

МОДЕЛИРОВАНИЕ ЯЗЫКОВЫХ СТРУКТУР СО ЗНАЧЕНИЕМ УСЛОВИЯ И ПРИЧИНЫ ДЛЯ АНГЛО-РУССКОГО ЛИНГВИСТИЧЕСКОГО ПРОЦЕССОРА

Е.Б. Козеренко

Институт проблем информатики РАН

Россия, 117333, Москва, ул. Вавилова, 44 корп. 2

E-mail: kozerenko@mail.ru

Ключевые слова: лингвистический процессор, представление знаний, расширенные семантические сети, функционально-семантическое поле, взаимодействие синтаксиса и семантики, машинный перевод.

Данная работа посвящена вопросам моделирования языковых структур, значение которых относится к функционально-семантическим полям с обстоятельственным ядром, а именно – условия и причины. Эти поля выделяются, прежде всего, на базе группировки сложноподчиненных предложений обстоятельственного типа; вместе с тем, аналогичная семантика выражается и средствами простого предложения, деепричастными, причастными оборотами. В составе функционально-семантического поля могут выступать разнообразные “рассеянные” грамматические и лексико-грамматические средства. Построение функционально-семантических представлений – плодотворная основа для создания лингвистического процессора. На этапе формализации используется система представлений, сочетающая в себе черты унификационной и порождающей грамматик. Декларативное представление лингвистического процессора английского и русского языков предназначено для задач извлечения знаний из неструктурированных текстов (в нашем случае использовался нарративный дискурс научной прозы) и машинного перевода.

1. Введение

Внимание к вопросам моделирования языковых структур, значение которых относится к функционально-семантическим полям с обстоятельственным ядром, а именно – условия и причины объясняется их важностью при построении систем управления знаниями, поскольку на основе естественно-языковых конструкций со значением условия и причины строится система продукционных правил для некоторой предметной области. Эти поля выделяются, прежде всего, на базе группировки сложноподчиненных предложений обстоятельственного типа; вместе с тем, аналогичная семантика выражается и средствами простого предложения, деепричастными, причастными оборотами [1-2,4,6-8]. В составе функционально-семантического поля могут выступать разнообразные “рассеянные” грамматические и лексико-грамматические средства (отдельные формы, синтаксические и лексико-синтаксические конструкции) [1,5]. Исследования языкового материала проводятся в двух направлениях: от семантики – к форме и от формы – к семантике. При нашем анализе в центре внимания находится соотношение грамматических единиц и контекстов, в которых они реализуются.

Как было показано в [3], кодирующая техника языка в значительной мере детерминирована глубинной семантической структурой, и значительные преимущества имеет такой способ описания, при котором в качестве отправного пункта принимается семантический уровень и определенным семантическим единицам сопоставляются выражающие их кодирующие средства.

Построение функционально-семантических представлений – плодотворная основа для создания лингвистического процессора. На основе выделяемых значений происходит формирование единого семантического пространства, где средства выражения сходных значений в разных языках рассматриваются как синонимические средства.

На этапе формализации используется система представлений, сочетающая в себе черты унификационной и порождающей грамматик [6-7,9-12]. Результатом этапа формализации является декларативное представление

лингвистического процессора английского и русского языков для задач извлечения знаний из неструктурированных текстов (в нашем случае использовался нарративный дискурс научной прозы) и машинного перевода.

Предметом анализа являются связи временной соотнесенности действий с семантикой обусловленности и отношениями характеристики. Рассматриваются особенности взаимодействия доминирующих аспектуально-таксисных значений и отношений обусловленности между действием, обозначаемым основным глаголом в личной форме и второстепенной глагольной формой. В функциональной грамматике в семантическую сферу обусловленности включают отношения причины, цели, условия, уступки, следствия. В фокусе внимания нашего исследования находятся отношения условия и причины.

2. Средства выражения условной и причинной обусловленности аспектуально-таксисной ситуации

Выделяются и анализируются полипредикативные комплексы, представленные в тексте в виде сложноподчиненных предложений условия и причины, конструкций с неличными формами глаголов (причастиями, деепричастиями, отглагольными существительными, герундиями) а также отдельными формами и лексико-синтаксическими конструкциями, передающими аспектуально-таксисные ситуации, связанные с семантикой обусловленности (причинной и условной).

