

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2017»

Москва, 31 мая — 3 июня 2017

КАК НАМ БЫТЬ С КОНСТРУКЦИЯМИ ТИПА *КАК БЫТЬ?*¹

Иомдин Л. Л. (iomdin@iitp.ru)

Институт проблем передачи информации
им. А. А. Харкевича РАН, Москва, Россия

В статье рассматривается ряд сложных микросинтаксических единиц русского языка, образованных служебными словами и словами широкой семантики, такими как местоимения, союзы, частицы и связочные глаголы. Одна микросинтаксическая единица, *как быть*, описывается весьма подробно. Рассматриваются проблемы адекватного представления таких единиц в лингвистических ресурсах и в задачах автоматической обработки текстов. Особое внимание уделяется двум типам ресурсов: Микросинтаксическому словарю и аннотированному корпусу русских текстов «СинТагРус», в который встроена микросинтаксическая разметка.

Ключевые слова: микросинтаксис, синтаксические фраземы, микросинтаксический словарь, корпус с микросинтаксической разметкой

WHAT TO DO ABOUT CONSTRUCTIONS LIKE *WHAT TO DO?*

Iomdin L. L. (iomdin@iitp.ru)

Institute for Information Transmission Problems (Kharkevich
Institute), Russian Academy of Sciences, Moscow, Russia

¹ Работа выполнена при поддержке Российского гуманитарного научного фонда (грант № 15-04-00562). Автор выражает фонду искреннюю признательность.

The paper discusses a number of challenging microsyntactic units of Russian formed with functional and broad semantics words, like pronouns, particles, conjunctions and auxiliary verbs. One microsyntactic unit, “*kak byt'*” (≈ “what to do”), is described in full detail. Issues of adequate presentation of such units in linguistic resources and their treatment in natural language processing tasks are considered. Special emphasis is laid on two types of resources: the Microsyntactic dictionary, developed as a specific lexicographic resource that can be used in theoretical linguistics and natural language processing, and a deeply annotated corpus of Russian, SynTagRus, supplied with microsyntactic tagging. The Microsyntactic dictionary consists of lexical entries compiled according to a uniform pattern that comprises 9 zones: (1) name of microsyntactic unit; (2) type of unit (syntactic idiom or nonstandard construction); (3) lexical composition of the unit (individual words of the syntactic idiom or classes of words forming the nonstandard construction, if any); (4) analytical lexicographic definition of the unit; (5) government pattern; (6) syntactic properties of the unit presented in terms of dependency grammar, including the identification of syntactic relations that hold between the unit's constituent elements, and relations that incorporate this unit into the sentence; (7) synonyms and analogues; (8) comments and (9) illustrations. The microsyntactic markup of SynTagRus shows what units are present, how they fit in the structure and which of the polysemous entities occur in the tagged sentence.

Keywords: Microsyntax, syntactic idioms, microsyntactic dictionary, corpus tagged for microsyntactic elements

1. Вводные замечания. Микросинтаксические конструкции, образованные служебными словами

Среди конструкций, так или иначе относящихся к сфере фразеологии, особое место занимают такие, которые составлены из слов весьма широкой или размытой семантики. Часто это служебные слова — местоимения, союзы, предлоги, частицы, связочные глаголы. Эти конструкции отличаются несколькими существенными особенностями, из которых мы упомянем четыре:

- (1) повышенный уровень некомпозициональности (смысл конструкции может быть настолько сложным и специфичным, что его невозможно вывести из смысла входящих в нее элементов даже посредством цепочки семантических переходов);
- (2) повышенный уровень неоднозначности (такие фразеологические конструкции могут иметь более одного значения сами по себе и иметь омонимичные свободные словосочетания);
- (3) повышенная употребительность; образно говоря, у конструкций, собранных из служебных слов, не было бы шанса утвердиться в языке, если бы они не использовались его носителями на каждом шагу;
- (4) повышенный уровень синтаксической сложности (во многих из таких конструкций нелегко определить типы синтаксических связей между их составными элементами или даже сам факт наличия таких связей, а в некоторых случаях бывает трудно установить даже частеречный статус этих элементов).

Примерами этих конструкций являются такие микросинтаксические единицы, как составные частицы *пока что, только что, разве что*, многозначные наречные предложения *все равно* или *как раз*, союз *как будто*, многозначная единица *как бы* (частица или союз), конструкции типа *вот оно что, ни за что* и многие другие.

Одни из этих конструкций несомненно содержат несколько отдельных слов, другие (скажем, частицы *как бы* или *что ли*) в процессе языковых изменений движутся по направлению к единому слову подобно тому, как это произошло с союзом/наречием *якобы* (которое, согласно данным НКРЯ, могло писаться раздельно вплоть до начала XX века) и подобно партикулам Т. М. Николаевой (2008), таким как *либо* и *неужели*, исторически образовавшимся из нескольких отдельных формантов.

