

Проблемы автоматического анализа комитативных групп в русском языке

Александр Владимирович Архипов

МГУ им. М. В. Ломоносова

1. Комитатив и комитативная группа в русском языке

В русском языке комитативными группами (КГ) мы будем называть предложные группы вида «предлог *с* + ИГ в творительном падеже» (напр. *с вареньем, с женой, с удовольствием*).

С точки зрения типологии целесообразным представляется различие "комитатива в узком смысле" и "комитатива в широком смысле". В произвольном языке *L* комитатив в узком смысле предлагается выделять на основе определения комитативной стратегии соединения именных групп (далее СИГ) [Stassen 2000]: если один и тот же предикат предидируется одновременно относительно двух самостоятельно выраженных участников, ИГ, выражающие этих участников, могут иметь одинаковый «структурный ранг», и тогда мы говорим о сочинительной стратегии СИГ; или же иметь различный «структурный ранг», и тогда мы имеем дело с комитативной стратегией СИГ.

При комитативной стратегии СИГ одна из ИГ-участниц конструкции занимает обычно позицию подлежащего или же прямого дополнения, а другая — комитативная — всегда позицию косвенного дополнения.

Комитативными конструкциями (КК) в широком смысле являются все конструкции, формально тождественные КК в узком смысле (использующие те же морфосинтаксические средства, которыми в последних маркируется комитативная ИГ). Так, для русского языка в (1) имеем КК в узком смысле, а в (2, 3, 4) — КК в широком смысле.

- (1) *Петя* пришел к нам *со своей младшей дочерью*.
- (2) *Петя* целуется *с Машей*.
- (3) *Петя с Машей* по четвергам ходят в театр.
- (4) *Петя с улыбкой* попрощался.

2. Проблемы автоматического анализа комитативных групп

Вследствие широкого спектра возможных значений и синтаксических функций КГ, для систем автоматического понимания текста (САПТ) нетривиальной задачей является установление как синтаксической принадлежности КГ, так и их места в семантической и информационной структуре текста. Проблемы с толкованием КГ нередко являются предметом языковой игры, отдельные казусы описаны в художественной литературе (*пирог с околоточным надзирателем*).

Замысел статьи возник во многом благодаря спецкурсу Н.Н.Леонтьевой "Семантический компонент в системах автоматического понимания текста", читавшемуся осенью 2002 г. на ОТиПЛ МГУ. Ниже будут изложены некоторые соображения касательно возможностей возникновения неоднозначности при анализе предложений, содержащих комитативные группы, а также некоторые общие рекомендации по разрешению этой неоднозначности. В первую очередь, но не исключительно, изложенное может быть полезно при разработке многоуровневых систем АПТ (в частности, системы Н.Н.Леонтьевой), включающих отдельные уровни синтаксического и семантического анализа.

3. Функции комитативных групп в русском языке

В первом приближении, за вычетом некоторых частных случаев, комитативные группы в русском языке могут занимать синтаксические функции актанта, сирконстанта, распространителя ИГ, а также элемента квазисочиненной ИГ. В первых двух случаях КГ формально подчиняется предикату, в последних двух — вершине ИГ (имени или местоимению).

Определение синтаксической функции КГ необходимо в силу того, что от нее зависит, в какую часть семантического представления текста попадет информация о соответствующем участнике ситуации, и в этом синтаксический компонент САПТ помогает семантическому; далее семантический компонент определяет семантическое отношение, связывающее ее с этим узлом.

3.1. Комитативная группа синтаксически подчиняется (главному) предикату

A1. Собственно комитативная конструкция

Петя пришел к нам со своей младшей дочерью.

Конструкции такого типа мы будем называть собственно комитативными конструкциями, или КК в узком смысле.

Синтаксис. КГ не соответствует никакой валентности, определенной в модели управления (МУ) предиката. На поверхностном уровне КГ следует рассматривать как сирконстант.

Семантика. Участник, выражаемый комитативной группой, не соответствует никакой фиксированной (для всех употреблений) семантической роли. Скорее всего, в этом случае можно говорить о "расщеплении" одной из валентностей предиката, так что роль "копируется" с другого участника (выраженного "базовой" ИГ конструкции), например, агенса. Такие конструкции предлагается считать одним из способов "омножествления" одного из типовых участников ситуации наряду с сочинением ИГ, грамматическим и лексическим множественным числом.

(схема)

Давно отмечено, что значение предложения, по крайней мере пропозициональный компонент, сохраняется при следующих преобразованиях, что является аргументом в пользу тождества семантических ролей участников, выраженных базовой именной группой (БИГ) и КГ:

Петя вернулся с Мариной в Киев в четверг. □ Марина вернулась с Петей в Киев в четверг. □ Петя и Марина вернулись в Киев в четверг.

Кроме того, КГ может соотноситься отнюдь не только с одушевленным агенсом, но также, например, с экспериенцером или с пациенсом.

