

Некоторые аспекты использования лексических функций в системе машинного перевода «ЭТАП-3»

Андреева А.С.

В данной работе будут рассматриваться некоторые случаи использования лексической функции Oper1 в системе машинного перевода «ЭТАП-3».

Лексические функции (ЛФ), «характеризующие идиоматичную взаимозаменяемость и сочетаемость лексем» [ТКС 1984], используются в «ЭТАПе» для перевода большинства несвободных словосочетаний. Подробное описание лексических функций можно прочитать в [Мельчук 1974] или [ТКС 1984]. Лексические функции, используемые в современной версии «ЭТАПа», несколько отличаются от описанных в этих книгах, но основная идея остается прежней.

Функция Oper1 связывает существительное (аргумент функции) и глагол поддержки (ее значение), первым актантом которого является подлежащее, а вторым – главное дополнение, т.е. само существительное.

Примеры:

ПОПУЛЯРНОСТЬ: пользоваться

ОТЧАЯНИЕ: быть <в2>

АВАРИЯ: терпеть

АГРЕССИЯ: совершать / осуществлять

ВОЙНА: вести²

Анализ материала (см. таблицу 1) показал, что в простых случаях обработка лексических функций, в том числе функции Oper1, обеспечена в системе «ЭТАП-3» достаточно хорошо. Однако в переводах более сложных и более тонких контекстов возникало существенное количество ошибок. В данной работе будет рассмотрено несколько таких случаев и показано, как были устранены ошибки в обработке ЛФ.

Таблица 1.

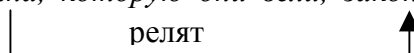
№	L1	L2
1.	Мы ведем борьбу за победу цивилизации.	We wage the struggle for a victory of the civilization.
2.	Они ведут беседу.	They carry on a talk.
3.	Мы выдвинули новый аргумент.	We have produced a new argument.
4.	Вчера Вася нанес визит Пете.	Yesterday Vasya paid a visit to Petya.
5.	Он внес большой вклад в развитие этой системы.	He has made a large contribution to the development of this system.
6.	Учитель имеет на детей большое влияние.	The teacher has a large influence on the children.
7.	Петя вел войну с соседом.	Petya waged a war with a neighbour.
8.	Вася задал Пете вопрос.	Vasya has asked Petya a question.
9.	John asked Pete a question.	Джон задал Пете вопрос.

Одним из контекстов, в которых система «не справлялась» с функцией Oper1, оказались сложные синтаксические конструкции, в которых аргумент и значение функции

находятся в разных клаузах. Это предложения типа *Война, которую они вели, закончилась довольно быстро.*

С точки зрения отношений между аргументом и значением функции подобные конструкции очень сильно отличаются от прототипических (*Дети задавали вопросы*). Если в простых предложениях существительное синтаксически подчиняется глаголу, то в рассматриваемых сложных конструкциях направление связи меняется на противоположное; более того, эта связь может быть и опосредованной. Синтаксическая цепочка, связывающая аргумент и значение лексической функции, достигает иногда 7 единиц (*The war they were going to begin to wage is very important.*), и вполне возможно, что это не предел.

Кроме того, естественным образом меняется характер синтаксической связи. В прототипических конструкциях это первая комплетивная связь (*Они вели ^{1-компл} войну.*), в рассматриваемых сложных предложениях придаточное присоединяется к главному релятивной связью (*Война, которую они вели, закончилась довольно быстро.*), остальная



часть цепочки, соединяющей значение и аргумент лексической функции, также состоит из определенных связей.

Еще одно отличие заключается в том, что в простых конструкциях глагол и существительное обычно стоят рядом, в сложноподчиненных предложениях же между ними находится как минимум одно слово. В принципе по сравнению с остальными различиями это не очень существенный и наиболее легко поддающийся обработке параметр, но в некоторых случаях его необходимо учитывать.

Все вышеупомянутые отличия сложных предложений, содержащих аргумент и значение лексической функции Oper1, от их простых аналогов и привели к тому, что система, довольно хорошо справлявшаяся с простыми конструкциями, не обеспечивала правильную обработку ЛФ в сложноподчиненных предложениях. Для того чтобы особенности сложных конструкций учитывались при переводе, необходимо было внести в систему некоторые изменения, которые будут описаны ниже.

В таблице 2 представлены примеры того, как «ЭТАП-3» переводил сложные предложения с придаточными относительными, в которых аргумент и значение функции Oper1 находились в разных клаузах.

Таблица 2.

№	L1	L2
1.	О битве, которую они вели, знает весь мир.	About a battle which they <u>led</u> , knows the whole world.
2.	О битве, которую они могли вести, ничего не сообщали.	About a battle which they could <u>lead</u> , nothing reported.
3.	Война, которую вели эти страны, закончилась.	The war which these countries <u>led</u> has ended.
4.	Война, которую они собирались вести, прекратилась.	The war which they were going to <u>lead</u> has stopped.
5.	Вопрос, который Вася собирался задать Пете, был трудным.	The question which Vasya was going to <u>give</u> to Petya was difficult.

Во всех представленных в таблице 2 примерах и в русском, и в английском языках существительные требуют при себе глагола поддержки, не являющийся тривиальным

вариант на этапе анализа, чем создавать правила выбора верного перевода L1 на этапе трансфера.