В данной группе конструкций интересную подгруппу образуют единицы, в составе которых присутствуют связочные глаголы, прежде всего *быть* и *стать*. Даже при самом беглом взгляде на единицы *так и быть, была не была, будь что будет, стало быть, не стало (такого-то), будешь (чай, кофе и т.п.)* становится ясно, что в их значение входят семантические компоненты, далеко выходящие за пределы семантики связочности.

Так, в предложениях

- (1) — *Удостоверений мы обычно не даём, — ответил кот, насупившись, — но для вас, **так и быть**, сделаем исключение* (М. А. Булгаков)

и

- (2) *Эй! вы там! Подайте мне корону. **Так уж и быть**, остаюсь на престоле* (Е. Л. Шварц)

вводный оборот *так и быть* обозначает **согласие говорящего выполнить просьбу слушающего после уговоров последнего и собственных колебаний** (то, что в «Золушке» Шварца король притворяется, что к нему поступила просьба не отрекаться и что он колебался, выполнить ли ее, только подтверждает верность толкования).

В предложениях

- (3) *В театре я не находила такой роли и поэтому однажды — **была не была!** — решила сниматься в кино* (Г. Александров),
- (4) *Не хочется попадаться немцам в руки. Ну ладно, думаю, **была не была!** Высунулся на улицу* (Ф. Искандер),
- (5) *Оставалась надежда только на «**была не была!**»* (Е. Евтушенко)

вводный оборот *была не была*, удивительным образом составленный из двух глаголов *быть* в женском роде, которые, несмотря на очевидную перспективность, имеют форму прошедшего времени, означает, что **говорящий готов совершить действие, допускающее два противоположных исхода и чреватое для него большим риском, нередко смертельным** (недаром этот оборот часто используется в военной прозе).

А конструкция *будь что будет*, в некоторых случаях означающая почти то же самое, что и *была не была*, а именно готовность идти на серьезный риск, как в примерах

- (6) *Несколько дней Глаша не была у себя в подвале, ночевала в прачечной, на гладильном столе, и как-то под утро, вдруг проснувшись, решила: будь что будет, пойду в полицию.* (Л. Юзефович),
- (7) *Схватился за голову, сел в машину. Эх, будь что будет! Поехал на малой скорости домой.* (Н. Мордюкова)

в целом значительно отличается от *была не была*, во-первых, выражая не столько готовность к риску, сколько **граничащее с фатализмом безразличие к результату действия или бездействия говорящего**, и, во-вторых, необязательной перспективностью: можно что-то сделать или решить, а потом сказать *будь что будет*. Интересно встречающееся порой столкновение этих конструкций в одном тексте, как в любовном стихотворении С. Гандлевского, где легкомысленность усиливается до безразудства:

- (8) *...И в окне зацвела темнота. Будь что будет, вернее, была не была — Выцветай, как под утро фонарь на столбе, Доцветай, как сирень на столе доцветала... И сиреневый сор я смахнул со стола.*

В предложении

- (9) *Так вот её-то внучонок, мой, стало быть, правнучек, прибежал ко мне* (П. П. Бажов)

вводный дискурсивный оборот *стало быть* обозначает, что говорящий усматривает **логическое следствие** между двумя ситуациями, эксплицитно или имплицитно содержащимися в тексте.

Любопытно, что у этого оборота есть два однословных синонима *следовательно* и *значит*, которые не вполне взаимозаменяемы. К примеру, *стало быть* в

- (10) *Ну вот, стало быть, я иконку на грудь пришпилил и побежал...* (М. А. Булгаков)

можно заменить на *значит*, но вряд ли — на *следовательно*.

Стало быть состоит из двух глаголов с жестким порядком следования и с жестко фиксированными формами — *стать* в среднем роде прош. времени и *быть* в инфинитиве.

В противовес такой жесткости словоформ конструкция *не стало*, управляющая подлежащим исключительно в родительном падеже и имеющая смысл 'умер', допускает колебание формы времени глагола *стать*; ср.

- (11) *Через полгода у него внезапно обнаружили рак, и вскоре его не стало* (А. Городницкий)

и

- (12) *Завтра, как только ее не станет, я тоже умру.* (Е. Шварц).

Добавим, что в этой конструкции *стать* может даже оказаться в инфинитиве в контексте модального глагола:

- (13) *Ей было не до нее, мама тогда была совсем плоха, и однажды Ольга вдруг ясно увидела, что мамы может не **стать*** (Г. Щербакова).

