

S T U D I A P H I L O L O G I C A



ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ им. А. А. ХАРКЕВИЧА РАН
ИНСТИТУТ РУССКОГО ЯЗЫКА им. В. В. ВИНОГРАДОВА РАН

Ю. Д. АПРЕСЯН, И. М. БОГУСЛАВСКИЙ,
Л. Л. ИОМДИН, В. З. САННИКОВ

ТЕОРЕТИЧЕСКИЕ ПРОБЛЕМЫ РУССКОГО СИНТАКСИСА

ВЗАИМОДЕЙСТВИЕ ГРАММАТИКИ И СЛОВАРЯ

Ответственный редактор
Ю. Д. Апресян



ЯЗЫКИ СЛАВЯНСКИХ КУЛЬТУР
МОСКВА 2010

УДК 811.161.1

ББК 81.2Р-4

А 77

Издание осуществлено при финансовой поддержке

Российского гуманитарного научного фонда

(РГНФ)

проект № 08-04-16172

Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л., Санников В. З.

А 77 Теоретические проблемы русского синтаксиса: Взаимодействие грамматики и словаря / Отв. ред. Ю. Д. Апресян. — М.: Языки славянских культур, 2010. — 408 с. — (Studia philologica).

ISSN 1726-135X

ISBN 978-5-9551-0386-0

В монографии рассматривается ряд трудных вопросов русского синтаксиса и намечаются способы их решения в формальной модели языка. В первой главе дан краткий обзор самой модели русского синтаксиса. В него вошли комментированный список используемых в ней синтаксических отношений и правила синтаксических чередований. Во второй главе рассматриваются различные типы синтаксических конструкций, наиболее тесно взаимодействующих с лексикой. Третья глава посвящена новой версии теории синтаксического управления. В последней, четвертой главе разбираются уроки функционирования формальной модели синтаксиса в системе машинного перевода с русского языка на английский.

ББК 81.2Р-4

*В оформлении переплета использован фрагмент картины
С. де Рокамболя «Вариации на тему Белый Квадрат» (2000)*

Электронная версия данного издания является собственностью издательства, и ее распространение без согласия издательства запрещается.

ISBN 978-5-9551-0386-0

© Авторы, 2010

© Языки славянских культур, 2010

СОДЕРЖАНИЕ

Введение.....	11
---------------	----

Глава 1

О модели русского синтаксиса

1.1. Вводные замечания.....	21
1.2. Синтаксические отношения	21
1.2.1. Актантные СинтО.....	24
1.2.1.1. Предикативное СинтО.....	25
1.2.1.2. Дательно-субъектное СинтО.....	25
1.2.1.3. Агентивное СинтО.....	26
1.2.1.4. Квазиагентивное СинтО.....	26
1.2.1.5. Несобственно-агентивное СинтО.....	26
1.2.1.6. Комплетивные СинтО.....	27
1.2.1.7. Присвязочное СинтО.....	27
1.2.1.8. Несобственно-комплетивные СинтО.....	28
1.2.1.9. Неактантно-комплетивное СинтО.....	28
1.2.1.10. Комплетивно-аппозитивное СинтО.....	29
1.2.1.11. Предложное СинтО.....	29
1.2.1.12. Подчинительно-союзное СинтО.....	29
1.2.1.13. Сравнительное СинтО.....	30
1.2.1.14. Сравнительно-союзное СинтО.....	30
1.2.1.15. Элективное СинтО.....	30
1.2.1.16. Сентенциально-предикативное СинтО.....	31
1.2.1.17. Презентативное СинтО.....	31
1.2.2. Атрибутивные СинтО.....	32
1.2.2.1. Определительные СинтО.....	32
1.2.2.2. Общеатрибутивные СинтО.....	33
1.2.2.3. Аппозитивные СинтО.....	34
1.2.2.4. Количественные СинтО.....	34
1.2.2.5. Обстоятельственные СинтО.....	36
1.2.3. Сочинительные СинтО.....	39
1.2.3.1. Сочинительное СинтО.....	40
1.2.3.2. Сентенциально-сочинительное СинтО.....	40

