

Разрешение синтаксической омонимии при анализе комитативных групп в русском языке

Александр Владимирович Архипов, Мария Михайловна Брыкина
 МГУ им. М. В. Ломоносова
sarkipo@mail.ru, mashukm@mail.ru

В работе предлагается квантитативный подход к определению синтаксического хозяина комитативной группы («с + ИГ_[твор]») в предложении и типа ее зависимости при автоматическом анализе текста. Выделяются несколько параметров, по которым оценивается сама комитативная группа и ее потенциальные хозяева. На основании суммы взвешенных оценок по этим параметрам определяется наиболее вероятный хозяин и тип зависимости. В случае реальной омонимии конкурирующие варианты анализа должны получить равную оценку вероятности. Демонстрируется необходимость учета нетривиальной семантической (лексикографической) информации, такой как принадлежность предиката к классу глаголов физического перемещения и позиции, семантический класс комитативной группы, и т. п.

1. Синтаксические функции комитативных групп в русском языке

В русском языке комитативными группами (КГ) мы называем предложные группы вида «предлог с + ИГ в творительном падеже» (напр. *с вареньем, с женой, с удовольствием*). Ярлык «комитативная группа» апеллирует в данном случае не столько к содержательной стороне, сколько к формальной: предложные группы «с + ИГ_[твор]» рассматриваются как основное средство выражения собственно комитативных конструкций (см. ниже); на этом основании мы применяем к ним название «комитативная группа», причем во всех употреблениях, в том числе не имеющих собственно комитативной семантики.

В дополнение к предложным группам с предлогом с, мы будем рассматривать как КГ предложные группы с составным предлогом *вместе с* (см. раздел 3.3).

Синтаксические функции КГ в русском языке были описаны, в частности, в «Синтаксическом словаре» Г. А. Золотовой [2001 (2-е изд.)] и, с несколько других позиций, в работе [Архипов 2003]; некоторые замечания к описанию КГ у Г. А. Золотовой см. также в [Архипов, в печати]. Ниже мы перечислим вкратце основные синтаксические типы употреблений комитативных групп. (Не считая собственно комитативной конструкции, выделяемой в особый тип, разделений среди сирконстантов мы в данной работе не проводим. Некоторые частные случаи для упрощения изложения опущены.)

А. КГ синтаксически подчиняется предикату

Тип А1-КОМ: собственно комитативная конструкция.

1. Ее всегда узнавали, когда она появлялась с ним на улице. (Ю. Домбровский)

1. Этим летом я жил с пастухами на альпийских лугах Башкапсара <...> (Ф. Искандер)
Тип А2-СИММ: КГ является актантном симметричного предиката.
2. Как будто и со второй женой он не ужился — Рита в точности не знала. (И. Грекова)
3. Рядом с воротами — неказистое зданье в роде кирпичного сарая. (И. Грекова)
Тип А3-АКТ: КГ является актантном несимметричного предиката.
4. — Ни беса лысого они со мной не сделают. (Ю. Домбровский)
5. С этим делом придется поработать, решил он <...> (В. Аксенов)
Тип А4-СИРК: КГ является сирконстантом.
6. По вагону шёл с кожаной сумкой и продавал газеты, журналы. (А. Солженицын)
7. С заискивающей надеждой заглядывала им в глаза. (А. Алексин)

Б. КГ не подчиняется предикату

Тип Б1-ОПР: КГ является определением имени.

8. — Что там? — вновь обратилась я к женщине с воспаленным лицом. (А. Алексин)
9. Ему ответил молодой, звонкий, с легкой наглечой голосок <...> (Ю. Домбровский)
Тип Б2-СОЧ: КГ выступает в качестве конъюнкта в квазисочинительной конструкции.
10. И Клара с Иннокентием тоже пошли за ними. (А. Солженицын)
11. — Уж мы с мамочкой постарались. (В. Аксенов)

2. Проблема распознавания синтаксической функции КГ при автоматическом анализе

Присутствие в тексте комитативных групп может представлять для систем машинного перевода и автоматического понимания текста проблемы двух видов. Во-первых, это установление синтаксического хозяина КГ; во-вторых, установление типа зависимости КГ от этого хозяина. Вопросы различения семантических типов КГ при известной синтаксической функции (напр., различения типов сирконстантов) мы в настоящей статье затрагивать не будем.