А вот причастные и деепричастные варианты этой конструкции начисто исключены: **Этот больной, не ставший (не став) сегодня утром...*

Наконец, разговорная конструкция с глаголом *быть*, обозначающим в первом приближении **потребление еды и питья**, обладает рядом уникальных сочетаемостных, семантических и прагматических свойств.

В частности, она содержит глагол *быть* исключительно в личной форме и будущем времени; ср.

- (14) *Я буду грушу и шарлотку,*

но не **Я был грушу и шарлотку*, **Я могу <хотел бы> быть грушу и шарлотку* и не **Будучи только грушу и шарлотку, ты останешься голодным.*²

Во вторых, этот глагол, в отличие от любых базовых лексических значений *быть* — экзистенциального, связочного, локативного — допускает в нашей конструкции прямое дополнение в винительном падеже (ср. пример (14)) или — при отрицании — в родительном падеже (ср. (15)),

- (15) *Я сказал: «Не буду каши, буду есть сегодня мед»* (Б. Симкин).

которое может заполняться словом со значением пищи, питья, а также, неожиданным образом, со значением курева: *Будешь сигаретку?* (Par extension, здесь могут появляться и слова со значением меры и количества, а также местоимения (*Буду рюмку коньяку; Буду пару пирожков; Вы что будете; Вы что-нибудь будете?* — *Ничего не буду, спасибо*). Расширение этого семантического класса невозможно: нельзя сказать что-либо вроде **Я буду инъекцию аспирина*.

В третьих — и это представляется самым любопытным — смысловой компонент 'потреблять (еду или питье)' не исчерпывает значения нашей конструкции. Она употребляется тогда, когда потребитель оказывается в той или иной **актуальной ситуации обслуживания**: ему предлагают еду (или питье, или закурить), а он принимает или не принимает это предложение. Предлагать — вербально и невербально — может кто угодно: официант в ресторане, хозяйка, кормящая семью обедом, или приятель. При этом неважно, каков статус этого потребителя: это может быть обычный человек, маленький ребенок, не умеющий говорить, или даже животное: можно сказать *Мальчи вряд ли будет эту молочную смесь* или *Кот ни за что не будет Вискас, даже не предлагай ему*. (Разумеется, в последних двух случаях вряд ли будет уместен глагол в первом или во втором лице). Невозможно представить высказывания типа *Смотри, бабушка завтракает*. **Сейчас она будет чай с молоком*. **Я скоро приду домой*

² Данное ограничение можно было бы объяснить эллипсисом полнозначного глагола, пусть и лексически неопределенного: *буду грушу и сок < буду есть грушу и пить сок*, если бы значения таких выражений были тождественными. Однако, как мы увидим ниже, это далеко не так.

и буду пельмени (отсутствует обслуживающее лицо) или даже *#Я скоро приду домой, жена мне приготовит обед, и я буду пельмени* (отсутствует актуальность).³

Приведенные примеры убедительно демонстрируют, насколько прихотливо и непредсказуемо могут быть устроены фраземные конструкции, сформированные из служебных слов, и какого детального описания они требуют. Разумеется, эта цель становится еще более сложной при решении компьютерно-лингвистических задач, в частности, задачи глубокого семантического анализа.

2. Микросинтаксический словарь и конструкция «как быть»

Предпримем теперь попытку описать одну из таких конструкций с несколько большей степенью подробности. Это описание выполняется в рамках создаваемого под руководством автора Микросинтаксического словаря русского языка, который мыслится как отдельный лингвистический ресурс, призванный вывести на новый уровень описание синтаксической фразеологии. Важным дополнением к этому ресурсу служит производимая в настоящее время микросинтаксическая разметка глубоко аннотированного корпуса текстов СинТагРус (см. Маракасова — Иомдин 2016, Iomdin 2016.)

Микросинтаксическая разметка корпуса будет кратко охарактеризована в следующем разделе.

Статьи Микросинтаксического словаря составляются по следующей схеме, в значительной мере опирающейся на опыт создания нового Активного словаря русского языка под ред. Ю. Д. Апресяна (2014):

- а) имя микросинтаксической единицы;
- б) тип микросинтаксической единицы (синтаксическая фразема или нестандартная синтаксическая конструкция)⁴;
- в) лексический состав единицы (конкретные слова синтаксической фраземы или классы слов, входящих в состав нестандартной синтаксической конструкции, если таковые имеются);
- г) лексикографическое толкование единицы (по возможности, аналитическое);
- д) модель управления (если она имеется);

³ Эта конструкция исследовалась автором в некоторых публичных выступлениях и лекциях. Результаты исследования были учтены (и несколько переосмыслены) в статье Апресян 2010 (с соответствующей ссылкой на эту неопубликованную работу) и в написанной Ю. Д. Апресяном словарной статье глагола БЫТЬ Активного словаря русского языка. Сам автор публикует, однако, эти результаты впервые.

