определённым членом предложения. Минимальная синтагма, грамматическое или фонетическое слово является такой единицей [5]. Минимальная синтагма состоит из знаменательного слова и примыкающих к нему детерминантов и служебных слов. Служебные слова нередко сигнализируют о функции члена предложения, особенно в английском языке с его небогатой системой флексий.

Синтагматическое членение текста нетрудно увязать с его логико-грамматической структурой [6]. Попытки использовать ЯТО для анализа логико-грамматических структур как письменного, так и устного текста уже предпринимались [4].

Интонограмма звучащего текста даёт возможность зафиксировать «смысловые вершины» [7], то есть ключевые слова в тексте, а также определить актуальное членение предложения на «данное» и «новое». Отыскать эти единицы в письменном тексте значительно труднее. С помощью ЯТО можно зафиксировать и варианты переводных значений слов (субстрата). В последней версии ЯТО А.И. Уёмов предложил ряд атрибутивных системных параметров (АСП) и соответствующих им системных закономерностей, которые можно использовать как варианты слов и концептов, а также определить стратегию их поведения. Такой потенциал ЯТО позволяет сделать вывод о возможности использования его как компонента метаязыка при машинном переводе. Конечно, внедрение логического системного компонента ЯТО в качестве составляющей метаязыка требует большой работы многих исследователей: логиков-системщиков, лингвистов, математиков, программистов. Но эта работа необходима, о чем еще раз свидетельствует анализ ошибок машинного перевода, проведенный нами.

Группе студентов V курса специальности «Английский язык» Крымского гуманитарного университета было предложено перевести отрывки из английских художественных произведений с помощью МП программ Google, Promt, On-line. Затем студентам должны были провести анализ ошибок машинного перевода по следующим параметрам: грамматические ошибки: а) морфологические, б) синтаксические; системно-логические ошибки: а) концептуальные (искажение смысла высказывания), б) структурные (нарушение логической последовательности высказывания); в) субстратные (выбранное слово не соответствует заданному концепту).

Было проанализировано 18 отрывков, состоящих из 233 предложений. Всего было отмечено 1594 ошибки. Из них 687 были выделены как грамматические, 907 — как системно-логические. Из числа грамматических ошибок 41% составили морфологические: неправильные флексии и неверное определение части речи. 59% ошибок студенты определили как синтаксические (неправильное определение структуры и функции предложения).

Что касается системно-логических ошибок, то 54% составили субстратные ошибок (неверно выбранный вариант перевода), 26% — концептуальные ошибки (неправильно передан смысл высказывания), 19% составили структурные ошибки (нарушена логическая последовательность высказывания).

В качестве примера проанализируем перевод отрывка из произведения Г.Х. Андерсена “Leaf from Heaven” [9]: “High up in the clear, pure

air flew an angel, with a flower plucked from the garden of heaven. As he was kissing the flower a very little leaf fell from it and sunk down into the soft earth in the middle of the wood. It immediately took root, sprouted and sent shoots among other plants. “What a ridiculous little shoot!” said one.” It must be a kind of garden plant” said another.”

Первым представим перевод отрывка, сделанный МП программой on-line Google: «Высоко в ясном чистом небе летел ангел с цветком сорвал с садом. Как он целовал цветок очень мало лист упал в мягкую землю в середине леса. Он сразу пустил корни и разослал побегов среди других растений. «Что смешного мало стрелять», — сказал один. «Это должен быть своего рода завод сада», — сказал другой». Сравним этот вариант с правильным переводом: «Высоко в ясном, чистом небе летел ангел с цветком, сорванным в райском саду. Когда он целовал цветок, очень маленький листок упал с него в мягкую землю в середине леса. Он сразу пустил корни и разослал побеги среди других растений. «Какой смешной маленький росток! — сказала одно растение. «Это должно быть какое-то садовое растение», — сказала другое».