Важность обоих видов проблем легко продемонстрировать. Так, в результате анализа собственно комитативной конструкции типа *Петя пошел в кино с Машей* и *Петя и Маша* должны быть проинтерпретированы как участники ситуации ‘пойти в кино’, причем с одинаковыми ролями; напротив, в случае *Петя пошел в кино с чемоданом* объект ‘чемодан’ не должен интерпретироваться как самостоятельно перемещающийся агенс.

При переводе комитативной группы на французский язык, например, подчиненные предикату КГ (типы А1–А4) следует переводить через предлог *avec*, а КГ в роли определения (тип Б1-ОПР) — через предлог *à* (*au, aux*): *Il a dansé la valse [avec une brune [aux boucles d'oreilles d'argent]]* ‘Он танцевал вальс [с брюнеткой [с серебряными серьгами]]’; квазисочиненные же конструкции (тип Б2-СОЧ) следует переводить сочиненными: *Отец с сыном уехали из города => Le père et le fils ont quitté la ville.*

В существующих системах автоматического анализа текста значительная часть трудностей может быть решена на этапе синтаксического анализа, в частности, путем расширения и уточнения распознаваемых моделей управления. Это позволит избежать многих ошибок при понимании и переводе, таких как неидиоматичные переводы (*С Новым годом!* => *With a New Year!* вместо *Happy New Year!*), нераспознавание валентностей (*Он изменял жене с секретаршей* => *He was betraying a wife and a secretary*) и т. п. (примеры из одной из известных систем машинного перевода).

В то же время, в некоторых контекстах возникающие неоднозначности трудно разрешить без обращения к разнообразным фоновым знаниям, энциклопедической информации, минимально — к фреймам, соотносимым с участниками ситуации. Например, для интерпретации определений при вложенных ИГ, ср. *парень в голубой рубашке с зелеными глазами vs парень в голубой рубашке с зелеными рукавами*, необходимо учитывать, что глаза обычно бывают у людей (и животных), а рукава — у одежды, предметом какой являются рубашки.

Носители языка, как правило, вообще не отдают себе отчет в существовании подобной омонимии, замечая ее только при внимательном «вчитывании» в текст. Между тем, смоделировать механизмы ее разрешения на основании одной только морфосинтаксической информации весьма непросто. Цель дальнейшего рассуждения — показать факторы, влияющие на интерпретацию комитативных групп человеком, пути их формализации и значимость их учета для адекватного анализа (иными словами, возможные потери при отказе от учета того или иного фактора).

3. Схема анализа КГ

В ходе разработки данного вопроса мы отказались от несколько прямолинейного подхода, основанного на «да-нет»-вопросах и простом ветвлении [Архипов 2003]. Предлагаемая в настоящее время более гибкая схема анализа включает следующие этапы: 1) поиск потенциальных синтаксических хозяев КГ; 2) для каждого из потенциальных хозяев анализируются несколько параметров, влияющих на возможность/вероятность того или иного типа отношений с КГ; 3) выбор наиболее вероятной(-ых) возможности(-ей).

На третьем этапе в принципе возможны две стратегии принятия решений. Первая стратегия — статистическая: на основании размеченного вручную корпуса примеров выбирается вариант анализа, наиболее частотный при данном сочетании значений параметров. При второй стратегии параметрам присваиваются различные веса, и вероятность каждого варианта анализа получает количественную оценку; в итоге выбирается вариант, который получает наибольшую оценку.

Наиболее эффективным представляется сочетание обеих стратегий: данные о важности тех или иных параметров получаются путем статистического исследования корпуса примеров и затем воплощаются в количественной оценке значений параметров.

Схема анализа была реализована нами в виде интерактивного макроса для MS Excel, роль которого сводится к облегчению механической работы пользователя-лингвиста и автоматизации третьего этапа анализа (количественной оценки). Тестирование схемы проводилось на материале около 400 примеров употребления предложных групп «с + ИГ_[твор]» из Русского Национального Корпуса (<http://corpora.yandex.ru>).

3.1 Необходимая предварительная информация

Предполагается, что необходимость в Схеме, разрешающей омонимию КГ, появляется после того, как произведен полный морфологический и частичный синтаксический анализ текста. Для работы Схемы требуется выполнение следующих исходных условий:

12. Текст должен быть разбит на предложения, выделены простые предложения в составе сложных. Поиск потенциальных синтаксических хозяев КГ проводится внутри простого предложения (клаузы).
13. В клаузе должны быть определены вершинный предикат и подлежащее (если есть).
14. Должны быть распознаны именные и предложные группы (в частности, комитативные).
15. Должны быть распознаны однородные актанты и определения.

