

Некоторые лексико-функциональные соответствия в русском и английском языках

Андреева А. С.
МГУ им. М. В. Ломоносова
andreyevs@mtu-net.ru

В работе рассматриваются соответствия между лексическими функциями (ЛФ) типа Func и Oper в русском и английском языках, где Func (X) - глагол поддержки, при котором X является подлежащим и Oper (X) - глагол поддержки, при котором один из актантов (не X) является подлежащим, а X – главным дополнением. Работа проводится с использованием словаря системы машинного перевода «ЭТАП-3».

В наборах русских и английских слов, имеющих лексические функции определенного вида, наблюдаются существенные различия. Так, например, слов, имеющих лексическую функцию Func, в английском языке примерно в 2 раза меньше, чем в русском.

Поскольку функции Func и Oper являются «обратными» друг к другу, то в русском и английском языках между ними имеются закономерные соответствия. В качестве одного из довольно простых соответствий можно привести тот факт, что у всех названий болезней и недугов в русском языке есть лексическая функция Func1 (*быть у – у него был тяжёлый грипп*), а в английском - Oper1 (*have / suffer from – he had serious flu*).

В связи с этим возникает задача выделить как можно более широкие лексико-семантические классы предикатных слов в обоих языках, соответствие между которыми можно задать некоторой парой функций (Func1 в русском - Oper1 в английском, Func2 в русском – Oper2 в английском и т.д.), и сформулировать достаточно общие правила, которые можно было бы успешно использовать в системе автоматического перевода.

В качестве дополнительной задачи рассматривается внесение этой информации в систему ЭТАП-3 и улучшение таким образом ее работы.

В данной работе рассматриваются соответствия между лексическими функциями типа Func и типа Oper в русском и английском языках. Работа проводится с использованием словаря системы машинного перевода «ЭТАП-3».

Лексические функции (ЛФ) – особые отношения, «характеризующие идиоматичную взаимозаменяемость и сочетаемость лексем» [ТКС 1984]. Лексические функции очень удобны для применения в формальных системах обработки естественного языка, в т.ч. в системах машинного перевода. В частности, они широко используются в системе «ЭТАП-3». В данной работе я сосредоточусь на двух классах лексических функций: классе Oper и классе Func.

Func (X) обозначает глагол поддержки, при котором X является подлежащим. Существует 4 типа Func'ов:

1. Func0(X) - X имеет место (глагол поддержки, при котором X выполняет функцию подлежащего).

Примеры:

- АВАРИЯ: происходить / случаться
- БАНКЕТ: состояться¹
- ВОЙНА: идти

1. Func1(X) - X имеется у P1¹ или характеризует P1 (глагол поддержки, при котором X выполняет функцию подлежащего, а P1 - функцию главного дополнения).

Примеры:

- ВЛАСТЬ: быть <y1>
- БЕДА: случаться <c2> / стрястаться <c2>
- СТРАХ: терзать / владеть

2. Func2(X) - X воздействует на P2, касается P2 или состоит в P2 (глагол поддержки, при котором X выполняет функцию подлежащего, а P2 - функцию главного дополнения).

Примеры:

- ВОЗМЕЗДИЕ: настигать
- РАСПОРЯЖЕНИЕ: касаться
- АГРЕССИЯ: направлять <против>

3. Func3(X) - X воздействует на P3, касается P3 или состоит в P3 (глагол поддержки, при котором X выполняет функцию подлежащего, а P3- функцию главного дополнения).

Примеры:

- ПРИКАЗ: предназначаться (кому-либо)
- ВЫВОД: следовать <из> / вытекать <из>

Oper (X) обозначает глагол поддержки, при котором один из актантов (не X) является подлежащим, а X – главным дополнением. Существует 3 типа Oper'ов:

1. Oper1(X) - делать X, иметь X или быть в состоянии X (глагол поддержки, при котором P1 выполняет функцию подлежащего, а X - функцию главного дополнения).