Как мы видим, в машинном переводе не определен заранее концепт, то есть тема текста. Отсюда пошли лексические и грамматические ошибки. Сначала слово “shoot” правильно переводится как «побег», но затем предлагается слово «стрелять». Сначала слово “plant” верно переводится как «растение», но затем оно представлено как «завод». Смысл текста полностью искажен. Не соблюден принцип одинакового значения слова в одном тексте. Это очень грубая, но легко устранимая ошибка. Заранее не были определены и переведены ключевые слова текста. Видимо, необходимо при переводе руководствоваться не одним предложением, а более крупным отрезком текста для определения ключевых слов и логической структуры предложения.

Проанализируем перевод отрывка из произведения Г. Уэльса «Человек невидимка» [10].

Оригинал текста “The Invisible Man by H. G. Wells”: “The stranger came early in February one wintry day, through a biting wind and a driving snow, the last snowfall of the year, over the down, walking as it seemed from Bramblehurst railway station and carrying a little black portmanteau in his thickly gloved hand. He was wrapped up from head to foot, and the brim of his soft felt hat hid every inch of his face but the shiny tip of his nose; the snow had piled itself against his shoulders and chest, and added a white crest to the burden he carried”.

Перевод № 1 (Сократ Персональный 4.1): «Новичок приходил в начале февраля один зим-

ний день, через хлесткий ветер и ведущий снег, последний снегопад года, над вниз, проходит как он кажется с железнодорожной станции Bramblehurst и несущий небольшим черным чемоданом в его густо gloved руке. Он был завернут по из головы футу, и край своей мягкой фетровой шляпы прятал каждый дюйм его лица, но лоснящийся конец его носа; снег имел сам против его плеч и ящика, и добавивших белый пик к бремени он понес».

Перевод № 2 (on line переводчик Google): «Вошел незнакомец в начале февраля один зимний день, через колючий ветер и снег вождения, последний снегопад год более вниз, ходьба, как казалось со станции Bramblehurst железнодорожных и проведение маленького черного чемодана в его густо рукой в перчатке. Он был обернут с головы до ног, и края его мягкой фетровой шляпе спрятался каждый дюйм его лицо, но блестящий кончик носа; снег свалили себя от плеч и груди, и добавил, белый гребень, чтобы бремя он нес».

Перевод № 3 (переводчик Translate.ru): «Незнакомец приехал в начале февраля один зимний день, через пронизывающий ветер и ведущий снег, последний снегопад года, по вниз, идя, поскольку это казалось от железнодорожной станции Bramblehurst и переноса небольшого черного портманто в его плотно руке в перчатке. Он был обернут с головы до пят, и край его мягкой фетровой шляпы скрыл каждый дюйм его лица, но солнечный кончик его носа; снег сложил себя против его плеч и груди, и добавил белый гребень к бремени, которое он нес».

Правильный перевод: «Незнакомец пришел одним ветреным февральским днем. Кусал злой ветер и валил снег, последний снегопад года. Он шел по склону, вероятно от железнодорожной станции Блумбелхерст и нес небольшой черный чемодан в руке. Он был закутан с ног до головы, края его мягкой фетровой шляпы закрывали лицо, кроме блестящего кончика носа, снег лежал на его плечах и груди, добавляя крест к бремени, который он нес».

Как видим, смысл этого отрывка можно понять в переводе с большим трудом. Виной этому не только неверно выбранные варианты перевода слов («Foot» — «нога», а не «фут» — единица измерения; «Fire» — не «огонь», а «камин» и так далее). Причиной неадекватности перевода является, прежде всего, неопределенность логико-синтаксической структуры предложений. Для устранения такой неопределенности используем приемы и символы ЯТО.