3.2 Поиск потенциальных хозяев комитативной группы

Поиск производится внутри клаузы (простого предложения) К, к которой относится исследуемая комитативная группа. Подчиненные клаузы, распространяющие клаузу К, исключаются из рассмотрения, если известно, что КГ не входит в их состав.

В рамках клаузы К потенциальными синтаксическими хозяевами КГ считаются: (1) все финитные и нефинитные глагольные словоформы (аналитические словоформы считаются за одну вершину); (2) выраженные не глаголами предикаты (предлоги/наречия, прилагательные, предложные группы), которые могут иметь множественную валентность (напр. *вместе, рядом, одновременно, заодно, один, в обнимку* и т. п.) или специфическую валентность на КГ (*хорошо/плохо/трудно с чем*); (3) все номинализации (*встреча, движение, намерение, удар, работа*); (4) все вершины именных групп, не отделенные от КГ потенциальными хозяевами других типов (допускается отделение определениями).

3.3 Оценка комитативной группы

Комитативная группа оценивается по следующим параметрам:

- Одушевленность. Неодушевленная КГ с большей вероятностью имеет функцию сирконстанта, чем собственно комитативную, и наоборот. Одушевленные КГ редко служат определениями одушевленных ИГ (ср. однако *дама с собачкой*).
- Личность. Личные местоимения не образуют КГ-определения.
- Собственное или нарицательное. КГ с именами собственными редко используются в качестве определений и сирконстантов (ср. однако *фильм с Брюсом Уиллисом*).
- Однородность характеристик с базовой ИГ. Наиболее вероятно квазисочинение двух личных, или двух одушевленных, или двух неодушевленных имен.
- Семантический класс. Существует несколько семантических кластеров имен, предрасположенных к использованию в функции сирконстанта; многие из них являются номинализациями. Среди них можно назвать обозначения чувств и эмоций (*с удовольствием, с грустью, с ненавистью*), жестов и характеристик поведения (*с улыбкой, с непринужденностью*), говорения и звуков (*со словами, со смехом, с криком, с шумом, с лязгом*); в целевых сирконстантах нередки КГ *с целью, с намерением*, и т. п.
- Наличие в КГ составного предлога *вместе с*. Фактически КГ с предлогом *вместе с* анализируется в два прохода. При первом проходе устанавливается зависимость КГ с предлогом *с* от симметричного предиката *вместе* (тип А2-СИММ). При втором проходе происходит поиск хозяина для новой, вторичной КГ с предлогом *вместе с*. Количество возможных синтаксических функций для вторичных КГ ограничено собственно комитативной конструкцией (тип А1-КОМ) и, реже, квазисочиненной конструкцией (тип В2-СОЧ) и функцией сирконстанта (тип А4-СИРК).

3.4 Оценка потенциальных хозяев комитативной группы

1. Оценка потенциальных хозяев-глаголов. Допустимыми типами зависимости КГ от глагольной вершины являются типы А1–А4. Предлагается учитывать следующие признаки вершины:

- Имеется ли в модели управления валентность на КГ (отрицательный ответ исключает типы А2-СИММ и А3-АКТ);
- Если в МУ предиката имеется множественная валентность (т. е. предикат симметричен), проверяется число базовой ИГ, соответствующей этой валентности. Единственное число базовой ИГ означает отдельный способ заполнения множественной валентности с помощью КГ; множественное число базовой ИГ, однако, не означает обратного.

- Позицию по отношению к КГ (для типа А2-СИММ). Для симметричных предикатов статистически предпочтительным является контактное расположение (КГ непосредственно следует за глаголом или же за базовой ИГ, которая, в свою очередь, следует за глаголом).
- Семантический класс глагола (для типа А1-КОМ). Наиболее вероятно появление собственно комитативной конструкции при глаголах движения и позиции, при глаголе *жить* и его производных, при обозначениях занятий (*гулять, играть, отдыхать*) и деятельности (*работать, учиться, воевать*).

Последний признак наиболее трудно поддается формализации. Представляется, однако, что по крайней мере часть лексикографической информации, требуемой для его проверки (например, принадлежность к классу глаголов движения), должна быть так или иначе доступной в любой системе автоматического понимания или перевода, для которой целесообразно вводить специальные методы различения омонимии КГ. В крайнем случае в качестве промежуточного решения можно задать требуемые глаголы списком.