Примеры:

- ПОПУЛЯРНОСТЬ: пользоваться
- ОТЧАЯНИЕ: быть <в2> пребывать <в2> / предаваться
- АВАРИЯ: терпеть / потерпеть
- АГРЕССИЯ: совершать / осуществлять

2. Oper2(X) - подвергаться X-у или быть в области действия X-а (глагол поддержки, при котором P2 выполняет функцию подлежащего, а X - функцию главного дополнения).

Примеры:

- ПРИЕМ: встречать
- УВАЖЕНИЕ: пользоваться / вызывать² / внушать
- АГРЕССИЯ: подвергаться

3. Oper3(X) - подвергаться X-у или быть в области действия X-а (глагол поддержки, при котором P3 выполняет функцию подлежащего, а X - функцию главного дополнения).

Примеры:

- ДОКЛАД: быть <на2>/слушать (*Мы были на докладе*)
- ПРИКАЗ: получать <от>
- СООБЩЕНИЕ: получать

Описания лексических функций приводятся по справочной документации к системе машинного перевода «ЭТАП-3» и по статье [Aprésian et al 2003].

¹ PN – участник ситуации номер N. Т.е. P1 – первый, P2 – второй, P3 – третий.

Из описания функций видно, что Func и Orig являются «обратными» друг к другу, поэтому неудивительно, что в русском и английском языках между ними имеются регулярные соответствия. Задачей данной работы является выделение как можно более широких лексико-семантических классов предикатных слов в обоих языках, соответствие между которыми можно задать некоторой парой функций из приведенного выше списка. На материале таких классов могут быть сформулированы достаточно общие правила, подходящие для использования в системе автоматического перевода

Я ограничусь рассмотрением нетривиальных пар, состоящих из разных функций. Существует огромное количество слов, являющихся переводами друг друга и имеющих в русском и в английском языке идентичный набор ЛФ вида Func и Orig. С теоретической точки зрения эти соответствия по большей части малоинтересны, а их автоматическая обработка не представляет особой сложности. Те же случаи, когда наборы лексических функций у русских и английских слов определенной группы различаются, дают возможность для интересных теоретических наблюдений, кроме того, требуют нестандартных технических решений для их правильной автоматической обработки.

Прежде всего стоит заметить, что слов, имеющих лексическую функцию Func, в русском языке примерно в 2 раза больше, чем в русском. При этом практически не бывает так, чтобы у английского слова была функция типа Func, а у его русского перевода ее не было. Т.е. примерно в половине случаев, в которых некий смысл передается в русском языке функцией из группы Func, английский язык вынужден обходиться другими средствами.

Func0

Первое из таких средств – использование вместо значения лексической функции (по определению нестандартного) обычного бытийного глагола. Это очень характерная ситуация для тех английских слов, у русских аналогов которых имеется функция Func0 (X имеет место). В то время как русский язык использует различные слова, часто являющиеся метафорами движения или положения в пространстве (*стоять, идти, проходить* и т.п.), английский язык считает, что вполне можно обойтись простым глаголом *to be*.

В частности, именно так дело обстоит у группы слов, обозначающих отрезки времени. Сюда относятся слова *время* и *период*, а также названия времен года, месяцев, дней недели, частей дня (*утро, вечер, день, ночь*) и других отрезков времени от *секунды* до *тысячелетия*. Из этих слов времена года, месяцы и части дня могут *стоять* и *идти*, все остальные – только *идти*.

Примеры:

- *Стоял холодный февраль.*
- *Шел 1995-й год.*

В техническом отношении со словами этой группы поступают следующим образом. При переводе с русского языка на английский оба глагола заменяют глаголом *to be*. При переводе в обратную сторону лексическую функцию не восстанавливают: фразы типа *Был холодный февраль* и *Был 1995 год* звучат вполне нормально, а на превращение их в более красивые пришлось бы потратить массу усилий при минимальном улучшении качества текста.