Второстепенное действие (придаточное предложение условия, причины; деепричастие; причастие) не только сопряжено во времени с основным действием, выраженным личной формой глагола в главном предложении, но и обуславливает осуществление основного действия. Выражение значения обусловленности является семантическим эффектом взаимодействия таких составляющих как лексические, видовые значения личной формы глагола в главном предложении и глагольной формы, выражающей второстепенное действие, а также порядка слов и вспомогательных элементов – подчинительных союзов, предлогов. В сложноподчиненных предложениях обусловленность и связь во времени выражаются дискретно – посредством соотнесения видовых форм сказуемых в главном и придаточном предложениях, а также с помощью союзов и союзных слов, например:

If we come to this point, we shall speak only in terms of numerical values.

Обусловленность может быть выражена эксплицитно (*если...то*), имплицитно (*при переходе...*), размыто (*...вздрыгнул, когда услышал...*). Последний случай формально неотличим от простой временной соотнесенности:

Когда солнце встало, работа была сделана.

В данном случае восход солнца не является причиной основного действия.

Второстепенная глагольная форма, представленная препозитивным деепричастием совершенного вида, часто совмещает значения причинной и временной обусловленности. При постпозитивном деепричастии совершенного вида преобладает причинное значение:

Они не стали продолжать эксперимент, установив причину эффекта.

Рассматриваемые аспектуально-таксисные ситуации можно отнести к следующим видам:

- 1). Основное действие обусловлено второстепенным действием с признаком перфектности (например, в русском языке - деепричастием совершенного вида, в английском языке – перфектным причастием или предложением *after* в сочетании с причастием простым) и непосредственно следует за ним;
- 2). Основное действие – процесс (выражено глаголом несовершенного вида) обусловлено предшествующим фактом (*Убрав все оборудование, исследователи прекратили работу*);
- 3). Основное действие происходит по причине и на фоне второстепенного процесса (*The rate of the reaction was low with water being cooled*);
- 4). Основной процесс происходит по причине второстепенного процесса и одновременно с ним (*If being frozen, water becomes a colourless solid*);
- 5). Основное действие совершается по причине повторяющегося действия, при этом они совмещены в едином периоде времени (*Provided the network is trained repeatedly, the recognition results would be accurate*).

При рассмотрении отношения условия установлено следующее: в роли второстепенного глагольного элемента, выражающего условие по отношению к основному, в русском языке обычно употребляются деепричастия совершенного вида, выражающие действие, достигшее предела, в английском – конструкции с перфектными формами причастия или причастием вторым с перфектной импликацией. Основное действие, обусловленное другим, либо относится к будущему, либо является виртуальным (гипотетическим), либо нереальным.

Основным лингвистическим материалом исследований были следующие языковые конструкции:

1). Полностью оформленные сложноподчиненные предложения с придаточными причины и условия используют подчинительные союзы «так как», «если», «после того как», например, «Если воду охладить, скорость реакции снижается»; “*If water is cooled, the rate of the reaction is low*”. Полнооформленные придаточные предложения условия могут вводиться подчинительными союзами:

if - если; (“*If the temperature is low the reaction will stop*”);

unless – если...не (к глаголу) (“*Unless otherwise stated, the value is constant*”);

provided (providing that) – при условии, если ... (“*Two-layer networks with sigmoidal hidden units can approximate arbitrarily well any functional continuous mapping from one finite-dimensional space to another, provided the number of hidden units is sufficiently large*”).

При этом союз, присоединяющий главное предложение, выражающее следствие (*so, then*), часто может отсутствовать.

Придаточные причины вводятся союзами: *as* – так как; *since* – поскольку; так как; потому что.

2). Конструкции с неличными формами глаголов (“*with water being cooled, the rate of the reaction was low*”; “*the other conditions being equal, the acceleration will be the same*”). При переводе этих конструкций следует такое наблюдение: если в предложении с зависимым причастным оборотом сказуемое стоит в пассивной форме, то оборот иногда удобно переводить неопределенно-личным предложением, а не повторять подлежащее всего предложения: “*Unless otherwise stated, the value will be considered constant*” «Если не указано иначе, значение будет считаться постоянным»). Особое внимание уделялось сочетаниям предлогов, союзов с неличными формами глаголов: “*With the isomerization preceding the reaction, the yields were very low*” (перевод – с придаточным предложением: «Поскольку изомеризация предшествовала реакции, результирующий выход был очень низким»); “*Due to the central atom being screened, the effect was obtained*” («Благодаря тому, что центральный атом был экранирован, эффект был получен»).

