

РУССКИЕ ИМПЛИКАТИВНЫЕ ГЛАГОЛЫ И ИХ СЕМАНТИЧЕСКИЙ АНАЛИЗ В РАМКАХ ЛИНГВИСТИЧЕСКОГО ПРОЦЕССОРА ЭТАП-3

Рыгаев И. П. (irygaev@gmail.com)

Институт лингвистики РГГУ; Лаборатория компьютерной лингвистики Института проблем передачи информации им. А. А. Харкевича РАН, Москва, Россия

Данная работа посвящена исследованию имплицативных глаголов в русском языке и моделированию их семантики в лингвистическом процессоре ЭТАП-3. О чём речь? Прочитайте короткий текст: *"Вася соблазнил Машу поехать с ним на юг. Коля попросил Лену выйти за него на работу."* и ответьте на вопросы по нему: *"Поехала ли Маша на юг с Васей (Да/Нет/Неизвестно)? Вышла ли Лена на работу вместо Коли (Да/Нет/Неизвестно)?"*. Человеку очень легко ответить на эти вопросы. Для компьютерной программы это не так легко, т.к. прямые ответы в тексте отсутствуют, а синтаксическая структура двух предложений очень близка. Суть в семантике глаголов *"соблазнить"* и *"попросить"* - первый позволяет сделать вывод об истинности зависимого высказывания, второй - нет. Глаголы первого типа и называются имплицативными. В работе исследованы более 100 имплицативных глаголов русского языка, проанализированы их семантические и синтаксические свойства, поведение в различных контекстах, и настроены правила в лингвистическом процессоре ЭТАП-3, формирующие вывод импликации. Это позволило в явном виде представить информацию, которая неявно присутствовала в тексте, и таким образом улучшить семантический анализ текстов, в частности для целей вопросно-ответной системы.

Ключевые слова: имплицативные глаголы, семантический анализ, извлечение информации, умозаключения на основе здравого смысла

RUSSIAN IMPLICATIVE VERBS AND THEIR SEMANTIC ANALYSIS IN ETAP-3 MT SYSTEM

Rygaev I. P. (irygaev@gmail.com)

RSUH Institute for Linguistics; Laboratory of Computational Linguistics, A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

This paper is dedicated to the analysis of the implicative verbs in Russian and modeling of their semantic behavior in ETAP-3 Machine Translation System. What does that mean? Read this short text: *"Вася соблазнил Машу поехать с ним на юг. Коля попросил Лену выйти за него на работу."* (*Vasya tempted Masha to spend the holiday with him. Kolya asked Lena to come to work for him.*) and answer a couple of questions: *"Did Masha spend the holiday with Vasya (Yes/No/Unknown)? Did Lena come to work for Kolya (Yes/No/Unknown)?"* These are very easy questions for a human, but for a computer program they are not so easy, since there are no direct answers in the text and the syntactic structures of two sentences are almost identical. The key is in the meaning of verbs *"соблазнить"* (*tempt*) and *"попросить"* (*ask*) – the former allows to make an inference about the truth of the dependent proposition (at least in Russian), while the latter does not. The verbs of the former type are called implicative verbs. In this paper I analyzed more than hundred implicative verbs in Russian, determined their semantic and syntactic features and their behavior in different contexts, and implemented the rules in ETAP-3 MT System to generate the appropriate inferences. This permitted to extract and explicitly represent the information, which was implicitly contained within the text and thereby to enhance the semantic analysis of texts in ETAP-3, particularly for the question answering purposes.

Key words: implicative verbs, semantic analysis, information extraction, commonsense reasoning

Введение

Прочитайте короткий текст:

- (1) Вася соблазнил Машу поехать с ним на юг.
Коля попросил Лену выйти за него на работу.

и ответьте на вопросы по нему:

- (2) Поехала ли Маша на юг с Васей (Да/Нет/Неизвестно)?
Вышла ли Лена на работу вместо Коли (Да/Нет/Неизвестно)?

Человеку очень легко ответить на эти вопросы. Для компьютерной программы это не так легко, т.к. прямые ответы в тексте отсутствуют, а синтаксическая структура двух предложений очень близка.

Разница в семантике глаголов «*соблазнить*» и «*попросить*». Первый позволяет сделать вывод об истинности зависимого высказывания, второй - нет. Глаголы первого типа называются имплицативными. Именно они интересны для систем семантического анализа и вопросно-ответных систем, т.к. позволяют извлекать дополнительную информацию из текста.

