

Крылов С. А. (krylov@rinet.ru)

**Композиционно-интегральный металингвистический
подход к семантической интерпретации текстов**

Krylov Sergej A. (krylov@rinet.ru)

**An attempt at compositional-integral meta-linguistic
approach to the semantic interpretation of texts.**

The main subject is an original approach to the problem of interpretation of the content of Russian discourse segments (DSs). It can be called compositional-integral metalinguistic approach. The content plane of SE is regarded at a specific grade of abstractness and at the same time on a specific level of «depth» of its interpretation, which can be called «conventional-semantic» (CS-) level.

The notion of «CS-notation» should be explained.

The principle of compositionality can be stated as follows: CS-representation (CSR) of a given DS is built out of the CSR of the particular meanings (PMs) of all linguistic units (LU), that appear in this DS: from CSR of lexical PMs of words that constitute its lexical content, CSR of grammatical PMs of word-forms, CSR of PMs of prosodic units used in this DS, CSR of PMs of derivational means used in this DS, CSR of syntactic relations between word-forms belonging to the given DS, CSR of PMs of syntactic patterns (predicative as well as non-predicative ones) and CSR of PMs of phraseological expressions used in this DS.

The CSR of a given LU differs from its morphological notation in one important feature: the morphological representation (MR) of two DSs, where different PMs of the two DSs are actualized, should coincide (by definition), but their CSRs should differ (by definition). So, if a LU can be used in two different PMs, then each one of its PMs gets its own CSR which differs from the CSRs of other PMs of this LU. Due to this property CSR is similar to such ways of meaning representation as «deep syntactic representation», «surface semantic representation», «deep structure language» and so on.

But the notion of CSR differs from all the above mentioned ways of content representation of DS in one very important feature – namely, in the semiotic type of the meta-language (ML) we use. One can propose to use as a semantic ML not a set of some notional signs that express some ideas, but a set of non-elementary individual names (more precisely, descriptions), or, in other terminology, «logical proper names», that denote strictly defined (individual) entities. These entities belong to a very special domain – namely, the domain of metalinguistic facts (DMLF). The DMLF, generally speaking, is giant and has no clear borders. That is why the only way of getting explicitness during the work with elements of DMLF is the following: we select some explicitly given subset from the DMLF, elements of which (according to my intuitive sense) belong to the «central» or «nuclear» zone inside the DMLF. One can arbitrarily call this subset the «canonical» DMLF (= CDMLF) and try to work with it on the base of the a priori adopted hypothesis about its validity for the use as

a tool for metalinguistic interpretation of DSs.

Предметом внимания является специфический подход к проблеме интерпретации плана содержания (ПС) речевого отрезка (РО) на русском языке (РЯ), который можно назвать композиционно-интегральным металингвистическим подходом (КИМП). При этом ПС РО рассматривается на некоторой определённой ступени абстракции и одновременно на некотором определённом уровне «глубины» интерпретации – так называемом «условно-семантическом» (УС-) уровне представления.

Понятие УС-записи (= УС-транскрипции = УС-представления) РО следует пояснить.

Принцип композиционности состоит в том, что условно-семантическое представление (УСП) РО включает в себя УСП частных значений (ЧЗ) языковых единиц, входящих в этот РО: из УСП лексических ЧЗ слов, образующих его лексический состав, УСП грамматических ЧЗ словоформ, УСП грамматических ЧЗ просодических единиц, реализуемых на сочетаниях словоформ, УСП ЧЗ деривационных средств, употреблённых в данном РО, УСП ЧЗ синтаксических связей слов, входящих в данный РО, УСП ЧЗ синтаксических конструкций (как предикативного, так и непредикативного типа) и УСП ЧЗ фразеологических единиц, употреблённых в данном РО.

Сначала объясним, почему УСП вообще трактуется именно как «семантическое» (а не «морфологическое») представление. УСП отдельной языковой единицы отличается от её морфологической записи (морфологического представления, МП) тем, что морфологические представления двух РО, в которых реализуются разные ЧЗ этой языковой единицы, должны по определению совпадать, тогда как их УСП должны по определению различаться. Таким образом, если некоторая языковая единица способна употребляться в нескольких разных ЧЗ, то каждое из них получит своё УСП, отличающееся от УСП других ЧЗ. Очерченное свойство сближает УСП с такими способами записи ПС РО, как «глубинно-синтаксическое представление», «поверхностно-семантическое представление», «язык глубинных структур» и т. п.