К группе слов, обозначающих отрезки времени, примыкает другая группа – слова, обозначающие события, продолжающиеся некоторое время. Это такие слова как *событие, инспекция, акция, антракт, аукцион, банкет, бомбардировка, заседание, захват, зачет* (в значении *экзамен*) и другие². Как и в первой группе, у них в русском языке есть лексическая

² Стоит отметить, что группы слов, имеющих набор лексических функций определенного вида, обычно вместе с некоторым рядом понятий включают и их гипероним: *время* и *период*

функция Func0, а в английском нет никакого аналога. В отличие от первой группы, в качестве значения функции выступают не только глаголы движения (*идти, проходить*) но и производные от них (*состояться, происходить*).

Примеры:

- *Захват дворца произошёл очень быстро.*
- *Аукцион состоится 25 марта.*
- *Бомбардировки шли трое суток.*
- *Зачет проходит в 2 этапа: первый этап – студенты заполняют зачетки, второй этап – преподаватель проходит по рядам и расписывается.*

Ситуация с английскими аналогами у этой группы тоже немного сложнее, чем у первой. На месте значения лексической функции у них может использоваться глагол *to be*, но наряду с ним употребляется и словосочетание *to take place*. Статус его при этом остается весьма туманным. С одной стороны, он вполне отвечает определению значения лексической функции Func0. С другой стороны, значением лексической функции, во всяком случае в рамках принятых в системе ЭТАП-3 решений, может быть одно слово, в крайнем случае – слово и предлог, но не словосочетание, состоящее из 2-х полнозначных слов. Таким образом, с теоретической точки зрения можно считать, что у части слов второй группы лексическая функция Func0 имеется и у русских, и у английских слов, однако отразить этот факт в системе автоматического перевода именно с помощью механизма лексических функций весьма трудно.

Что касается автоматической обработки, то с этой группой слов можно поступить следующим образом. При русско-английском переводе все русские глаголы превращаются в *take place*. При работе в обратном направлении при наличии в тексте выражения *take place* порождается тот русский глагол, который является значением лексической функции Func0 соответствующего слова³. Если же в английском тексте присутствует глагол *be*, как и в предыдущем случае, он переводится как *быть*.

Следует заметить, что в описываемую группу попадают не все слова, обозначающие события, продолжающиеся некоторое время. Как уже было отмечено выше, в данной работе рассматриваются только такие лексемы, у переводов которых лексическая функция определенного вида отсутствует. Существует довольно большая группа слов рассматриваемого типа, у которых ЛФ Func0 имеется и в русском, и в английском языках. Во-первых, это слова, английские варианты которых сочетаются с *to be on*⁴. (Пример – *The discussion is on.*) Во-вторых, всякие нежелательные происшествия, которые по-русски могут *случаться* или *происходить*, а по-английски, соответственно, *occur* или *happen*. Это слова типа *авария, происшествие, катастрофа* и т.п. Правда, у последних в качестве компенсации имеется приятный сюрприз. У английских слов этого ряда, помимо ЛФ Func0, есть Func2 – *occur to*. Это один из немногих случаев, когда лексическая функция типа Func имеется в английском языке и отсутствует в русском.

для группы слов, обозначающих отрезки времени, *событие* для событий, продолжающихся некоторое время, *погода* для группы погодных явлений и т.д.

³ Можно было бы порождать в русском языке для этих случаев конструкцию *X имеет место* – это намного легче, чем восстанавливать значение лексической функции, – но, по моему мнению, эта конструкция слишком искусственная, чтобы допустить ее появление в выходном тексте.

⁴ Хотя следует заметить, что сочетание типа *X is on*, вообще говоря, употребляется реже, чем *X takes place*, и поэтому, может быть, стоит вообще не считать *to be on* за лексическую функцию, т.к., если оно таким образом оформлено в словаре, то будет автоматически порождаться при наличии Func0 в русском предложении.

Пример:

- *A calamitous catastrophe occurred to Eli's house. [3]*

Еще одна группа слов, имеющих в русском языке лексическую функцию Func0 и не имеющих ее в английском, включает в себя лексемы, описывающие погоду. Это слова типа *погода, ведро, жара, мороз, холод*. В русском языке они сочетаются с глаголом движения *стоять*, а английский язык считает, что простого бытийного глагола *to be* им вполне достаточно.