1.2.3.3. Сочинительно-союзное СинтО	40
1.2.3.4. Коммуникативно-сочинительное СинтО	40
1.2.3.5. Кратное СинтО	41
1.2.4. Служебные СинтО	41
1.2.4.1. Аналитическое СинтО	41
1.2.4.2. Пассивно-аналитическое СинтО	41
1.2.4.3. Вспомогательное СинтО	42
1.2.4.4. Количественно-вспомогательное СинтО	42
1.2.4.5. Соотносительное СинтО	42
1.2.4.6. Эксплетивное СинтО	43
1.2.4.7. Пролептическое СинтО	43
1.2.4.8. Эллиптическое СинтО	43
1.3. Чередования в синтаксисе (к проблеме альтернологии)	43
1.3.1. Вводные замечания	43
1.3.2. «Основные» и «дополнительные» синтаксические средства	44
1.3.3. Чередования в морфологии и синтаксисе	45
1.3.4. Примеры синтаксических чередований. Чередования «комбинаторные» и «исторические»	46
1.3.5. Чередования в морфологии и синтаксисе: различия между ними	49
1.3.6. Какие из морфологических характеристик могут подвергаться чередованиям?	51
1.3.7. Альтернология и понятия «управление» и «согласование» ...	53
1.3.8. Об отображении синтаксических чередований в формальных моделях русского языка	54
1.3.9. О соотношении понятий «чередование», «морфонология», «синтморфология»	56
1.3.10. О списке русских синтаксических чередований	57
1.3.11. Краткие итоги	58

Глава 2

Конструкции малого синтаксиса

2.1. Конструкция типа негде спать: синтаксис, семантика, лексикография	59
2.1.1. Вводные замечания	59
2.1.2. Лингвистическая природа местоимений типа <i>некого</i>	62
2.1.2.1. Орфографические свойства	63
2.1.2.2. Просодические свойства	63
2.1.2.3. Морфологические и синтаксические свойства	65

2.1.2.4. Лексические свойства	73
2.1.2.5. <i>Не</i> ² -единицы — синтаксические агломераты	73
2.1.3. Синтаксическая структура предложений с <i>не</i> ² -единицами ...	74
2.1.3.1. Количество синтаксических позиций в <i>не</i> ² -предложениях	74
2.1.3.2. Древесная синтаксическая структура <i>не</i> ² -предложений ...	77
2.1.4. Синтаксические позиции конструкции	84
2.1.4.1. Первая позиция	84
2.1.4.2. Вторая и четвертая позиции	91
2.1.4.3. Третья позиция	95
2.1.4.4. Пятая позиция	106
2.1.5. Заключение	110
2.2. Пролептические и близкие к ним конструкции	113
2.2.1. Вводные замечания	113
2.2.2. Обособленно-атрибутивные конструкции	115
2.2.3. Пролептические конструкции	119
2.2.4. Спорные случаи	124
2.2.5. Формальное представление пролептических конструкций ..	128
2.3. Гипотеза о двух синтаксических началах	129
2.3.1. Проблема	129
2.3.2. Решение	132
2.3.3. Обсуждение	136
2.4. Синтаксические фраземы: между лексикой и синтаксисом	141
2.4.1. Синтаксис и микросинтаксис	141
2.4.2. Микросинтаксические конструкции, образованные вопросительными местоимениями	147
2.4.2.1. Вводные замечания	147
2.4.2.2. К-слова в русском языке	147
2.4.2.3. Классы неопределенных местоимений, формируемые К-словами	148
2.4.2.4. Синтаксические фраземы, начинающиеся с К-слова	149
2.4.2.5. Синтаксические фраземы, заканчивающиеся К-словом	153
2.4.3. Синтаксическая фразема ВСЁ РАВНО	156
2.4.4. Синтаксические фраземы, мотивированные полнозначными словами	162
2.4.4.1. Вводные замечания	162
2.4.4.2. Составной предлог В СИЛУ	163
2.4.4.3. Наречие степени ОТ СИЛЫ	167
2.4.4.4. Наречная конструкция В Х-ОВУЮ СИЛУ	169