Однако понятие УСП существенно отличается от вышеуказанных способов фиксации ПС РО по одному весьма существенному параметру – а именно, по семиотическому типу используемого метаязыка. В качестве семантического метаязыка предлагается использовать не набор тех или иных нарицательных имён, выражающих те или иные понятия, а набор неэлементарных индивидуальных («сингулярных») имён (дескрипций) (или, как иногда говорят, логических собственных имён), обозначающих строго определённые (индивидуальные) объекты, принадлежащие весьма специфической предметной области – а именно, предметной области металингвистических фактов (ПОМФ)¹. В этом и состоит специфика УСП по сравнению с другими разновидностями «семантических представлений».

Сама ПОМФ, вообще говоря, представляет собой весьма обширное множество и к тому же имеет не вполне чёткие границы. Поэтому единственный способ добиться эксплицитности при оперировании с элементами ПОМФ состоит в том, чтобы выделить из неё некоторое эксплицитно заданное подмножество, элементы которого будут, в соответствии с моим интуитивным ощущением,

¹ Ср. подход, при котором ПОМФ не задаётся эксплицитно.

относиться к ядерной (центральной) зоне в её составе. Условно назовём такое множество «каноническим ядром» (КЯ) и попытаемся работать с ним на основе исходно принимаемой рабочей гипотезы о его пригодности для цели металингвистической интерпретации РО.

При КИМП элементами ПОМФ являются металингвистические факты (МФ) – то есть высказывания лингвистов о ЧЗ тех или иных языковых единиц (ЯЕ)². В частности, при построении КЯ ПОМФ целесообразно включить в его состав по крайней мере следующие типы высказываний о ЧЗ:

1. Высказывания о ЧЗ лексических единиц РЯ, содержащиеся в основных доступных нам толковых словарях (В. И. Даля, Д. Н. Ушакова, С. И. Ожегова, МАС, БАС, БТС).

2. Высказывания о ЧЗ граммем РЯ, содержащиеся в основных доступных нам «академических» грамматиках (1952, 1970, 1980, «Пражская» грамматика 1979), а также в некоторых обобщающих работах той же степени всеохватности (напр., А. В. Исаченко).

3. Высказывания о ЧЗ синтаксических конструкций (как предикативных, называемых «структурными схемами предложений», так и непредикативных, называемых «типами словосочетаний», «типами синтаксических связей», «типами оборотов» и т. п.) РЯ, содержащиеся в синтаксических разделах таких грамматик, а также в некоторых обобщающих работах (напр., А. М. Пешковского, Г. А. Золотовой, М. В. Всеволодовой).

4. Высказывания о ЧЗ интонационных конструкций, содержащиеся в разделе «Фонетика» таких грамматик. Ср. подход, при котором интонационные средства языка исключаются из числа ЯЕ (таковы, в частности, все подходы, при которых текст понимается как буквенная цепочка). Если анализируемый текст представлен в виде буквенной цепочки, допускающей (в данном - фиксированном - контексте!) различные интонационные оформления, должен трактоваться как пучок **нескольких** омографических РО и, соответственно, получать столько интерпретаций, сколько возможностей интонирования он допускает.

5. Полуэксплицитные высказывания о ЧЗ знаков препинания русского языка, содержащиеся в основных справочниках по русской пунктуации (А. Б. Шапиро, Д. Э. Розенталь, Н. С. Валгина) в форме высказываний о правилах употребления этих знаков. Ср. подход, при котором пунктуационные средства языка исключаются из числа ЯЕ (таковы некоторые формальные модели морфологии и синтаксиса).

6. Высказывания о ЧЗ словообразовательных средств РЯ, содержащиеся в составе раздела «Словообразование» в указанных грамматиках. Ср. подходы, (а) при котором производные слова, не входящие в словарь, предлагается исключать из рассмотрения при разборе текста, и (б) при котором из предмета семантического анализа исключаются словообразовательные средства.

7. Высказывания о ЧЗ графических средств РЯ, содержащиеся в основных справочниках по русской орфографии (А. Б. Шапиро, Д. Э. Розенталь, В. Ф. Иванова) в форме высказываний о правилах употребления этих графических средств (прописных букв, графического удвоения *н*, слитности и раздельности написания маркера отрицания *не* и т. п.). Ср. подходы, при которых у графических средств

² Ср. подход, при котором содержание описаний (словарей, грамматик, научных монографий), рассчитанное на одушевлённого адресата, игнорируется на основании ссылки на усматриваемую в таких описаниях «неформализованность», якобы присущую таким описаниям.

не усматривается никакой семантики (то есть что предметом орфографических описаний является лишь орфографическая правильность, то есть что графические средства исключаются из числа двусторонних ЯЕ).