Примеры:

- *Стояла жуткая жара.*
- *Все лето стояла хорошая погода.*

При автоматической обработке с этими словами можно поступать так же, как со словами из первой группы (обозначающими отрезки времени): с русского на английский переводить с использованием глагола *be*, при обратном переводе лексическую функцию не восстанавливать:

Стояла хорошая погода – The weather was good. – Была хорошая погода.

Стояла жара. – There was heat. – Была жара.

Интересно заметить, что слова, обозначающие разные виды осадков (*град, дождь, ливень, снег*), тоже имеют лексическую функцию Func0, правда, и в русском, и в английском языке. Однако у них есть своя интересная особенность. Если в русском языке для разных слов используются разные глаголы, то в английском разнообразия намного меньше – почти у всех слов этой группы одинаковые значения ЛФ Func0..

Примеры:

- *Падает град. The hail falls.*
- *Идет дождь. The rain falls.*
- *Выпал снег. The snow fell.*
- *Метет метель. The blizzard falls.⁵*

То же самое наблюдается у слов, обозначающих чувства и ощущения. И в русском, и в английском языке они имеют лексическую функцию Func0, причем в английском в качестве ее значения выступает ряд глаголов, каждый из которых может сочетаться со всеми или с большинством членов группы (*overtake, overcome, overpower*) а в русском практически для каждого слова используется свой оригинальный глагол: грусть одолевает или снедает, дрожь бьет, зависть гложет, гнев владеет, дрема и сон охватывают, а жаба душит.

Из всех функций вида Func и Oper ЛФ Func0 является самой простой. У нее у единственной есть только одна переменная – сам X. Именно поэтому при наличии этой функции у слова в русском языке существует только 2 варианта в английском: либо она там тоже есть, либо ее нет. Функция Func0 не может соотноситься с другими функциями типа Func или Oper, т.к. у всех у них по 2 переменных. А вот у всех остальных функций – вида FuncX и OperX, где X>0 – возможны более интересные сочетания.

Func1 и Oper1

Самой распространенной комбинацией функций типа Func и Oper является, пожалуй, группа следующего вида:

⁵ Правда, еще можно сказать *The blizzard blows*, поскольку метель является не столько осадками, сколько ветром, а слова, обозначающие разные виды ветра, в качестве значения ЛФ Func0 имеют глагол *дуть/blow*.

Таблица 1.

русская лексема	английская лексема
Func1:быть<y1>/иметься<y1> Oper1:иметь	Oper1:have

Такой набор лексических функций имеют очень многие слова. Единственным, что их объединяет, является тот факт, что называемые ими объекты можно иметь. Это и предметы (*билет, купон*), и результаты мыслительной деятельности (*идея, план, лицензия*), и разного рода характеристики личности (*обаяние, опыт, принцип, репутация*), и просто абстрактные понятия (*возможность, право*).

С технической точки зрения проблему в данном случае создает перевод конструкций типа «У X-а есть Y». Для того, чтобы он работал, необходимо специально правило, заменяющее глагол и переставляющее актанты. Что касается перевода конструкций с глаголом *иметь* (в обе стороны), то об этом я скажу чуть позже.

В таблице 1 приведен наиболее частотный для данной группы набор лексических функций и их переменных, однако он далеко не единственный. У некоторых русских слов описываемой категории отсутствует Oper1. Таких слов немного, и обозначают они абстрактные понятия (*мотив, достижение*). Вряд ли их стоит выделять в какую-то особую семантическую группу. Для большинства русских слов, с которыми возможны конструкции *У меня есть X* и *Я имею X*, первая является предпочтительной. Вторая при этом начинает просто отмирать за ненужностью. Таким образом, слова можно распределить по шкале в зависимости от того, на какой степени отмирания находится конструкция *Я имею X*. Например, *Я имею возможность* – абсолютно нормальная фраза, *Я имею подозрение* – уже хуже, а *Я имею достижения* – совсем плохо. При этом *Я имею X* практически никогда не бывает лучше, чем *У меня есть X*.