Первым шагом при переводе предложения должно быть определение структуры предложения: является ли предложение простым, сложно-

подчиненным или сложносочиненным. Для этого необходимо разбить текст на синтагмы, функционально значимые единицы предложения. Это группы слов, выполняющие функции основных членов предложения [6]. Устанавливаем количество слов в группах главных членов предложения, подлежащего и сказуемого, и определяем функции других синтагм. В работе используем пока лишь один символ ЯТО: «а» — «неопределенная вещь». Если поместим «(а)» в скобки и приписать справа «(а)» еще одно «а», то это будет означать, что вещь «(а)» обладает свойством «а». Если мы поместим символ «а» слева от скобки, «а (а)», то это будет означать, что вещь обладает отношением «а». Если любую из этих формул мы поместим в квадратные скобки, то это будет понятийная формула — вещь со свойствами, не утверждение. Общая схема предложения (a){(a) a...} a}. Такой анализ поможет определить логическую структуру предложения. Проанализируем первое предложение из отрывка приведенного выше отрывка (таблица 1).

Обозначения: подлежащее — S; сказуемое — P; дополнение — obj; обстоятельство — adv; определение — adj; adv. V — глагольное обстоятельство; / — синтагматическая пауза внутри предложения; // — конец предложения.

На основе проведенного анализа предложение можно проанализировать так: простое распространённое, осложнённое обстоятельствами 2, 3, 4, 6, 7, 8, 9, 10, 11; обстоятельства 7, 8, 10 глагольные, то есть несут вторичную предикацию. В предложении имеется только одно подлежащее «незнакомец» — «Stranger». Но оно осложнено причастными оборотами «walking...» и «carrying...», которые являются носителями вторичной предикации и дают нам возможность построить отдельные предложения с этим же подлежащим. Такую возможность МП программы не реализовали, так как арсенал переводных структурно-синтаксических, как и субстратно-лексических вариантов в этих программах беден. Они не удовлетворяют требованию разделения текста на синтагмы, анализа функции каждой синтагмы и определения синтаксической и логической структуры предложения.

В качестве выводов можно предложить следующее.

Программы машинного перевода должны предусматривать следующие процедуры при анализе исходного текста:

— определение концептуальной структуры текста: стиль, жанр, предметная область;

— выяснение логико-синтаксической структуры предложений и всего текста в целом, для чего необходимо разбить предложения на синтагмы и

определить функцию каждой синтагмы, а потом, исходя из результатов синтаксического анализа, обозначить логическую структуру предложений и всего текста;

– подбор правильных вариантов значений переводных слов путем опоры на концептуально-структурные характеристики текста;

– учет логической структуры предложений и текста, которые должны быть наполнены семантическим содержанием в соответствии с нормами языка перевода.

Такие возможности предоставляет переводчику ЯТО (язык тернарного описания). Преимущество этого языка перед другими логическими языками состоит в том, что он базируется на философской триаде вещь – свойство – отношение, а также на категориях определенности, что позволило создать

глубокий формализованный аппарат системного анализа.

Недостатком всех логических моделей является при переводе отсутствие чётко определенной единицы перевода. Мы предлагаем использовать в качестве единицы перевода синтагму. Применение ЯТО как компонента метаязыка машинного перевода будет способствовать устранению логических смысловых ошибок. С его помощью можно зафиксировать концепт и соответствующие структурные и субстратные единицы. Задача разработчиков МП программ – дать языковые варианты концевых, структурных и субстратных значений для различных языков. Это позволит объединить различные формы машинного перевода: фразеологическую, автоматизированную, статистическую системы.

Таблица 1

Анализ логической структуры предложения с помощью ЯТО
(языка тернарного описания)