2. Оценка потенциальных хозяев-имен. Допустимыми типами зависимости КГ от именной вершины являются типы Б1–Б2. Значимыми признаками вершины являются:

- Позиция по отношению к КГ. Для квазисочинительной конструкции приемлемо только положение базовой ИГ непосредственно перед комитативной группой; в функции определения КГ также может предшествовать определяемому существительному; фактором, наиболее однозначно характеризующим тип Б1-ОПР, является обрамление КГ с одной стороны базовой ИГ, а с другой — согласованным определением: Ему ответил молодой, звонкий, с легкой наглечей голосок.
- Выраженность местоимением или полной ИГ. Личные и вопросительные (кто, что) местоимения не могут иметь при себе КГ-определения. (В отличие от них, неопределенные кто-то, что-то принимают определения относительно свободно: кто-то с автоматом; что-то с чем-то).
- Собственное или нарицательное. Имена собственные принимают определения значительно реже нарицательных.
- Согласование сказуемого (при базовой ИГ-подлежащем). Если базовая ИГ имеет единственное число, а сказуемое — множественное, очевидно, мы имеем дело с квазисочинительной конструкцией. Если же сказуемое стоит в единственном числе, квазисочинительная интерпретация исключается.

3. Оценка потенциальных хозяев-номинализаций. Номинализации, совмещающие именные и глагольные свойства, могут относиться к любому из типов А1–А4 и Б1–Б2.

Соответственно, оценка происходит по совокупности признаков, описанных в пп. 2 и 3.

4. Оценка прочих предикатов. Предикаты, выраженные прочими частями речи, претендуют на типы зависимости А2-СИММ и А3-АКТ. Для них оцениваются те же параметры, что указаны в п. 1, кроме семантического класса.

4. Ранжирование значений параметров

Наибольший вес имеют, во-первых, параметры (вернее, их значения), блокирующие тот или иной тип зависимости (например, неконтактное расположение блокирует квазисочинительную интерпретацию). Из «разрешающих» признаков наиболее значимым является наличие у предиката специфической валентности на КГ, чуть менее сильным — наличие множественной валентности (которое может быть решающим только в случае единственного числа соответствующего актанта, напр. **Петя подрался. — Петя подрался с Васей*). Для типа Б1-ОПР сильным фактором является обрамление КГ однородными определениями и самой ИГ-вершиной, как в примере (10). Вероятность типа Б2-СОЧ максимальна при следующем сочетании параметров: ИГ-потенциальный хозяин является подлежащим в единственном числе, а глагол имеет согласование по множественному числу. Следующим по весу параметром является принадлежность КГ или потенциального

хозяина к характерному для того или иного типа отношений семантическому классу. Приблизительно равный или наименьший вес, в зависимости от типа зависимости, имеет наличие составного предлога *вместе с*, одушевленность ИГ (КГ) и выражение ИГ (КГ) именем собственным или нарицательным. Минимальную прибавку к оценке также дает иногда непосредственное следование КГ за потенциальной вершиной.

5. Иллюстрации

Проанализируем один из примеров собственно комитативной конструкции:

16. Предложили эвакуироваться на Урал с заводскими семьями. (И. Грекова) [А1-КОМ]
 В этом предложении имеются три потенциальные хозяина КГ: *предложили*, *эвакуироваться* и *Урал*. Тип Б1-ОПР не получает высокой оценки, поскольку базовая ИГ является именем собственным; тип Б2-СОЧ — из-за неоднородности характеристик ИГ и КГ. (Необходимо иметь в виду, что существительное *семья*, хоть и является морфологически неодушевленным, обозначает совокупность людей; аналогичная информация должна быть доступна также о существительных, обозначающих коллективы и организации, таких как *оркестр*, *милиция*, *университет* и под. Среди имен собственных следует по возможности различать одушевленные и неодушевленные). Из двух предикатов ни один не имеет валентности на КГ, таким образом возможности сводятся к типам А1-КОМ и А4-СИРК. Позиция КГ не создает предпочтения для отнесения к тому или другому предикату. Комитативная группа не предрасположена к использованию в качестве сирконстанта, но, обозначая людей, более склонна к собственно комитативным конструкциям. Наконец, решающим фактором оказывается семантика предикатов: *эвакуироваться* как глагол физического перемещения является наиболее вероятным хозяином КГ.