Различия между сочинением и КК усматриваются прежде всего в коммуникативной структуре. ИГ, образующие комитативную конструкцию, могут приобретать различный коммуникативный статус независимо друг от друга, в то время как сочиненные ИГ имеют однородные характеристики. В типичных случаях () БИГ является частью темы, а КГ частью ремы высказывания.

Еще одним отличием КК от сочиненных ИГ является возможный в КК (но не обязательный) семантический компонент совместности.

Лексическая сочетаемость. Наиболее часто КК в узком смысле употребляются с глаголами движения; не сочетаются с экспериенцерами ментальных предикатов (**Я знаю с Петей экономику; *Игорь боится с Петей инопланетян; ср. Мы с Петей знаем экономику; Игорь с Петей боятся инопланетян*).

Сочетаются с глаголами направленного восприятия (слушать vs слышать и т.п.).

*В это время год назад он любовался закатом с Машей. □ *В это время год назад он*

восхищался закатом с Машей.

A2. Комитативная группа при симметричных предикатах

Примеры:

Петя целовался с Машей. / Петя и Маша целовались / Они целовались.

В Англии скрестили две породы кошек. / В Англии эту породу скрестили с другой.

Они играли в шахматы. / Петя и Маша играли в шахматы. / Петя играл в шахматы с Машей.

Синтаксис. Так называемые симметричные предикаты (напр. *целоваться, толпиться, пересекаться, соединять* и т.д.) характеризуются тем, что одну из (семантических) валентностей в них должен заполнять множественный участник. Выражена эта множественность может быть любым способом: лексически, с помощью граммы множественного числа, с помощью сочинительной конструкции или же комитативной.

Семантика. Как и в случае КК в узком смысле, участник, выраженный КГ, имеет такую же семантическую роль, как и "базовый" участник. Разница состоит в том, что, во-первых, КГ присутствует в МУ предиката, во-вторых, предложения с симметричными предикатами не содержат соединения ИГ (в определении Стассена), поскольку из *Петя целовался с Машей* не следует независимых друг от друга *?Петя целовался* и *?Маша целовалась*: предикат *целоваться* не определен для единичных участников, по крайней мере, в первом значении.

Следует иметь в виду, что многие симметричные предикаты имеют омонимичные несимметричные варианты (ср., например, *Они тепло поздоровались [друг с другом]* vs *Я с ним поздоровался, а он со мной нет*), в связи с чем возможны проблемы с атрибуцией КГ-сирконстанта, напр. *Он поздоровался с ядовитой усмешкой*.

Особый случай представляют собой сочетания "друг с другом" и "сам с собой". Синтаксически их, видимо, следует для наших целей рассматривать как творительный падеж лексем *друг##друг* и *сам##себя* с предлогом *с*.

"Сам с собой" употребляется при симметричных предикатах и указывает на то, что множественный участник предиката формируется из участника, выраженного БИГ, и его "копии". "сам с собой" (актант при симметричных предикатах)

Синт.: БИГ <-- ГЛ --> с --> сам##себя [твор. пад.]

Сем.: ГЛ --> МНУ(БИГ, сам##себя)

Что же касается сочетания "друг с другом", оно употребляется не только при симметричных, но вообще при совместных предикатах, и должно быть скорее отнесено к сирконстантам (см. п. А4).

A3. Комитативная группа □ актант несимметричного предиката

Комитативом в русском языке управляют самые разнообразные предикаты, и семантическая роль участника, выраженного КГ, может быть различна.

пациенс

происходить, случаться, приключаться

Недавно со мной произошла прелюбопытная история.

быть, стать, оказаться (в безличной форме) + предикатив (или вопр. мест.)

Что с тобой? С ним все будет в порядке.

С договором все оказалось очень непросто.

делать, сотворить

Что же мне с тобой делать?

пациенс/причина

работать, возиться, копошиться

Он провозился с документами и опоздал на последний автобус.

контрагент

изменять, обманывать

Он изменял жене с секретаршей.

содержание

относиться

До этого случая он относился ко мне с презрением.

А4. Комитативная группа □ сирконстант

КГ-сирконстанты чаще всего являются свернутыми пропозициями: *с улыбкой (улыбнулся), с восхищением (восхищался), с испугом (испугался).*

Кроме того, КГ может называть некоторый параметр ситуации, выраженной предикатом:

Он бежал от нас с быстротою страуса. (Ф. Искандер)

Особый случай: "друг с другом" (см. тж. п. А2).

Напарники поделились друг с другом своими соображениями.

В отличие от "сам с собой", "друг с другом" употребляется (а) не только при симметричных предикатах, и (б) тогда, когда валентность множественного участника уже заполнена, поэтому само это сочетание валентности не занимает, а является сирконстантом.

Синт.: БИГ <-- ГЛ --> с --> друг##друг [твор. пад.]