В связи с вышесказанным возникает следующая ситуация. Все русские слова данной группы имеют лексическую функцию Func0. Также все английские слова данной группы имеют лексическую функцию Oper1. Большинство русских слов также имеют ЛФ Oper1, причем для некоторой их части конструкция с ЛФ Func0 предпочтительнее.

В любом случае придется использовать правило, переводящее конструкцию *X has Y* в конструкцию *У X-а есть Y* – для тех слов, у которых ничего другого нет. Для всех прочих слов существует 2 возможности организовать перевод с английского на русский: 1) как обычно, т.е.: если в русском языке ЛФн(Xp)=Y, а в английском языке ЛФн(Xa)=Z, то в предложении, где а) Xp переводится как Xa, б) имеется Y и с) ЛФн(Xp)=Y, Y будет переводиться как Z, - или 2) как у слов, не имеющих функции Oper1 в русском языке.

Первый вариант хорош легкостью реализации и своей «нормальностью»: ровно для такой организации работы и создавалась вся система лексических функций. Второй вариант имеет свои преимущества: во-первых, он будет единым для всех слов данной группы, во-вторых (что более важно), при его использовании мы получим более правильные предложения, чем если мы будем следовать общей схеме. Поэтому, я думаю, следует предпочесть второй вариант. В том, что он будет работать не так, как все остальные лексические функции, нет ничего страшного. Нужно будет просто дать специально указание системе, что в данном случае используется особая схема.

Например, предложение *This child has rich imagination* следует переводить не как *Этот ребенок имеет богатое воображение*, а *У этого ребенка богатое воображение*.

Помимо слов, имеющих в качестве значений функций Func1 и Oper1 бытийные глаголы и глаголы обладания, существуют, естественно, и слова с другими значениями этих функций. Их не так много, поэтому нет смысла разбивать их на большое количество групп. Собственно говоря, имеет смысл выделить слова, обозначающие болезни (поскольку их примерно в 2 раза

больше, чем остальных слов данного типа), а оставшиеся поместить в группу «все остальное».

В эту вторую группу попадают прежде всего объекты, в обладании которыми возможна динамика – т.е. те, которые могут не только иметь хозяина, но и менять его. Это объекты, которые можно *подавать* (*жалоба, заявка, заявление*), *выдавать* (*лицензия, разрешение*), *вручать* (*награда*). Разновидностью таких объектов можно считать названия запахов (от *аромата* до *вони* во главе с самим *запахом*), которые в некотором смысле тоже меняют владельца – частицы пахучего вещества попадают из одного места в другое. В эту же группу попадают некоторые чувства, которые в русском языке могут просто быть у их владельца (*интерес, сочувствие, доверие*) – поэтому с их Func1 надо что-то делать, – и небольшое количество слов, взаимодействие которых с субъектом действия, в котором они принимают участие, описывается нестандартным глаголом (*говорить с акцентом, оказывать влияние, задавать вопрос* и т.п.)

С технической точки зрения их обработка довольно проста: значение функции Oper в русском языке превращается в ее значение в английском и наоборот, а Func переводится в Oper с помощью специального правила, подробное описание которого имеется выше. Если же у русского слова функция Oper отсутствует, то приходится делать 2 специальных правила: одно для превращения русского Func'a в английский Oper, другое для обратной операции. Точно таким же способом обрабатываются слова, относящиеся к группе «болезни». В нее входят названия собственно болезней (*грипп, коклюш*), болезненных состояний (*кашель*) и ненормальных состояний частей организма (*рана, опухоль*). У большинства из них имеется лексическая функция Func1 в русском языке (*быть у*) и функция Oper1 в английском (*have / suffer1 from*). Примерно у четверти есть еще русский Oper1 – *болеть* или *страдать2*.