Рассмотрим следующие предложения:

17. *Петя посмотрел на девочку с угрозой.* [А4-СИРК]
 18. *Петя посмотрел на девочку с бантиками.* [Б1-ОПР]
 19. *Петя посмотрел на девочку с мамой.* [Б2-СОЧ/Б1-ОПР]

При неизменном характере предиката, согласовании, типах ИГ и положении комитативной группы мы встречаем в этих примерах три разные синтаксические функции КГ. В примере (16) имеется реальная омонимия между квазисочиненным прочтением (*посмотрел на девочку и на маму*) и определительным (*посмотрел на ту девочку, которая [была] с мамой*), причем квазисочиненное прочтение предпочтительно.

Предикат *посмотрел* не имеет ни специфической валентности на КГ, ни множественной валентности, следовательно, типы А3-АКТ и А2-СИММ исключаются. Он также не принадлежит к классу предикатов, склонных к участию в комитативной конструкции. КГ *с угрозой* предрасположена к сирконстантному употреблению (*угроза*: неодушевленное, номинализация, характеризует поведение человека). ИГ *девочку* хоть и расположена непосредственно перед КГ, но является одушевленной, тем самым вероятность типа Б1-ОПР средняя. Значение признака одушевленности у этой ИГ и КГ не совпадает, тем самым уменьшается вероятность квазисочиненной интерпретации. Итоговая шкала предпочтительности функций КГ в примере (14) выглядит следующим образом:
 А4-СИРК > Б1-ОПР > Б2-СОЧ > А1-КОМ.

Для примера (15) отличие состоит в том, что КГ *с бантиками* не принадлежит к классу типичных сирконстантов; в силу этого на первый план выходит именная вершина *девочку* как имеющая наиболее выгодное расположение: Б1-ОПР > А4-СИРК > Б2-СОЧ > А1-КОМ.

Одушевленная КГ *с мамой* в примере (16) имеет меньшую вероятность быть определением при одушевленной ИГ и еще меньшую вероятность выступать в роли сирконстанта. Напротив, вероятность квазисочиненной интерпретации в силу однородности характеристик ИГ и КГ повышается: Б2-СОЧ > Б1-ОПР > А1-КОМ > А4-СИРК.

Как видно, учет семантического класса КГ оказывается важен для различения сирконстантной и определительной функции. Вопрос о возможности сведения этого параметра к более общим лексикографическим характеристикам (типа ‘номинализация’, ‘–конкретность’ и т. п.) или, наоборот, к более или менее исчерпывающему списку имен еще предстоит изучить. Еще одной иллюстрацией важности учета семантики КГ могут служить следующие предложения:

20. *Они многие годы боролись с несправедливостью.* [А3-АКТ]
 21. *Они боролись с Васей два часа.* [А2-СИММ]
 22. *Они боролись с удовольствием.* [А3-АКТ/А4-СИРК]

В примере (18) представлен симметричный предикат *бороться*, множественная валентность которого заполнена отдельно: подлежащим *они* и КГ *с Васей*. В предложении (17) находим омонимичный несимметричный предикат: перифраза *[*Они с несправедливостью*] *боролись многие годы* недопустима. На это указывает тип комитативной группы (неодушевленная, абстрактная; не номинализация). В то же время КГ в примере (19) также является неодушевленной и абстрактной и не является номинализацией, но обозначает психологическое состояние человека, и потому предрасположена к сирконстантной функции. При этом, строго говоря, предложение (19) омонимично: интерпретация КГ с удовольствием как актанта глагола *бороться* возможна, и трудно сформулировать признак, который исключил бы актантную трактовку в (19), сохранив ее в примере (17).

Литература

1. Золотова Г. А. (2001) Синтаксический словарь: Репертуар элементарных единиц русского синтаксиса. Изд. 2-е, испр. — М.: Эдиториал УРСС.
2. Архипов А. В. (2003) Проблемы автоматического анализа комитативных групп в русском языке. // И. М. Кобозева, Н. И. Лауфер, В. П. Селегей (ред.). *Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции Диалог'2003 (Протвино, 11–16 июня 2003 г.)*. М.: Наука. С. 36–42.
3. Архипов А. В. (в печати) Грамматика и семантика русского комитатива (предлог *с* + творительный падеж): комментарии к «Синтаксическому словарю». // *Труды Международной конференции «Грамматика славянского предлога» (Москва, 28–30 октября 2003 г.)*. М.: МГУ.