Сем.: БИГ <--(акт)-- ГЛ --(сирк)--> "с + друг##друг"

3.2. Комитативная группа подчиняется имени

Б1. Квасисочинительная конструкция

[Петя с женой] уехали в Прагу.

Квасисочинительные конструкции морфологически выглядят как комитативные, однако проявляют все типичные свойства сочинительных. БИГ и КГ образуют единую составляющую, т.е. вызывают согласование предиката по множественному числу, не допускают линейных разрывов, не подвержены топиализации, фокализации и под.:

**Петя с кем уехали в Прагу?*

**Петя даже с женой не поехали в Прагу.*

Б2. Конструкции с поглощенным референтом (Plural Pronoun Constructions)

[Мы_(i+j) с братом_(j)] ловили рыбу напротив дока. (Ф. Искандер)

Конструкции с поглощенным референтом синтаксически являются разновидностью квазисочинительных конструкций. Их особенность заключается в особом референциальном устройстве. БИГ в этих конструкциях является личным местоимением множественного числа (*мы, вы, они*), причем референт КГ входит в референт местоимения.

В подавляющем большинстве случаев КГ содержит одушевленное существительное (местоимение). Конструкции с поглощенным референтом допускают вставку частиц, напр. *же, ведь, то* и т.д., между БИГ и КГ.

Б3. Комитативная группа □ модификатор ИГ

На каждом столике стоял плакатик с указанием факультета. (Ф. Искандер)

В этом классе конструкций КГ играет роль определения к существительному. Наиболее частотными отношениями являются: часть-целое, содержимое-контейнер, изображение/текст/информация-носитель.

КГ линейно примыкает к определяемой ИГ или следует за другими модификаторами

4. Примерный алгоритм распознавания функции комитативных групп в предложении

Рассматриваются различные «ключевые» грамматические и лексические признаки ИГ

(как комитативной, так и соотнесенной с ней «базовой» ИГ, если она имеется) и вершинного предиката, которые помогают проинтерпретировать текст, содержащий КГ. Среди «ключей» — характеристики личности / одушевленности, абстрактности / конкретности ИГ, модель управления предиката (в частности, признак симметричности), наличие уточняющих наречий и другие параметры. Важную роль играет близость характеристик «базовой» и комитативной ИГ.

1. Частные случаи

1.1. "друг с другом" (тип А4, сирконстант, "уточнение референции").

1.2. "сам с собой" (тип А2, актант при симметричном предикате).

2. Общие случаи

2.1. КГ □ актант?

(1) Имеется ли КГ в модели управления (главного) предиката?

ДА - (2).

НЕТ - (6).

(2) Предикат симметричный?

ДА - (3).

НЕТ - тип А3.

(3) Имеют ли КГ и БИГ однородные характеристики (личность/одушевленность, конкретность/абстрактность)?

ДА - тип А2.

НЕТ - (4).

(4) Имеется ли омонимичный несимметричный предикат?

ДА - (5).

НЕТ - (10).

(5) Соответствуют ли характеристики ИГ требованиям несимметричного предиката?

ДА - тип А3.

НЕТ - (10).

2.2. КГ подчиняется имени?

(6) Следует ли КГ непосредственно за к.-л. ИГ?

ДА - (7).

НЕТ - (10).

(7) Эта ИГ □ личное местоимение мн.ч., а КГ является одушевленной?

ДА - тип Б2.

НЕТ - (8).

(8) Эта ИГ является подлежащим, имеет однородные характеристики с КГ, а предикат согласуется по множественному числу?

ДА - тип Б1.

НЕТ - (9).

(9) Эта ИГ является подлежащим, имеет единственное число, а предикат согласуется по множественному числу?

ДА - тип Б1.

НЕТ - (10).

2.3. КК в узком смысле, сирконстант или модификатор ИГ?

(10) КГ содержит предикатное имя (категория □ речементальные предикаты, эмоция, чувство, поведение, жест) - тип А4.

(11) КГ содержит предметное имя, неодушевленное; входит с предшествующей ИГ в отношения часть-целое, содержимое-контейнер - тип А7.

(12) КГ содержит предметное имя, имеет однородные характеристики (личность; личность или одушевленность; неодушевленность) с одним из актантов предиката - тип A1.

Приведенный выше алгоритм является лишь примерным, требует дальнейшей формализации и уточнения, а также настройки под конкретную систему. Возможна модификация алгоритма на основе статистических методов с использованием тренировочного корпуса, содержащего предложения с комитативными группами.

В сложных случаях необходима "ручная" корректировка работы алгоритма.

Автор выражает благодарность Н.Н.Леонтьевой и И.М.Кобозевой за обсуждение отдельных проблем, затронутых в статье.

При подготовке были использованы материалы морфологически размеченного корпуса русских художественных текстов (<http://corpora.yandex.ru>).