2.4.4.5. Предикативное наречие В СИЛАХ 1 и предикативная наречная фразеосхема В СИЛАХ 2	171
2.4.5. Синтаксические фраземы с экспрессивным значением	174
2.4.5.1. Какого чёрта?	174
2.4.5.2. На кой чёрт?	179
2.4.5.3. Чёрт с ним!	181
2.4.5.4. Ну его к чёрту!	184
2.4.5.5. Чёрта с два	189
2.5. Конструкции с тождественными словоформами	191
2.5.1. Вводные замечания	191
2.5.2. Редупликация — морфология, словообразование, синтаксис?	192
2.5.3. Основные синтаксические характеристики КТС	193
2.5.4. Основные семантические группы КТС	194
2.5.4.1. Весело, весело было	195
2.5.4.2. Весело-весело было	199
2.5.4.3. Он рос и рос	200
2.5.4.4. Снег да снег кругом	202
2.5.4.5. Я верил и не верил	202
2.5.4.6. Бывают аварии и аварии	203
2.5.4.7. Ну, медведь и медведь	204
2.5.4.8. (Ну), упал и упал	204
2.5.5. Повторы вопросительных местоимений	205
2.5.5.1. Куда, куда?	205
2.5.5.2. Кто да кто?	206
2.5.5.3. (Ну), кто-кто, а Коля придет	206
2.5.5.4. Кого-кого он (только) не знает	208
2.6. Иллокутивное употребление как вид синтаксического эллипсиса	209
2.6.1. Вводные замечания	209
2.6.2. Объем и границы иллокутивных употреблений	212
2.6.3. Кандидаты в иллокутивные употребления	214
2.6.4. Иллокутивные употребления или синтаксический эллипсис?	222
2.6.5. Иллокутивное употребление. Попытка уточнить понятие	224
2.7. Сочинительные союзы и синтаксические конфликты	226
2.7.1 Вводные замечания	226
2.7.2. Каноническое и неканоническое сочинение	228
2.7.2.1. Принцип однофункциональности сочиненных членов	228
2.7.2.2. Вторичная союзная связь	230

2.7.3. Союзные соединения и неоднородные союзы	237
2.7.4. Неканонические сочинительные конструкции с союзными соединениями	240
2.7.4.1. Смещение	244
2.7.4.2. Асимметричное сокращение	250
2.7.5. Асимметричная конструкция типа «Перенос без опущения»	257
2.7.5.1. Разновидности этой конструкции	257
2.7.5.2. Конструкция «Перенос без опущения» как способ разрешения синтаксического конфликта	262
2.7.6. Источники конфликта	271
2.7.6.1. Семантический источник конфликта	273
2.7.6.2. Синтаксический источник конфликта	273
2.7.6.3. Сочетаемый источник конфликта	275
2.7.6.4. Коммуникативный источник конфликта	276
2.7.7. Другие способы разрешения конфликта	277
2.7.8. Заключение	279

Глава 3

Трехуровневая теория управления: лексикографический аспект

3.1. Вводные замечания	281
3.2. Некоторые этапы в эволюции лингвистической теории управления	281
3.2.1. Традиционный взгляд на управление	282
3.2.2. Системный лексико-синтаксический подход к управлению	283
3.2.3. Первый семантический этап в развитии теории управления	285
3.2.4. Второй семантический этап в развитии теории управления	289
3.3. Принципы установления семантических актантов глагольной лексемы	290
3.3.1. ФКП и ее рефлекс в грамматической и лексической системах языка	290
3.3.2. ФКП и глагольное управление	303
3.3.2.1. Общая характеристика	303
3.3.2.2. Многоактантные глаголы	308
3.3.2.3. Многоактантные существительные	329
3.3.2.4. Замечание о семантических ролях	335
3.3.3. Аналитические толкования лексем	337
3.3.4. Совмещение ФКП и аналитических толкований (типов и портретов)	340