Единственное слово, выбившееся из общего ряда – это гипероним, *болезнь*. У нее, помимо полного набора (Func1 и Oper1 в русском языке и Oper1 в английском) есть еще русский Func0 – *течь, протекать*. Таким образом, слово *болезнь* попадает сразу в 2 группы: группу «болезни» и группу «события, продолжающиеся некоторое время». Кроме нее, существует еще очень небольшое количество слов, попадающих более чем в одну группу. Например, слово *каникулы*: оно относится и к объектам, которые можно иметь, и к событиям, продолжающимся некоторое время.

Func2 и Oper2

В русском и английском языках имеется небольшая группа слов, представляющих интерес в связи с наличием у них непарных функций Func2 и Oper2. Это существительные, которые описывают ситуацию, имеющую как минимум 2-х участников, действия одного из которых направлены на другого⁶. При этом почти у всех этих слов есть либо отрицательная коннотация, либо отрицательный элемент значения – в общем, что-нибудь отрицательное. В данную группу входят лексемы *агрессия, амнистия, запрет, инструкция, кара, критика, наказание* и, возможно, некоторые другие. У всех этих слов в русском языке имеется лексическая функция Func2 (*Кара настигла бандитов, Агрессия направлена против сербов*) и Oper2 (*Вы получили четкие инструкции, Фильм вызвал критику религиозных организаций*). При этом в английском языке у половины слов есть функция Oper2 (*запрет, инструкция, кара, наказание*), у остальных нет ничего.

⁶ Конечно, вообще таких слов в языке довольно много, но интерес для данной работы представляют только те из них, у которых имеется непарный набор лексических функций вида Func2 и Oper2. Слова, имеющие в обоих языках Func2, или Oper2, или обе эти функции, в рассмотрении не участвуют.

С теми четырьмя, у которых в английском словаре ЭТАПа-3 есть Oper2, можно поступать так же, как в случае с парой Func1 – Oper1. Значения функций Oper являются переводами друг друга, а русский Func2 пересчитываются в английский Oper2.

Со второй половиной слов все сложнее.

В таблице 2 даны значения лексических функций Func2 и Oper2 для этих трех существительных.

Таблица 2.

№	русское слово	лексические функции русского слова	английское слово
1.	агрессия	_func2:направлять <против1> _oper2:подвергаться	aggression
2.	амнистия	_func2:касаться / затрагивать _oper2:подпадать <под1>	amnesty1
3.	критика	_func2:направляться <на1> _oper2:подвергаться/вызывать2	criticism critique1

В этой группе для каждого слова приходится принимать отдельное решение. *Критику* и *амнистию*, скорее всего, нужно заменять на глаголы – *amnesty* и *criticize*, соответственно полностью перестраивая всю структуру предложения. Слово *агрессия* на глагол не заменяется. С ним придется поступать следующим образом. Oper2 переводить как *to be a victim of*, а Func2 заменять на конструкцию типа *It was aggression against...*

Что касается упомянутых в описании функций Func3 и Oper3, они встречается довольно редко и чаще всего бывают парными.

Итак, в данной работе были рассмотрены русские и английские слова, имеющие лексические функции вида Func и Oper. На основе анализа материала были выделены группы слов, обладающих одинаковыми несимметричными наборами функций вышеназванных типов. Также были предложены способы обработки этого материала в системах автоматического перевода.

В качестве продолжения данной работы предполагается рассмотреть лексические функции типа Incerfunc, Inceroper, Finfunc и Finoper, обозначающих начало и конец ситуаций, описываемых функциями вида Func и Oper соответственно.

Литература

1. [ТКС 1984] Мельчук И.А., Жолковский А.К. Толково-комбинаторный словарь современного русского языка (Вена, 1984)
2. [Apresian et al 2003] Jurij Apresian, Igor Boguslavsky, Leonid Iomdin, Alexander Lazursky, Vladimir Sannikov, Victor Sizov, Leonid Tsinman) ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning – Text Theory. Paris, Ecole Normale Superieure, Paris, June 16-18 2003, pp. 279-288.
3. www.wcg.ndirect.co.uk/Article43.htm