⁴ Синтаксическая фразема — это лексикализованная единица, образованная из конкретных слов (таковы все рассмотренные в разделе 1 единицы); нестандартная синтаксическая конструкция не привязана к конкретным словам (например, конструкции с повторяющимися элементами, такие как *X-овать не X-овал* ‘не X-овал, делал что-либо более слабое, чем X’, ср. *Дружить не дружили, но были приятелями*). Подробнее см. Иомдин 2013, 2014.

- е) синтаксические свойства единицы, формулируемые в терминах грамматики зависимостей и включающие идентификацию синтаксических отношений между элементами единицы и синтаксических отношений, встраивающих данную единицу в предложение;
- ж) синонимы и аналоги;
- з) комментарии;
- и) иллюстрации.

Приведем (в несколько сокращенном варианте, в неформальном виде и с необходимыми комментариями) статью Микросинтаксического словаря, представляющую синтаксическую фразу *как быть*. Эта синтаксическая фраза выступает в таких предложениях, как *Как нам быть с завтрашним семинаром?*, *Никто не знает, как быть в таких ситуациях* и т.д.

Имя: КАК БЫТЬ

Тип: синтаксическая фраза

Лексический состав: БЫТЬ, КАК⁵

Лексикографическое толкование:

Как X-у быть с Y-ом в P?

Пресуппозиция:

Лицо X является участником ситуации P, в которой имеет место факт, предмет или лицо Y. Говорящий считает, что Y создает для X-а трудности в P'.

Ассерция:

Говорящий хочет знать, что лицо X должно сделать, чтобы Y перестал создавать для X-а трудности в ситуации P'.

Модель управления.

В семантической структуре микросинтаксической единицы *как быть* имеется три переменных, соответствующих (1) валентности ситуации P, (2) валентности агенса X и (3) валентности объекта («помехи») Y.

В высказываниях, содержащих *как быть*, могут быть насыщены все три эти валентности или часть их, а может быть не насыщена ни одна:

- (16) *Как преподавателю (X) быть на экзамене (P) со студентами (Y), которые пользуются шпаргалками?* (выражены X, P и Y);
- (17) *Посоветуйте, как быть в такой ситуации (P)* (не выражено ни X, ни Y);
- (18) *Но как мне (X) быть с моей грудной клеткой И с тем, что всякой косности косней (Y)?* — Б. Л. Пастернак (не выражено P);
- (19) *Не знаю даже, как и быть* (не выражено ни P, ни X, ни Y).

Тем самым все валентности *как быть* являются факультативными.

⁵ КАК2 — местоименное наречие, в отличие от союза КАК1. Идентификация лексических единиц производится по комбинаторному словарю лингвистического процессора ЭТАП-3 (Iomdin et al. 2012), на основе которого строится и аннотированный корпус СинТагРус (Дяченко и др. 2015).

Примечание: В полной семантической структуре, содержащей эту единицу, непременно должна фигурировать еще одна переменная, соответствующая слушающему, или контрагенту X2, от которого ожидается, что он сообщит X-у запрашиваемую им информацию. Однако эта переменная характеризует любые вопросительные высказывания и не должна входить в валентную структуру конкретных языковых единиц, в том числе и синтаксической фраземы как *быть*, несмотря на её ингерентную вопросительность.

2.1. Синтаксические свойства

Обязательные элементы фраземы — глагол БЫТЬ, принимающий форму инфинитива, и вопросительное наречие КАК2. Последнее подчиняется инфинитиву БЫТЬ по обстоятельственному синтаксическому отношению (СинтО)⁶.

Валентность субъекта X выражается именной группой в дательном падеже. В синтаксической структуре эта именная группа подчиняется глаголу БЫТЬ по дательно-субъектному СинтО.

Валентность объекта Y выражается группой, вводимой предлогом С, который в синтаксической структуре подчиняется глаголу БЫТЬ по 1-му комплетивному СинтО (т.е. отношению, которое вводит первое дополнение). Тем самым имплицитно предполагается, что в пределах данной микросинтаксической единицы БЫТЬ имеет особое лексическое значение, не являющееся ни экзистенциальным, ни связочным, ни локативным (ср. аналогичное решение, принятое для БЫТЬ в рассмотренных в разделе 1 конструкциях типа *буду кофе и булочку*). Эта лексическая единица семантически и синтаксически близка к глаголу ПОСТУПАТЬ (хотя, разумеется, не тождественна ей).