3.3.4.1. Глаголы со значением агрессивных физических действий	340
3.3.4.2. Глаголы со значением побуждения	342
3.3.4.3. Путативные глаголы	342
3.3.5. ФКП и некоторые трудные вопросы управления	343
3.4. Типы соответствия семантических и синтаксических актантов	347
3.4.1. Прямое соответствие актантов	347
3.4.2. Словарная диатеза и мена диатезы	347
3.4.3. Расщепление актантов	351
3.4.4. Смещение актантов	353
3.4.5. Выражение двух семантических ролей одним актантом	356
3.4.6. Абсолютивная конструкция	361
3.4.7. Синтаксическая невыразимость семантического актанта	362
3.4.8. Появление нового семантического актанта	364
3.4.9. Актанты и сирконстанты	365
3.4.9.1. Временные значения	366
3.4.9.2. Целевые значения	367
3.4.9.3. Причинные значения	368
3.4.9.4. Значение способа	370
3.5. Новая номенклатура семантических ролей	370

Глава 4

Уроки русского синтаксиса

(из опыта работы системы машинного перевода)

4.1. Вводные замечания	380
4.2. Урок 1 — Синтаксический	381
4.3. Урок 2 — Лексико-синтаксический	384
4.4. Урок 3. Полузнание хуже незнания	388
4.5. Урок 4. Границы возможного	389
Литература	393

ВВЕДЕНИЕ

Эта монография¹ — результат исследований авторов, занятых, помимо теоретической лингвистики, разработкой крупных компьютерно-лингвистических приложений, в первую очередь полифункционального лингвистического процессора ЭТАП-3 (двунаправленная система англо-русского и русско-английского автоматического перевода, конверторы на язык UNL и деконверторы с него, компьютерная система перифразирования и ряд других), а также первого в мире синтаксически размеченного корпуса русских текстов (к настоящему времени объем корпуса — более 40 000 предложений; он доступен по адресу <http://www.ruscorp.org.ru/search-syntax.html>); см. о них, в частности, [Апресян и др. 1989, 1992, 2005], [Aprésjan et al. 2003], [Boguslavsky et al. 2000]. Работа над приложениями постоянно давала мощные импульсы к постановке и решению теоретических задач, отчасти представленных в книге, отчасти выходящих за ее пределы. Эти импульсы были двоякого рода.

Главный импульс шел от самой функционирующей системы ЭТАП-3. Она служила, в сущности, гигантским экспериментальным полигоном, на котором ее создатели могли проверять как адекватность положенной в ее основу общей лингвистической теории «Смысл ⇔ Текст» И. А. Мельчука, так и правильность формальных описаний конкретных явлений из области морфологии, синтаксиса и словаря обоих рабочих языков. У нас уже был случай писать, как такие глобальные эксперименты позволяют совершенствовать и модель в целом, и решения частных вопросов. Один свежий пример.

В порядке экспериментальной проверки того, как работают некоторые синтагмы (синтаксические правила) в условиях многократной

¹ Работа, результаты которой отражены в этой книге, была поддержана грантами Российского гуманитарного научного фонда № 04-04-00263а, № 06-04-00289а, грантами Российского фонда фундаментальных исследований № 05-06-80361, № 08-06-00344 и грантом для государственной поддержки ведущих научных школ Российской Федерации № НШ-3205.2008.6. Всем этим Фондам авторы выражают признательность за финансовую поддержку, без которой книга не была бы написана.

омонимии актантных зависимых в форме винительного и родительного падежей, компьютеру была предложена для анализа простенькая фраза *Его кабинет меня не устраивал*. В качестве первого варианта компьютер выдал неправильную синтаксическую структуру, в которой словоформа *его* была сочтена прямым дополнением глагола *устраивал*, а словоформа *меня* была связана со словом *кабинет*; остальные связи были установлены правильно. Полученной неправильной структуре соответствует вполне правильный смысл $\approx$ 'Принадлежавший мне кабинет, в котором он собирался разместиться, ему чем-то не понравился'. Однако на литературном русском языке этот смысл следовало бы выразить предложением *Мой кабинет его не устраивал*.