Таким образом, в словаре системы есть множество омонимов, в т.ч. омонимичных глаголов. Как и в обычном словаре, в «ЭТАПе» они записываются в виде СЛОВО1, СЛОВО2 и т.д. Если при переводе системе встречается такая лексема, то по умолчанию она выбирает первый омоним, если это не противоречит контексту или специальным условиям.

Среди слов, являющихся аргументами и значениями лексических функций, немало имеющих омонимы. Следовательно, для того, чтобы при переводе учитывалась их связь, необходимо, чтобы система из всех возможных омонимов выбирала те, которые связаны лексической функцией. Иначе она будет опознавать ЛФ только в том случае, если ее аргумент и значение являются омонимами № 1, что происходит далеко не всегда.

Именно правило выбора из всех возможных омонимов тех, которые связаны лексической функций, и было причиной ошибки в переводе. Как и в случае с опознанием лексической функции, это правило было рассчитано только на простые случаи и не учитывало контексты, в которых аргумент и значение функции находились в разных клаузах.

Для исправления этой ошибки понадобилось написать новое правило, которое отвечало за выбор правильных омонимов, связанных лексической функцией, в сложных контекстах. Новое правило распространяло действие старого на те случаи, когда предполагаемые аргумент и значение функции были связаны синтаксической цепочкой не более чем из 8 элементов, вершиной было существительное и непосредственное зависимое присоединялось к нему по релятивной связи. Результат работы системы после создания нового правила можно видеть в таблице 4.

Таблица 4.

№	L1	L2
1.	Война, которую они собирались вести, прекратилась.	The war which they were going to wage has stopped.
2.	Война с соседом, которую вел Петя, закончилась капитуляцией соседа.	The war with a neighbour which Petya waged has ended with a capitulation of the neighbour.
3.	Битва, которую они вели, известна всем.	The battle which they fought is known to everybody.
4.	Битва, которую они собирались вести, известна всем.	The battle which they were going to fight is known to everybody.

Таким образом, «ЭТАП-3» стал правильно обрабатывать лексическую функцию Oper1 в сложных синтаксических конструкциях.

Следует заметить, однако, что не всегда рассматриваемые сложные контексты с функций Oper стали переводиться правильно. Так, в некоторых случаях система не справляется с взаимодействием функций. Например, предложение *Война, которую они собирались начать вести...*, переводится как *The war which they were going to begin to conduct...*, поскольку лексическая функция Inceroper (*начать войну*) выявляется раньше, чем функция Oper, и вместо последней срабатывает первая. Эту ошибку тоже надо исправлять, однако такая работа выходит за рамки данного исследования.

Анализ предложений, включающих слова, связанные лексическими функциями, позволил выявить некоторые достаточно частные, но от этого не менее интересные лингвистические факты. Так, при переводе некоторых фраз возникали крайне странные ошибки (см. таблицу 5).

Таблица 5.

№	L1	L2
1.	Они вели битву в течение трех дней.	They fought <u>with</u> a battle during three days.
2.	Они вели битву с противником.	They fought <u>with</u> a battle with an adversary.
3.	Они вели битву.	They fought <u>with</u> a battle.

Причина ошибки заключалась в следующем. Второй актант английского глагола *fight* присоединяется к нему с помощью предлога *with* (*Pete fights with Tom.*). В словосочетании *fight a battle*, являющемся переводом выражения *вести битву*, слово *battle* является вторым актантом слова *fight*, следовательно, перед ним должен стоять предлог *with*. Однако это не так. Отсюда следует, что некоторые глаголы в конструкциях с лексическими функциями реализуют иную модель управления, чем в свободных словосочетаниях.

Это их свойство следует учитывать и в системе машинного перевода. Чтобы исправить ошибку в переводе, потребовалось написать правило, которое устраняло из поверхностно-синтаксической структуры выходного текста порожденные ранее предлоги в том случае, когда эти предлоги были не нужны. Правило является трафаретным, т.е. срабатывает только при появлении в предложении слов, в словарной статье которых есть указание его использовать, так что нужные предлоги оно не уничтожает.

Результат работы системы после применения этого правила отражен в таблице 6.

Таблица 6.

№	L1	L2
1.	Они вели битву в течение трех дней.	They fought a battle during three days.
2.	Они вели битву с противником.	They fought a battle with an adversary.
3.	Они вели битву.	They fought a battle.

Таким образом, в данной работе были рассмотрены некоторые конструкции, в которых используются слова, связанные лексической функцией *Oper*, и было показано, как можно обеспечить правильную обработку этих конструкций в системе машинного перевода.

Также в работе были показаны структуры, для правильной обработки которых требуется дополнительная работа. Эта работа, а также рассмотрение других особенностей конструкций с лексическими функциями, не представленных в данной статье, может стать продолжением проведенного исследования.

Библиография

- 1) [Мельчук 1974] Мельчук И.А. Опыт теории лингвистических моделей «Смысл ⇔ Текст», М., Наука, 1974. Переиздано в: Мельчук И.А. Опыт теории лингвистических моделей «Смысл ⇔ Текст». М., «Языки русской культуры», 1999
- 2) [ТКС 1984] Мельчук И.А., Жолковский А.К. Толково-комбинаторный словарь современного русского языка (Вена, 1984)