Валентность ситуации Р выражается достаточно разнообразно. Помимо групп, вводимых локативными предлогами В, НА (ср. примеры (17) и (18) выше), а также ПРИ, как в

(20) *Петр Ильич — тот в письмах к благодетельнице фон Мекк высказывал свои соображения о музыке и искусстве, изливал душу, а как быть при телефонах?* (В. Орлов)

она выражается еще и придаточными условия и цели, вводимыми союзами ЕСЛИ, КОГДА и ЧТОБЫ:

(21) *Но как же быть, если отказать высказанным Бердяевым суждениям в принудительной силе?* (Л. И. Шестов);

(22) *Долго сидела она с Лидией Александровной... и советовалась о жизни: как ей быть, когда Николай придет?* (Ю. Трифонов);

⁶ В описании единиц Микросинтаксического словаря используется синтаксический формализм (в том числе и арсенал синтаксических отношений), разработанный для лингвистического процессора ЭТАП-3 (см. о нем, например, Iomdin et al. 2012). Тем самым обеспечивается совместимость этого словаря со всеми компонентами процессора.

- (23) *Как быть с таким человеком, чтобы дом показался ему именно домом, а не дворцом?* (Н. Г. Чернышевский).

Валентность Р представляется разумным в синтаксической структуре подчинять глаголу *БЫТЬ* по 2-му комплетивному отношению.

2.2. Синонимы и аналоги

Точного синонима у синтаксической фраземы *как быть* нет. Аналогом является полусвободное сочетание *что делать: Как мне быть с опоздавшими? ≈ Что мне делать с опоздавшими?*

2.3. Комментарии

1. Микросинтаксическая единица *как быть* абсолютно не допускает отрицания при глаголе: **Как мне не быть с этой проблемой?* Выражение *Как не быть* — это реализация совершенно другой конструкции, имеющая значение, близкое к 'естественно быть', ср.

- (24) *А девушки там рядом есть? — Как не быть...*
(А. Стругацкий, Б. Стругацкий),

где *быть* имеет экзистенциальное значение, и

- (25) *В своей поэзии Бунин сумел сделать много прекрасного. Как не быть ему благодарным?* (В. Ходасевич),

где *быть* выступает в связочном значении: и в (24), и в (25) фигурируют обычные лексические значения *быть*, а не то фразеологическое \approx 'поступать', которое имеет место в данной синтаксической фраземе. Добавим, что скорее всего, мы имеем тут дело с другой синтаксической фраземой *как не + X-овать* (\approx 'естественно, что имеет место X'), которая формирует риторический вопрос.

2. Синтаксическая фразема *как быть*, в отличие от только что рассмотренной *как не + X-овать*, в общем случае формирует обыкновенный, риторический вопрос. Тем не менее из материала этого вопроса практически невозможно породить стандартный ответ, заменив в нем вопросительное слово. Трудно представить себе диалог типа *Как быть? — Быть так-то и так-то*. Редким исключением является диалог типа *Как быть? — (А) никак не быть*.

3. Единица *как быть*, как ясно из примеров в зоне модели управления, перенимает обычное для инфинитива управление дательным падежом субъекта. Что касается управления предложением *с*, вводящим объект, то такое управление свойственно глаголу *быть* исключительно в составе данной микросинтаксической фраземы. Вне этой фраземы стоящий при *быть* предлог *с*, вводящий именную группу в творительном падеже, выступает в самостоятельном значении, чаще всего в значении совместности, а само *быть* утрачивает связь с идеей поведения, ср.

(26) *Где нам быть с братом?*

(27) *Дети могли быть с няней*

и даже

(28) *Как нам быть с детьми в таком маленьком номере?*

4. Глагол *быть* в составе данной синтаксической фраземы может выступать в качестве вершины главного или придаточного предложения, но иногда может и подчиняться глаголу во втором лице из небольшого списка; ср.

(29) *Как вы посоветуете <порекомендуете> быть в такой ситуации?*

(30) *Как прикажете быть с соблюдением авторских прав?*

В (30) имеет место своеобразная формула, в которой обычный смысл глагола *приказывать* выветривается, так что как *прикажете быть* означает примерно ‘как вы ответите на вопрос „как быть“’.

Изредка вершиной предложения, в которое входит как *быть*, может выступать глагол *было* в прошедшем времени, обращая всю ситуацию в прошлое, ср.

(31) *Что же. Как было быть князю. Он мужчина еще не старей, в поре, про которую сказано: «Не добро быть человеку одну...» (Ф. Н. Плевако).*

(32) *Вот как было быть С.А., когда этот граф днем писал одно, а по ночам делал другое, а потом наутро в дневнике плохими словами осыпал жену за то, что она, видите ли, вызывает у него «дурные» мысли (из отзыва о книге Софьи Андреевны Толстой).*

5. Два основных словесных элемента фраземы как *быть* почти не имеют собственных синтаксических зависимых, не являющихся зависимыми всей фраземы в целом. Исключения составляет частица *же* при как и частица *–то* при *быть*, вместе или порознь: последовательности как *же быть*, как *быть-то*, как *же быть-то* весьма частотны. Спорадически встречаются и другие зависимые, ср. наречие *правильно* в (33)

(33) *И как правильно быть, если мать продала нас цыганам, нет, хуже — бросила собакам? (А. И. Солженицын).*

Такие случаи представляются достаточно экзотичными.