Если бы экспериментатор попросил выдать ему другие варианты синтаксического анализа фразы *Его кабинет меня не устраивал*, среди них непременно оказался бы и правильный вариант, со словом *его* в качестве зависимого существительного *кабинет*, словом *меня* в качестве прямого дополнения глагола *устраивал* и смыслом $\approx$ 'Принадлежавший ему кабинет, в котором я собирался разместиться, мне чем-то не понравился'. Дело, однако, в том, что неправильная структура, выданная компьютером, получилась «незаконно», исключительно в результате неполноты нужного для ее анализа синтаксического правила. Эксперимент обнажил эту некорректность и подтолкнул к ревизии правила.

В исходной формулировке правило гласило, что первый актант существительных типа *кабинет* может выражаться формой родительного падежа. Действительно, для фраз типа *Его кабинет отца не устраивал* синтаксическая структура, полученная компьютером для фразы *Его кабинет меня не устраивал*, была бы вполне правильной. Такие фразы на письме реально двусмыслны: а) либо $\approx$ 'Принадлежавший отцу кабинет, в котором он собирался разместиться, ему чем-то не понравился', б) либо $\approx$ 'Принадлежавший ему кабинет, в котором отец собирался разместиться, отцу чем-то не понравился'. Каждому из этих осмыслений соответствует своя синтаксическая структура². Однако во фразе *Его кабинет меня не устраивал* никакой омонимии нет и, значит, нет альтернативы структуре, в ко-

² В устной речи двум разным синтаксическим структурам и, следовательно, двум разным осмыслениям соответствуют две разные просодии (// обозначает паузу): *Его // кабинет отца не устраивал* VS. *Его кабинет // отца не устраивал*.

торой *меня* является прямым дополнением глагола. Почему же *меня* (родительный падеж *я*, ср. *Без меня, пожалуйста, С меня сто рублей*) не может выполнять такую же синтаксическую функцию при слове *кабинет*, как формы родительного падежа других существительных, включая личные местоимения третьего лица (ср. *Кабинет его (ее, их) был великолепен*, но не **Кабинет меня был великолепен*, надо: *Мой кабинет был великолепен*)?

Дело в том, что в русском языке действует довольно жесткий запрет на использование форм родительного падежа личных местоимений первого и второго лица (*я, ты, мы и вы*) в функции отсубстантивных (квази)агентивных и атрибутивных зависимых, не распространяющийся на личные местоимения третьего лица. Он описан в ряде специальных работ (см., например, [Виноградов 1947: 335], [Падучева 1984: 54]), но не в академических грамматиках ([Грамматика 1960] и [Грамматика 1980]), на которые мы в первую очередь ориентировались при разработке формальной модели русского синтаксиса. В первоначальной версии правил распознавания (квази)агентивных и атрибутивных зависимых он не был должным образом учтен, что и привело к ошибке.

Однако исследование ошибки не ограничилось просто коррекцией правила. Уточнение любого правила всегда сопровождается продумыванием большого языкового материала, так или иначе с ним связанного.

В ходе этой работы было замечено, что неправильность словосочетаний указанного типа несколько уменьшается, если их вершиной являются прототипические предикаты, без намека на значение предметности. Словосочетания типа [?]*участие меня (тебя) (в этой работе обязательно)* несколько более приемлемы (или, может быть, несколько менее неприемлемы), чем аналогичные словосочетания с предметными существительными типа **кабинет (портфель) меня (тебя, нас, вас) (был просторный и светлый)*.

Кроме того, внутри самого класса предикатов предикаты со значением процесса оказались более приемлемыми в роли синтаксических хозяев квазиагентивных и атрибутивных зависимых вида «личное местоимение первого или второго лица в форме родительного падежа», чем другие семантические типы предикатов. Так, *появление меня на пороге* лучше, чем [?]*приход меня (тебя, нас, вас) (на собрание)*; ср. также допустимое словосочетание *становление (формирование) меня как ученого* и менее приемлемые [?]*Работа (карьера) меня как ученого*.

Еще один фактор, действующий в пользу легитимизации словосочетаний с личными местоимениями первого и второго лица в качестве квазиагентивных и атрибутивных, — это синтаксическое «утяжеление» местоимения. Это обстоятельство было отмечено в упомянутой выше работе Е. В. Падучевой; ср. ее пример *Он ставит под сомнение компетентность вас как биографа* при неправильности **Он ставит под сомнение компетентность вас* (надо: *Он ставит под сомнение вашу компетентность*). В ходе продумывания более широкого материала выяснилось, что и этот фактор ярче проявляется на материале прототипических предикатов и «глохнет» на материале предметных имен; ср. *Приход меня и моего друга ничего не решит* VS. **Компьютер меня и моего друга ничего не решит*.