Исходный текст	Сократ	Google	Translate ru.	Логические Структуры	Верный Перевод
1. The stranger Come / SP	Новичок проходил	Вошел незнакомец	Незнакомец приехал	(a) a – SP	Незнакомец пришел
2 In February/adv	В феврале	В начале февраля	В начале февраля	(a . a advt)	Одним ветреным февральским днем
3. one wintry day / adv	Один зимний день	Один зимний день	Один зимний день	[(a) a]a; advt	
4. Through a biting wind and a driving snow / adv	Через жесткий ветер и ведущий снег	Через колючий ветер и снеговождения	Через пронизывающий ветер и ведущий снег	[(a) a I a(a)]adv V	Кусал злой ветер и валил снег
5. the last snowfall of the year / adj	Последний снегопад года	Последний снегопад года	Последний снегопад года	{{(a)a} a} adv	Последний снегопад года
6. over the down/ adv	Над, вниз	Более, вниз	Но вниз	(a) adv.	Он шел по склону
7. walking adv, V	Проходить	Ходьба	Идя	A adv V.	
8. as it seemed / adv, V	Как он кажется	Как казалось	Поскольку это казалось	[(a)a adv V.	Вероятно
9. from Bramblerhurst railway station / adv	С железнодорожной станции	С станции в Bramblerhurst	От железнодорожной станции Bramblerhurst	{{(a)a}a} adv.	От железнодорожной станции Блумберхерст /
10. and carrying a little black portmoneou / adv, V	Но и несущий небольшой черный чемодан	И проведение маленького черного чемодана	И переноса небольшого черного портманто	A { (a) a , a } adv.V	И нес небольшой черный чемодан
in thickly gloved hand // adv	В его густой деловой руке	В его густо рукой в перчатке	В его плотно руке в перчатке	{{(a)a}a} // adv.	В руке, одетой в плотную перчатку

Литература

1. Ахманова О.С. Словарь лингвистических терминов / О.С. Ахманова. — М. : Советская энциклопедия, 1966. — 608 с.
2. Варбанец П.Д. Статистические оценки структурно-синтагматического членения английского научного текста / П.Д. Варбанец, И.И. Русанова // Придніпровський науковий вісник. — 1998. — № 9. — С. 9—19.
3. Кунцос О.И. Использование языка тернарного описания в англо-русском машинном переводе / О.И. Кунцос // Параметрическая общая теория систем и её применение. — Одесса : Астропринт, 2008. — С. 24—37.
4. Русанова И.И. Определение семантической структуры переводного текста с помощью статистических методов / И.И. Русанова // Наука та освіта. — 2004. — № 4—5. — С. 193—197.
5. Русанова И.И. К проблеме логического моделирования языковых и речевых конструкций / Русанова И.И., Уёмов А.И. // Тезисы докладов и сообщений 14-го Всесоюзного семинара (АРСО). — Каунас, 1986. — С. 36—41.
6. Торсуева И.Г. К типологии интонационных структур / И.Г. Торсуева // Лингвистика и методика. — М. : Наука, 1993. — С. 25—31.
7. Уёмов А.И. Монарные атрибутивные системные параметры в параметрической общей теории систем / А.И. Уёмов // Параметрическая общая теория систем и её применение. — Одесса : Астропринт, 2008. — С. 24—37.
8. Уёмов А.И. Системный подход и общая теория систем / А.И. Уёмов. — М. : Наука, 1978. — 272 с.
9. Andersen H.C. Leaf from Heaven / H.C. Andersen [Электронный ресурс]. — Режим доступа: <http://hca.gilead/org.il/heaven/html>.
10. Wells H.G. The Invisible Man / Wells H.G. [Электронный ресурс]. — Режим доступа: <http://www.online literature.com/wellsg/invisible/>.

Русанова І.І. До проблеми удосконалення програм машинного перекладу. — Стаття.

Анотація. У статті пропонується використання логико-системної мови ЯТО як компоненту метамови машинного перекладу. Необхідність такого компоненту підтверджується аналізом помилок машинного перекладу.

Ключові слова: машинний переклад, системно-логічний, концепт, структура, статистичний, синтагма, помилки.

Rusanova I.I. Improving Machine Translation Programs. — Article.

Summary. The article considers the problem of the use of the logical systemic language as a component of the metalanguage for machine translation. The necessity of the component is suggested by the analysis of the mistakes in machine translation.

Key words: machine translation, systemic logical, concept, structure, statistic, syntagma, mistakes.