Наиболее сложными с точки зрения компьютерного моделирования являются ситуации, когда отношение причинной обусловленности передается только предлогами (*при, with*) в сочетании с существительным (т.е. предложными фразами), но при этом важным семантическим фактором, влияющим на ситуацию условной и причинной обусловленности, является семантика существительного, к которому относится данный предлог: это либо отглагольное существительное, либо существительное, имеющее в данном употреблении процессную или акторную импликацию («*при температуре ниже 0 градусов вода замерзает*» - т.е. имеется в виду, что температура ниже 0 градусов каузирует замерзание воды). Таким образом при постановке задачи «глубинного понимания», а именно если необходимо отразить во внутреннем представлении суть семантических различий фразы типа приведенной выше и фразы «*при театре работает студия*» необходима семантическая грамматика, которая позволит отобразить отличия формально неотличимых структур. На данном этапе исследований это удастся сделать для отглагольных существительных, для всех остальных случаев необходима кропотливая ручная работа по семантической разметке словарных входов. Для существительных типа «*температура*», «*давление*», и т.п. совершенно очевидно, что их параметрический характер и дает семантический эффект каузации.

3. Значения реальности обуславливающего действия

Ситуации условной и причинной обусловленности необходимо рассматривать в сочетании со значениями реальности/нереальности второстепенного действия. В нашей модели исходной точкой определения типов реальности служит система английского языка, и выделяется 3 типа значений:

1). Реальное условие – сказуемые в обоих (главном и придаточном) предложениях стоят в изъявительном наклонении (переводятся без частицы «бы») в любом времени, например:

If we raised temperature, the reaction became more active. – Если мы поднимали температуру, реакция протекала более активно.

If we read much, we know much. – Если мы много читаем, мы знаем много.

We shall make the experiment if we get raw materials. – Мы проведем этот эксперимент, если получим сырье.

2). Маловероятное условие (но теоретически еще возможное) - сказуемые в обоих предложениях стоят в сослагательном наклонении в форме Indefinite, и это означает, что предполагаемое действие еще может произойти, например:

If we worked harder, we could finish the project on time.

3). Нереальное условие – сказуемые в обоих предложениях стоят в сослагательном наклонении в форме Perfect, это означает, что предполагаемое действие уже произойти не может, например:

If he had worked harder, he would have finished the project on time. – Если бы он работал настойчивее (тогда), он закончил бы проект в срок.

В русском языке сослагательное наклонение имеет только одну форму – глагол в прошедшем времени с частицей «бы» - поэтому условные предложения второго и третьего типа переводятся одинаково, и, чтобы передать разницу между (2) и (3) можно вводить поясняющие слова, например, «этого не произошло...», и т.п. В условных предложениях 2 –го и 3-го типа “if” иногда опускается, и на его место выносятся глаголы: “had”, “could”, “were”, “should”. В этих случаях единственным формальным признаком, отличающим такие предложения от вопросительной формы, служит отсутствие вопросительного знака. При переводе надо ввести подчинительный союз «если». В русском языке тоже возможно бессоюзное присоединение условных придаточных предложений, например:

Had I time, I should go there. – Имей я время, я бы пошел туда.

Were the speed of the rocket equal to that of light, its mass would be infinite. – Будь скорость ракеты равной скорости света, ее масса была бы бесконечной.

Однако, бессоюзное присоединение, уместное в разговорной речи, в стиле научной прозы должно быть заменено на союзное.

This fact points to the postulated relationship provided no inversion has taken place.

This is the case providing that the chain molecules are shortened. – Это тот самый случай при условии, если цепочки молекул укорочены.

Моделирование значений реальности/нереальности в большей степени имеет значение для задач перевода, поскольку для задач извлечения знаний из текстов система представления этих значений мало разработана.

4. Декларативные представления на основе расширенных семантических сетей

Компьютерное моделирование описываемых структур проводилось в рамках проектов по созданию интеллектуальной среды обработки знаний (ДИЕС, ИКС, ЛОГОС-Д) [6-8], а также экспериментальной системы машинного перевода. Декларативное представление правил синтаксического разбора строилось на основе гибридной унификационно-порождающей грамматики, на глубинно-семантическом уровне условные конструкции представлены в виде продукционных правил на языке расширенных семантических сетей.

Примеры синтаксических правил для некоторых конструкций:

SYNRULE S/if_cl: S{fin+, sub_conj+} -> S{fin+,sub_conj+} PUNCT+ S{fin+,cl+}

Since there is a continual loss of micrometeoritic space because of the radiation effects, there must be a continual replenishment: otherwise micrometeorites would have disappeared from interplanetary space.