Эти наблюдения отчасти совпадают с выводами, которые были сделаны независимо от обсуждаемого эксперимента и на совсем другом материале. Л. Л. Иомдин, занимаясь семантикой и синтаксисом выражения *(не) в чьих-л. силах*, заметил, что его первый (сентенциальный) актант, выполняющий синтаксическую функцию подлежащего, нормально выражается инфинитивом, но не отглагольным существительным: можно сказать *Помочь вам не в моих силах*, но не **Помощь вам не в моих силах* (подробнее см. раздел 2.4.4). Заметим, что и в этих случаях неправильность словосочетаний с существительным в роли подлежащего тоже падает по мере того, как нарастает идея процессности в значении существительного: *Оказание вам помощи не в моих силах* явным образом лучше, чем **Помощь вам не в моих силах*.

Иными словами, на материале разной природы обнаруживается действие похожих семантических сил. Оказывается, что простейшие синтаксические конструкции отражают глубокие различия между семантическими классами и подклассами фундаментальной классификации предикатов. Это выводит исследование на совершенно новый уровень и позволяет получить принципиально новое знание о языке.

Перейдем ко второму фактору, который стимулировал получение новых знаний о синтаксисе и семантике изучаемых языков в процессе компьютерных экспериментов. В ходе построения упомянутых систем автоматической обработки текстов авторам пришлось вплотную столкнуться с громадным материалом современных русских текстов, изобилующих сложнейшими синтаксическими (как, впрочем, и другими языковыми) явлениями. Многие из этих явлений до сих пор не получили достаточно надежной и устойчивой интерпре-

тации не только в грамматиках, но и в научных статьях и монографиях. Здесь тоже будет уместен один пример.

Работа по глубокой автоматической разметке текстов — морфологической и синтаксической (см. упомянутую выше статью [Апресян и др. 2005]) — была организована следующим образом: сначала компьютер анализировал текст на основе имеющихся у него словарей и правил, а затем результаты его анализа поступали на постредактирование профессиональному лингвисту, как правило, из числа разработчиков ЭТАПа. В результатах, полученных компьютером, наряду с правильными анализами встречались, естественно, и ошибки. Они были двух типов.

Во-первых, компьютер неправильно распознавал какое-то морфологическое, синтаксическое или лексическое явление; так, для предложения *Он увидел их семью своими глазами* была получена структура, которая интерпретировалась как ‘Он увидел их, используя (все) семь своих глаз’. Здесь словоформа *семью* была неправильно проанализирована (относительно нужного смысла) и морфологически (ТВОР вместо ВИН), и лексически (*семь* вместо *семья*), и синтаксически (количественное зависимое слова *глаза* вместо прямого дополнения глагола *увидеть*). Такие ошибки легко исправлялись при постредактировании.

Во-вторых, для многих фраз, на первый взгляд совершенно простых, компьютер не мог построить никакой структуры по той причине, что, как оказывалось при последующем изучении результатов, не располагал нужными правилами. В частности, большие сложности возникли при анализе некоторых идентифицирующих конструкций с местоименными словами *это* и *то*.

Они формируют два основных типа идентифицирующих конструкций:

(а) конструкции со связкой (в том числе нулевой) типа *Это Иван, Это был Иван*;

(б) конструкции с полнозначным глаголом типа *Это идет Иван, Это <то> стучит дождь* и т. п.; ср. у И. Бродского *С чего бы это вдруг? Серебряный висок? / Оскомина во рту от сладостей восточных? / Потусторонний звук? / Нет, то стучит песок, / пустыни талисман, в моих часах песочных*.

С конструкциями первого типа трудностей у нас не возникало — для них строились синтаксические структуры со словом *это* в качестве подлежащего и вторым существительным в роли именной части составного именного сказуемого.