6. Как говорилось в зоне толкования, смысл единицы как *быть* в ассертивной части содержит компонент ‘хочет знать’. Отсюда вытекает, что придаточное предложение с как *быть* может подчиняться далеко не любому предикату главного предложения, способному управлять инфинитивным оборотом с вопросительным словом. Нормально сказать *Спроси, как мне быть; Не скажу даже, как вам быть с билетом*, но ненормально выразиться **Мне все равно <безразлично, неважно>, как ему быть, *Я забыл, как мне быть*, поскольку при этом возникает смысловое противоречие.

2.4. Иллюстрации и примеры

Микросинтаксическая единица **как быть** весьма частотна и встречается в текстах всех жанров: в художественной литературе, нейтрально-деловой и научно-популярной прозе, поэзии, в разговорной речи. В научно-технической литературе и в официальных документах единица встречается реже.

Примеры из основной части Национального корпуса русского языка и корпуса СинТагРус

Если уже вы, то есть, не поможете в нашей просьбе, то уж не знаем, как и быть: просто хоть в петлю полезай. (Н. В. Гоголь).

*Проблема двуязычия легче разрешима, скажем, в Латвии или Армении, а **как быть** уйгуру или дунганину в Киргизии?*

*А **как же быть** с единой энергетической системой, связью или, скажем, добычей драгоценных металлов?*

Другие примеры из НКРЯ см. выше.

Примеры из поэтического корпуса НКРЯ

*Как нам быть с тем ужасом, который был бегом времени
когда-то наречён?* (А. Ахматова)

Как мне быть с мечтаньями твоими, Устремленными далеко ввысь?
(М. Светлов).

*Как же нам быть с этим знаком, Ставить его или нет? Надо, чтоб сами
глаголы Людям давали ответ* (Э. Успенский).

3. Микросинтаксическая разметка корпуса СинТагРус

Известно, что для решения разных компьютерно-лингвистических задач, например, для разрешения многозначности лексических единиц, полезно использовать лексически размеченные корпуса текстов, в которых многозначным словам приписывается определенное лексическое значение, заданное каким-либо лексикографическим ресурсом.

Однако корпуса текстов, хотя бы частично размеченные словесными значениями, очень полезны и при решении задач разрешения неоднозначности в рамках теоретической семантики и лексикографии без непосредственной привязки к проблематике компьютерной лингвистики (см., например, Б. Иомдин 2014). Между тем, если корпуса, размеченные значениями обычных слов, уже имеются в нашем распоряжении, то подобных ресурсов, в которых бы аннотировались какие-либо фразеологические единицы, насколько нам известно, до последнего времени не существовало.

В рамках исследования микросинтаксиса мы решили восполнить эту лауну и ввели в число размечаемых свойств корпуса СинТагРус указания о присутствующих в предложениях корпуса микросинтаксических единицах.

На снимке экрана (см. рис. 1) приводится фрагмент СинТагРуса, в котором встречается рассматриваемая в данной статье единица *как быть*. Для каждой фразы приводятся границы этой единицы и ее имя, а, заглянув в синтаксическую структуру предложения, можно увидеть, как единица встраивается в эту структуру. Эту информацию предполагается использовать как для разрешения многозначности лексических единиц, так и для задач глубокого семантического анализа.