Поскольку имеет место постоянное сокращение микрометеоритного пространства по причине радиационных эффектов, должно быть постоянное пополнение: иначе микрометеориты исчезли бы из межпланетного пространства.

(В связи с тем, что постоянно происходит сокращение микрометеоритного пространства, вызванное радиационными эффектами, по всей вероятности, происходит его постоянное пополнение, в противном случае, микрометеориты уже давно исчезли бы из межпланетного пространства).

SYNRULE S/vppt_if: S{fin-,sub_conj-} -> S{fin-,sub_conj-} PUNCT+ S{fin+,cl+}

Debugged, the program ran

Исправленная, программа заработала.

(Поскольку она была исправлена, программа заработала).

При этом используется подход «синтаксической нормализации» [6,7]. Суть этого подхода состоит в том, что структуры с неличными формами глаголов приводятся к форме сложноподчиненных предложений с придаточными условия и причины, что позволяет отобразить их в глубинно-семантические пропозициональные структуры.

В парадигме аппарата расширенных семантических сетей (РСС) пропозициональные структуры представляются в виде фрагментов семантической сети – n -местных именованных предикатов вида $R(A_1, A_2, \dots, A_n/x)$, позволяющих отразить объекты произвольной степени вложенности. Условные и причинные зависимости формулируются в виде продукционных правил вида IF x_1, \dots, x_n THEN y_1, \dots, y_n , которые автоматически генерируются из естественно-языковых структур со значением условия и причины в результате анализа текста, поступающего на вход лингвистического процессора.

Функции машинного перевода поддерживаются декларативным синтаксическим процессором, обеспечивающим трансфер с одного естественного языка на другой.

5. Заключение

Представляется полезным в полной мере осознавать, что применительно к компьютерному моделированию «глубинное понимание» - это метафора. Понимание и интерпретация – это прерогатива человеческого интеллекта, а компьютерные модели «понимания» - это ни что иное как более или менее аккуратные и адекватные системы последовательных вычислительно-лингвистических представлений и их преобразований.

Задача извлечения знаний из текста – это в данном случае построение продукционных правил на языке расширенных семантических сетей, которые описывают некоторый фрагмент реальной действительности. Задача перевода – сопоставление структур, передающих один и тот же смысл средствами различных естественных (в данном случае русского и английского) языков. В рамках экспериментальных проектов делается попытка совместить эти задачи через глубинно-семантические представления.

Дальнейшие исследования связаны с уточнением и развитием сопоставительных функционально-семантических описаний языковых явлений в английском и русском языках и разработкой их формализованных представлений.

Список литературы

1. Бондарко А.В. Принципы функциональной грамматики и вопросы аспектологии. М.: Эдиториал УРСС, 2001г. 208 с.
2. Якобсон Р.О. Шифтеры, глагольные категории и русский глагол. // Принципы типологического анализа языков различного строя. М., 1972. С.101, 106-108.
3. Кибрик А.Е. Очерки по общим и прикладным вопросам языкознания. М.:УРСС, 2001, с.123-124
4. Звегинцев В.А. Предложение и его отношение к языку и речи. М.: МГУ, 1976. 307 с.
5. Carpenter V. The generative power of categorial grammars and head-driven phrase structure grammars with lexical rules. // Computational linguistics. 1991. 17, 3. P. 301-313.
6. Козеренко Е. Б. Функциональная семантика в компьютерных решениях//Труды Международного семинара «Диалог'2002» по компьютерной лингвистике и интеллектуальным технологиям. г. Протвино Московской обл. 2002 т.1. Стр. 218-226.
7. Козеренко Е.Б. Унифицированные категориально-функциональные представления для синтаксической разметки полнотекстового документа//Системы и средства информатики. М.: Наука, 2002 Вып.11.
8. Kozerenko E.B. Portable Language Engineering Solutions for Multilingual Processors // Proceedings of the International Conference on Artificial Intelligence IC-AI'02// CSREA Press, 2002, pp. 447-453.
9. Halliday M.A.K., Functional Diversity in Language as seen from a Consideration of Modality and Mood in English. // Foundations of Language. 1970. 6. P. 332-361.
10. Kaplan R.M., Bresnan J. Lexical-functional grammar: A formal system for grammatical representation. // The mental representation of grammatical relations. / Ed. by Bresnan. Cambridge, MA: MIT Press, 1982. P.173-281.
11. Levin B. English Verb Classes and Alternations. Chicago: University of Chicago Press, 1993.
12. Kay M. Functional unification grammar: A formalism for machine translation. // COLING-84. Stanford, CA, 1984. P. 75-78.