8. Полуэксплицитные высказывания о наборе собственных имён РЯ (а также об их смысловых оттенках), содержащиеся в «Словаре русских личных имён» Н. А. Петровского и в «Русских фамилиях» Б. О. Унбегауна. Ср. подходы, (а) при которых собственные имена исключаются из числа ЯЕ; (б) при которых собственные имена исключаются из предмета семантического анализа и (в) при которых собственные имена трактуются как «не обладающие смыслом» (или: «...значением»).

9. Высказывания о значениях аббревиатур, содержащиеся, например, в «Словаре сокращений русского языка» Д. И. Алексеева. Ср. подход, при котором аббревиатуры исключаются из числа слов РЯ.

10. Полуэксплицитные высказывания о референтах индивидуальных имён (а также об их смысловых коннотациях), содержащиеся, например, в «Большом энциклопедическом словаре» последнего издания. Ср. подход, при котором референция собственных имён исключается из предмета семантического анализа.

11. Высказывания о семантических подклассах ЯЕ (например, морфем, слов, синтаксических конструкций, интонем и т. п.), содержащиеся в ряде работ обобщающего характера, например, в книгах «Русский язык. Энциклопедия» (изд. 2, 1997) или «Handbuch der Russisten» (Wiesbaden, 2. Aufl.).

12. Высказывания о типах семантических сдвигов (тропов, фигур, приёмов выразительности), содержащиеся в ряде работ обобщающего характера (в частности, в энциклопедии «Русский язык» или в книге Д. Э. Розенталя «Практическая стилистика русского языка»). КИМП противостоит подходам, при которых КЯ ПОМФ не задаётся эксплицитно.

КИМП основан на следующих общих логических принципах.

Принцип естественно-языковой интегральности: при построении УСП следует учитывать все типы ЯЕ, входящих в интерпретируемый РО: не только грамматику, но и лексику; не только лексику, но и грамматику; не только сегментные, но и несегментные ЯЕ.

Принцип металингвистической замкнутости: для интерпретации используются только отсылки к объектам, входящим в КЯ ПОМФ (т. е. недопустимы отсылки к объектам, не входящим в КЯ ПОМФ). Любая трудность, связанная со стремлением последовательно соблюдать этот принципа, должна становиться предметом эксплицитного рассмотрения и эксплицитного вынесения решений. В частности, вопрос о расширении самого КЯ ПОМФ должен решаться столь же эксплицитно, как и вопросы о форме УСП любого РО.

Принцип металингвистической интегральности: при построении УСП следует описывать каждый обрабатываемый РО с помощью максимального набора истинных высказываний, которые можно о нём (и о его составляющих частях) сделать, не выходя за пределы выделенного канонического ядра ПОМФ. Из этого принципа следует, в частности, такое следствие: каждый шаг, связанный с расширением КЯ ПОМФ, должен сопровождаться пересмотром (естественно, в сторону обогащения) того множества истинных суждений, которые можно сделать данным речевом отрезке (а также обо всех его составных частях) на основе уже обновлённого (расширенного) канонического набора металингвистических фактов.

В рамках КИМП можно построить формальные аналоги некоторых широко обсуждаемых лингвистических понятий.

1) Формальным аналогом «семантического инварианта» (или так называемого «общего значения») является семантическое описание, хранящееся по адресу, способному быть предметом отсылки, такое, что предметом этого описания является одноэлементное или многоэлементное множества значений некоторой единой языковой единицы в том источнике, на который делается ссылка (единой – значит такой, которая трактуется в данном источнике не как пара омонимов, а как два разных значения одной и той же единицы). В случае многоэлементности указанного множества формальным аналогом семантического инварианта является дизъюнкция значений, приписываемых данной единице в данном источнике. Заметим, что наличие семантического инварианта у двух одноимённых РО в рамках КИМП доказывается достаточно просто – а именно, ссылкой на такой источник, где соответствующие изучаемым РО единицы трактуются не как омонимы, а как два разных значения одной многозначной единицы. Свидетельством же отсутствия семантического инварианта является трактовка соответствующих единиц как омонимов во всех источниках, где они описаны. Если же члены соответствующей пары описываются в двух источниках по-разному (в источнике А как омонимическая пара, а в источнике Б – как пара значений многозначной единицы), то этого факта достаточно для помещения изучаемой пары в «серую зону» переходных явлений между омонимией и многозначностью.

2) Формальным аналогом «ЧЗ» является описание отдельного ЧЗ данной ЯЕ в каком-либо источнике, хранящееся по адресу, способному быть предметом отсылки. Поэтому доказательством существования некоторого ЧЗ при металингвистическом подходе является возможность отсылки к такому источнику, где данное ЧЗ усматривается.