Sentence	ID	MicroSynt	Class
Как тут быть?	[1]	{(как быть, {1.Как...3.быть})}	
Как же мне быть с моим отношением к Шостаковичу?	[2]	{(как быть, {1.Как...4.быть})}	
Но как быть с тем, что не имеет объяснения, образа, понятия?	[3]	{(как быть, {2.как...3.быть})}	
Речь шла не о прорывных идеях в области экономической теории, а главным образом о том, что делать...	[4]	{(как быть, {28.как...29.быть})}	
Однако как быть с геополитическими интересами России, с завоеваниями и трудами многих поколений на...	[5]	{(как быть, {2.как...3.быть})}	
Как быть, если кредит "душит"?	[6]	{(как быть, {1.Как...2.быть})}	
"Сами подумайте, как мне теперь быть? - спрашивает Кирилл Пескин, машинист II класса подмосковного ...	[7]	{(как быть, {3.как...6.быть})}	
Если с веб-разработчиками все более-менее понятно, то как быть рядовым интернетчикам?	[8]	{(как быть, {10.как...11.быть})}	
Разумеется, возникает вопрос: даже если информационные продукты можно копировать совершенно бес...	[9]	{(как быть, {14.как...16.быть})}	LF
И если работу генов можно изучать по крайней мере на биоматериале, то как быть с самим производств...	[10]	{(как быть, {11.как...12.быть})}	
И если да, то как быть с авторитарными режимами, известными в истории почти всех европейских наро...	[11]	{(как быть, {5.как...6.быть})}	
Как быть?	[12]	{(как быть, {1.Как...2.быть})}	
Ну а как же быть с Гитлером?	[13]	{(как быть, {3.как...5.быть})}	
Вот узнал человек, что арахис не орех, и как ему быть?	[14]	{(как быть, {9.как...11.быть})}	
- Проблема двуязычия легче разрешима, скажем, в Латвии или Армении, а как быть уйгуру или дунганцам...	[15]	{(как быть, {11.как...12.быть})}	LF
А как же быть с единой энергетической системой, связью или, скажем, добычей драгоценных металлов?	[16]	{(как быть, {2.как...4.быть})}	

Рис. 1. Микросинтаксическая разметка корпуса СинТагРус

Следует отметить, что в случае с микросинтаксической единицей *как быть* материал СинТагРус'а обнаруживает редкое единообразие. Во всех предложениях, в которых содержится синтаксическая конфигурация, состоящая из инфинитива *быть* и подчиненного ему местоименного наречия *как*, присутствует именно указанная единица, а примеров типа (23)–(25) не встречается вовсе. Такая ситуация нетипична для микросинтаксической разметки корпуса: в других случаях мы чаще имеем дело с многозначностью самой единицы или случайным соположением входящих в нее словесных элементов (т.н. false positives разметки). Данное обстоятельство можно считать косвенным свидетельством фразеологической силы рассмотренной единицы.

Литература

1. *Активный словарь русского языка* (2014). Т. I. Авторы: Апресян В. Ю., Апресян Ю. Д., Бабаева Е. Э., Богуславская О. Ю., Галактионова И. В., Гловинская М. Я., Иомдин Б. Л., Крылова Т. В., Левонтина И. Б., Лопухина А. А., Птенцова А. В., Санников А. В., Урысон Е. В. Под ред. акад. Ю. Д. Апресяна. М.: Языки славянской культуры. 408 с.
2. *Апресян Ю. Д.* (2010). Принципы лексикографического описания многозначных слов. // *Исследования по лингвистике и семиотике. Сборник статей к юбилею Вяч. Вс. Иванова*. М., Языки славянских культур, с. 217–233.
3. *Дяченко П. В., Иомдин Л. Л., Лазурский А. В., Митюшин Л. Г., Подлеская О. Ю., Сизов В. Г., Фролова Т. И., Цинман Л. Л.* (2015). Современное состояние глубоко аннотированного корпуса текстов русского языка (СинТагРус) // *Национальный корпус русского языка. 10 лет проекту. Труды Института русского языка им. В. В. Виноградова*. М. Вып. 6. С. 272–299.
4. *Иомдин Б. Л.* (2014). Многозначные слова в контексте и вне контекста. Вопросы языкознания, № 4. С. 87–103.
5. *Иомдин Л. Л.* (2013). ЧИТАТЬ НЕ ЧИТАЛ, НО...: об одной русской конструкции с повторяющимися словесными элементами // *Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог»* (2013). М.: Изд-во РГГУ. Вып. 12. Т. 1. С. 272–284.
6. *Иомдин Л. Л.* (2014). *Хорошо меня там не было*: синтаксис и семантика одного класса русских разговорных конструкций // Сб. статей «Grammaticalization and Lexicalization in the Slavic Languages». По материалам Международного симпозиума «Грамматикализация и лексикализация в славянских языках», 11–14 ноября 2011 г. München-Berlin-Washington/D.C.: Verlag Otto Sagner. Band 55. P. 423–436.
7. *Маракасова А. А., Иомдин Л. Л.* (2016). Микросинтаксическая разметка в корпусе русских текстов СинТагРус // *Информационные технологии и системы 2016 (ИТиС'2016)*. Сборник трудов 40-ой междисциплинарной школы-конференции ИППИ РАН. Репино, Санкт-Петербург. С. 445–449.
8. *Николаева Т. М.* (2008). Непарадигматическая лингвистика (История «блуждающих частиц»). М.: Языки славянских культур. — 689 с.
9. *Iomdin Leonid* (2016). Microsyntactic Phenomena as a Computational Linguistics Issue. // *Grammar and Lexicon: Interactions and Interfaces. Proceedings of the Workshop*. Osaka, Japan. 2016, pp. 8–18. (<http://aclweb.org/anthology/W/W16/W16-38.pdf>). ISBN 978-4-87974-706-8
10. *Iomdin L., Petrochenkov V., Sizov V., Tsinman L.* (2012). ETAP parser: state of the art // *Computational Linguistics and Intellectual Technologies. International Conference (Dialog'2012)*. Moscow: RGGU Publishers. Issue 11(18). P. 830–843. ISSN 2221–7932.