В рамках данной работы была поставлена задача идентифицировать имплицативные глаголы русского языка и реализовать правила в лингвистическом процессоре ЭТАП-3 для формирования вывода импликаций, то есть для представления скрытой в тексте информации в явном виде.

Методы

На основе работ Л. Карттунена [Karttunen, 1971; Karttunen, 1973] об имплицативных глаголах в английском языке был составлен список из их русских аналогов, их синонимов и близких по смыслу глаголов. Этот список был дополнен при помощи поиска в Национальном корпусе русского языка глаголов с сентенциальными зависимыми и отбора тех из них, которые имеют имплицативное поведение. Использовались также другие источники.

В результате были идентифицированы более ста имплицативных глаголов русского языка. Для определения их «имплицативности» использовалась интроспекция автора. Была также предпринята попытка устроить опрос информантов, однако она провалилась. Результаты опроса получились противоречивыми, т.к. не удалось устранить влияние других факторов (контекста предложений) на оценку импликаций информантами.

Классификация

Уже в работах Карттунена отмечается, что не все глаголы ведут себя одинаково.

Во-первых, на поведение глаголов может влиять контекст – наличие или отсутствие отрицания в главной клаузе. Одни глаголы дают импликацию в обоих контекстах (*manage – суметь*), другие – только в положительном (*cause – быть причиной*), третьи – только в отрицательном (*be able – мочь*).

- (3) a. Sam *managed* to solve the problem => John solved the problem
 b. Sam *didn't manage* to solve the problem => John did not solve the problem
- (4) a. Gases *caused* the planet to warm up => The planet warmed up
 b. Gases *did not cause* the plane to warm up => ?
- (5) a. Sebastian *was able* to leave the country => ?
 b. Sebastian *was not able* to leave the country => Sebastian didn't leave the country

Во-вторых, одни глаголы дают импликацию той же полярности, что и главное предложение (*force – заставить*), другие – обратной полярности (*prevent – предотвратить*).

- (6) John forced Mary to stay home => Mary stayed home
 (7) John prevented Mary from leaving => Mary *didn't* leave

В русском языке ситуация осложняется наличием глагольного вида. Глаголы совершенного и несовершенного вида ведут себя по-разному с точки зрения импликаций (*добился – добивался*).

- (8) a. Иван добился увеличения штата сотрудников => Штат сотрудников увеличили
 b. Иван добивался увеличения штата сотрудников => ?

Кроме того было решено учитывать не только строгую импликацию, но и т.н. нестрогую или возможную импликацию, что ещё несколько усложнило задачу. Речь идёт о правдоподобных ожиданиях, таких как напрашивающиеся следствия (*invited inference*) [Geis and Zwicky, 1971] или имплицатуры [Grice, 1975].

- (9) Пете не разрешали смотреть телевизор => *Наверное*, Петя не смотрел телевизор (если явно не сказано обратное)

Таким образом, каждый из глаголов надо было проанализировать в четырёх контекстах: утвердительный совершенный, утвердительный несовершенный, отрицательный совершенный, отрицательный несовершенный. И в каждом из контекстов установить тип импликации – строгая, возможная или отсутствует и её направление – прямая или обратная.

По этим признакам глаголы были сгруппированы в 15 групп, представленных в таблице:

Классы	Прямая импликация	Обратная импликация
Фактивные ¹	детализация (<i>преуспеть</i>)	
Двусторонние имплицативы	факт (<i>случиться</i>) решение (<i>осмелиться</i>)	факт отсутствия (<i>предотвратить</i>)
Односторонние утвердительные	причина (<i>заставить</i>) свидетельство (<i>застать</i>)	забыть (<i>забыть</i>), причина отсутствия (<i>отговорить</i>)
Односторонние отрицательные	возможность (<i>мочь</i>) отсутствие противодействия (<i>разрешить</i>)	сомнение (<i>поколебаться</i>)
Имплицатурные ²	сильное намерение (<i>съездить</i>) слабое намерение (<i>попытаться</i>) старание (<i>постараться</i>)	отказ (<i>отказаться</i>)

¹ Фактивные глаголы всегда дают импликацию положительной полярности, независимо от контекста.