References

1. *Active dictionary of Russian. [Aktivnyj slovar' russkogo jazyka].* (2014). Vol I. Authors: Apresjan V. Yu., Apresjan Yu. D., Babaeva E. E., Boguslavskaya O. Yu., Galaktionova I. V., Glovinskaya M. Ya., Iomdin B. L., Krylova T. V., Levontina I. B., Lopukhina A. A., Ptentsova A. V., Sannikov A. V., Uryson E. V. Edited by Yu. D. Apresjan. Moscow: Jazyki slvjanskoj kultury []. 408 p. (In Russian).
2. *Apresjan Yu. D.* (2010). The principles of lexicographic descriptions of polysemous words [Principy leksikograficheskogo opisaniya mnogoznachnykh slov. // Issledovanija po lingvistike i semiotike Sbornik statej k jubileju Vjach. Vs. Ivanova. Moscow: Jazyki slavjanskix kultur, p. 217–233. (In Russian).
3. *Djachenko P. V., Iomdin L. L., Lazursky A. V., Mitjushin L. G., Podlesskaja O. Yu., Sizov V. G., Frolova T. I., Tsinman L. L.* (2015). The state-of-the art of a deeply annotated corpus of Russian texts (SynTagRus). [Sovremennoe sostojanie gluboko annotirovannogo korpusa tekstov russkogo jazyka (SynTagRus)] // Natsionalnyj korpus russkogo jazyka. 10 let proektu. Trydy Instituta russkogo jazyka im. V. V. Vinogradova. Moscow. Vyp. 6. p. 272–299. (In Russian).
4. *Iomdin B. L.* (2014). Polysemous words in context and out of context. [Mnogozhachnye slova v kontekste I вне konteksta. Voprosy jazykoznanija, N 4. p. 87–103. (In Russian).
5. *Iomdin L. L.* (2013). CHITAT' NE CHITAL, NO: On a Russian construction with repeated word elements. [CHITAT' NE CHITAL, NO: ob odnoj russkoj konstrukcii s povtorjajushchimisja slovesnymi elementami] // Kompjuternaja lingvistika I intellektualnye tekhnologii. Po materialam ezhegodnoj Mezhdunarodnoj konferencii "Dialog" (2013). Moscow: Izdatelstvo RGGU. vyp. 12. T. 1. p. 272–284. (In Russian).
6. *Iomdin L. L.* (2014). *Good thing I wasn't there: Syntax and Semantics of a class of Russian colloquial constructions. [Khorosho menja tam ne bylo: sintaksis i semantika odnogo klassa russkikh razgovornykh konstrukcij]* // In: Grammaticalization and Lexicalization in the Slavic Languages», November 11–14, 2011. München-Berlin-Washington/D.C.: Verlag Otto Sagner. Band 55. P. 423–436. (In Russian).
7. *Iomdin Leonid* (2016). Microsyntactic Phenomena as a Computational Linguistics Issue. // Grammar and Lexicon: Interactions and Interfaces. Proceedings of the Workshop. Osaka, Japan. 2016, pp. 8–18. (<http://aclweb.org/anthology/W/W16/W16-38.pdf>).
8. *Iomdin L., Petrochenkov V., Sizov V., Tsinman L.* (2012). ETAP parser: state of the art // Computational Linguistics and Intellectual Technologies. International Conference (Dialog'2012). Moscow: RGGU Publishers. Issue 11(18). P. 830–843.
9. *Marakasova A. A., Iomdin L. L.* (2016). Microsyntactic tagging in the SynTagRus corpus of Russian texts. [Mikrositaksicheskaja razmetka v korpuse russkikh tekstov SynTagRus] // Informacionnye tekhnologii i sistemy 2016 (ITiS'2016). Sbornik trudov 40-j mezhdistsiplinarnoj shkoly-konferencii IPPI RAN. Repino, Saint Petersburg. p. 445–449. (In Russian).
10. *Nikolaeva T. M.* (2008). (Nonparadigmatic linguistics. (The history of "straying particles") [Neparadigmaticeskaja lingvistika (Istorija "bluzhdajushikh chastic")]). Moscow: Jazyki slavjanskikh kultur. — 689 p. (In Russian).