² Новый класс, отсутствующий у Карттунена. Имплицатурные глаголы не дают строгой импликации ни в одном контексте, только возможную.

Таблица 1. Группы имплицативных глаголов в русском языке и их связь с классами по Карттунену.

Синтаксическое разнообразие

Кроме различий в семантике, глаголы сильно отличаются синтаксически, а именно следующими признаками:

1. Чем выражено зависимое, которое надо выносить в вывод импликации?
2. Какие трансформации с ним надо произвести, чтобы превратить в полноценное предложение?

При этом синтаксические признаки не коррелируют с семантическими группами глаголов, а являются независимыми критериями классификации.

Зависимое может быть выражено:

1. Инфинитивным оборотом с союзом *чтобы* или без него.
2. Именной группой, возглавляемой отглагольным существительным, в различных падежах и с различными предлогами.
3. Предложением, присоединяемым союзом *что*.

Некоторые глаголы допускают все три варианта:

- (10) а. Ему случилось *бывать* на юге
б. Стучилось *убийство*
с. Так случилось, *что* Дима опоздал

Самый простой из этих вариантов – третий. Зависимое предложение просто выносится “наверх”, и никаких преобразований с ним не делается.

В случае инфинитивного оборота надо преобразовать глагол к финитной форме нужного времени, уточнить его вид (совершенный вид «наследуется» от имплицативного глагола, несовершенный – нет) и сформировать при нём подлежащее. Подлежащее копируется из одного из актантов имплицативного глагола – его подлежащего, прямого дополнения или косвенного дополнения.

- (11) а. *Иван* догадался открыть дверь => *Иван* открыл дверь
б. *Иван* соблазнил *Марию* поехать на юг => *Мария* поехала на юг
с. *Иван* помог *Марии* решить задачу => *Мария* решила задачу

В случае именной группы было принято решение не приводить её к глагольной группе, а оставлять именной, добавляя “сверху” конструкцию “иметь место” в соответствующем времени. Приведение к глагольной группе потребовало бы, во-первых, преобразование зависимых существительного в зависимые глагола и, во-вторых, определение вида глагола по существительному. Обе задачи не имеют простых универсальных решений.

По аналогии с инфинитивом, иногда требуется восстановить при существительном недостающий актант из подлежащего или прямого или косвенного дополнения имплицативного глагола.

- (12) а. *Иван* осуществил подстановку переменных => Имела место подстановка переменных *Иваном*
б. Иван поймал *Марию* за рисованием картин => Имело место рисование картин *Марией*
с. Иван воспрепятствовал *Марии* в защите своих прав => Не имела места защита своих прав *Марией*

В случае существительного недостающий актант может быть не только агенсом, но и пациенсом. При этом синтаксическая позиция пациенса может варьироваться в зависимости от существительного. Информация о синтаксической позиции пациенса берётся из комбинаторного словаря ЭТАП-3.

- (13) а. Иван не избежал наказания => Имело место наказание *Ивана*
а. Иван не избежал мести => Имела место месть *Ивану*
а. Иван не избежал слежки => Имела место слежка *за Иваном*

Техническая реализация

На основе полученной семантической и синтаксической классификации были приспаны соответствующие признаки глаголам в комбинаторном словаре в системе ЭТАП-3 и написаны правила на языке FORET [Цинман, 1986], формирующие вывод импликации. Правила преобразуют исходное предложение с имплицативным глаголом в сложносочинённое, добавляя к исходному предложению вывод импликации:

- (14) Иван рискнул поставить всё на кон => Иван рискнул поставить всё на кон; Иван поставил всё на кон.

В комбинаторном словаре глаголу приписывается:

1. Семантическая группа (одна из 15)
2. Расположение имплицуемого зависимого
3. Способ трансформации актантов

Далее лингвистические правила на основе этих признаков делают следующее:

1. В зависимости от семантической группы глагола и контекста его использования определяют тип импликации (строгая, возможная или отсутствует) и её направление (прямая или обратная).
2. Находят имплицуемое зависимое, копируют его со всем поддеревом в новый узел, и связывают этот узел сочинительной связью с узлом имплицативного глагола.
3. В случае нестрогой импликации добавляют вводное слово «*наверное*».
4. В случае обратной импликации меняют полярность вывода – добавляют или удаляют отрицание.
5. В зависимости от способа трансформации актантов копируют актант со всем поддеревом от имплицативного глагола к узлу вывода.

На рисунке представлена синтаксическая структура для исходного предложения «*Андрей забыл позвонить родителям*». Узлы 1-4 относятся к исходной структуре. Узлы 5-8 являются результатом работы правил вывода.



В результате на выходе получается предложение «Андрей забыл позвонить родителям; Андрей не позвонил родителям».

Обсуждение и направления развития

Результат, полученный в этой работе можно считать лишь первым этапом по описанию имплицативных свойств конструкций русского языка. Уже сейчас видны несколько направлений развития исследований:

Перенос контекстных обстоятельств. В каких случаях и какие обстоятельства надо переносить в вывод?

- (15) *Вчера в баре* Вася заставил Петю заплатить за пиво => *Вчера в баре* Петя заплатил за пиво

Учёт других имплицативных конструкций (фразовые имплицативы, предикативные прилагательные и др.):

- (16) Пётр *сдержал слово* жениться => Пётр женился
 (17) Мария *была рада*, что осталась дома => Мария осталась дома

Обработка анафорических и нулевых местоимений:

- (18) Иванов добился повышения *себе* зарплаты => Имело место повышение зарплаты *Иванову*
 (19) Маша жалела, что \emptyset осталась дома => *Маша* осталась дома

Обработка кванторов:

- (20) *Никто* не забыл принести тетрадку => *Все* принесли тетрадку

Обработка модальных контекстов и контекстов «возможных миров»:

- (21) Зенит *мог добиться* победы над ЦСКА => *Могла* иметь место победа Зенита над ЦСКА
 (22) Пётр *считает*, что Иван *успел* сесть в поезд => Пётр *считает*, что Иван *сел* в поезд

Отдельным и, возможно, наиболее важным пунктом является формирование вывода импликаций не на естественном языке, а на формализованном языке семантических представлений. В лингвистическом процессоре ЭТАП-3 уже имеется такой язык, опирающийся на специально разработанную онтологию OntoEtap [Богуславский и др., 2015]. Возможно, многие из вышеописанных вопросов было бы удобнее решать в рамках семантических представлений.

Также в семантическом представлении нет нужды формировать вывод импликации отдельно. Достаточно лишь пометить нужное событие как фактически имевшее место, то есть высказывание о нём – как истинное. При условии, что исходное событие, выраженное имплицативным глаголом, также помечено как фактически имевшее место. Однако сейчас такие пометки отсутствуют. Семантическое представление ничего не говорит об истинности высказываний, в него входящих. Поэтому начинать надо с внесения в семантические представления информации об истинности высказываний.

Заключение

В работе исследованы более ста имплицативных глаголов русского языка, проанализированы их семантические и синтаксические свойства, поведение в различных контекстах, и настроены правила в лингвистическом процессоре ЭТАП-3, формирующие вывод импликации. Это позволило в явном виде представить информацию, которая неявно присутствовала в тексте, и таким образом улучшить семантический анализ текстов, в частности для целей вопросно-ответной системы. Намечены направления будущей работы.

Литература

- Богуславский И. М., Диконов В. Г., Иомдин Л. Л., Лазурский А. В., Сизов В. Г., Тимошенко С. П. 2015. Семантический анализ и ответы на вопросы: система в стадии разработки. *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 27 — 30 мая 2015 г.)*. Вып. 14 (21), т. 1: 62–80. — М.: Изд-во РГГУ, 2015.
- Цинман Л. Л. 1986. Язык для записи лингвистической информации в системе автоматического перевода ЭТАП (опыт «практической логики»). *Семиотика и информатика*. Вып. 27: 82–120.
- Geis, Michael L. and Arnold M. Zwicky. 1971. On Invited Inferences. *Linguistics Inquiry*, 2: 561–566.
- Grice, H. Paul. 1975. Logic and Conversation. *Syntax and Semantics*, Vol. 3, Speech Acts, ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press 1975: 41–58.
- Karttunen, Lauri. 1971. Implicative verbs. *Language*, 47:340–358.
- Karttunen, Lauri. 1973. La logique des constructions anglaises à complément prédicatif. *Langages*, 8:56–80. Published originally as “The Logic of English Predicate Complement Constructions” by the Indiana University Linguistics Club in 1971.