3) Формальным аналогом «речевого», «контекстного», «оказионального», «потенциального» значения РО (то есть такого значения, которое не свойственно компонентам данного РО как элементам языкового инвентаря, но которое данный РО способен выразить в благоприятном для этого контексте) является сочетание семантического описания одного из ЧЗ данной ЯЕ (хранящегося по адресу, способному быть предметом отсылки) в одном из источников с семантическим сдвигом определённого типа, применением которого к сочетанию компонентов данного РО можно данное потенциальное значение получить, таким, что описание семантических сдвигов этого типа хранится по адресу, способному быть предметом отсылки. В таком случае потенциальные значения будут подразделены на два класса: вычисляемые на основе канонического ядра ПОМФ и невычисляемые на этой основе. Внимательное изучение второго подкласса может стать основанием для включения в каноническую сферу новых источников, содержащих описания доселе неописанных семантических сдвигов.

Результатом применения КИМ-интерпретации к некоторому РО является его «семантическая разметка». Если описываемый РО по своей протяжённости равен целому тексту, то результатом его УС-интерпретации является «семантически размеченный» текст.

Ещё несколько слов о применении КИМП в типологии. При построении многоязычной текстовой базы данных, описывающей множество выровненных параллельных текстов, КИМП позволяет дать такое «внутриязыковое» описание данного текста, которое, с одной стороны, ориентировано на учёт содержательной стороны этого текста, а с другой стороны, может быть

эксплицитно сопоставлено с аналогичным независимо построенным УСП некоторого текста на другом языке, состоящего с исследуемым текстом в отношениях переводческой эквивалентности. Отсюда следует принципиальная возможность семантико-типологического сопоставительного изучения произвольного корпуса выровненных параллельных текстов - переводов одного и того же оригинала. УС-интерпретация оказывается полезной частью и желательной предпосылкой сопоставления текстов на разных языках, так как сопоставление планов выражения текстов, являющихся переводами друг друга, едва ли способно принести хоть сколько-нибудь нетривиальные результаты, а сопоставление планов содержания текстов возможно лишь при условии наличия эксплицитных описаний этих ПС. Между тем сама разработка средств такого описания должна опираться на некоторые принципиальные положения металингвистики. Посильным вкладом в выработку соответствующей теории и предлагается считать КИМП.

Отличие УСП от традиционного (в школьной и вузовской практике изучения русского языка) разбора «по членам предложения» и «по частям речи» состоит в следующих его особенностях:

1) в УСП присутствует не только грамматический, но и лексический разбор, предполагающий посильную «моносемизацию» слов в соответствии с их словарными значениями;

2) в УСП грамматические значения представлены не в виде указания на граммы («родительный падеж», «несовершенный вид»), а в виде указания на ЧЗ граммем («родительный падеж деятеля», «несовершенный вид в актуально-длительном значении» и т. п.);

3) при КИМП от анализирующего субъекта не требуется предварительного запоминания того теоретического аппарата, который используется при разборе РО;

4) при КИМП анализирующий субъект не обязан придерживаться никакой заранее заданной частнолингвистической теории в качестве «заведомо истинной»; все теории, входящие в каноническое ядро ПОМФ, признаются равноправными. Таким образом снимается сама дилемма «какую из теорий принимать за основу?». Вместо этой, «наивно-онтологической» трактовки лингвистических теорий предлагается гносеологический (эпистемологический) подход к ним.

Сплошная УС-интерпретация основана на постоянном участии человека. Никаким автоматическим устройством она выполняться не может. Понятно, что семантически размеченный корпус текстов разумно трактовать не как раз и навсегда завершённый продукт, а как постоянно пополняемую базу данных. Погоня за количественной протяжённостью такого корпуса едва ли существенно приблизит нас к формальной репрезентации ПС РО. Здесь уместно исходить из принципа «лучше меньше, да лучше». Однако уместно стремиться к тому, чтобы «полнота» охвата ПС была бы максимальной. Здесь, правда, возникает вопрос о разумной мере такой полноты. Принцип ограниченности КЯ ПОМФ плюс принцип металингвистической замкнутости позволяет дать на него наиболее естественный ответ.

Постепенное возрастание объёма семантически сплошь размеченного корпуса русских текстов рано или поздно приведёт к тому, что на основе такого корпуса станет возможным решение задачи создания частотных словарей, входами которых были бы не только «лексемы»³ = «вокабулы»⁴ (то есть

³ В смысле В. В. Виноградова, А. И. Смирницкого, Ю. С. Маслова, А. А. Зализняка и др.

единицы, вообще говоря, многозначные) (как это принято по сей день в практике частотной лексикографии), а «слово-значения»⁵ = «лексико-семантические варианты»⁶ = «семантические аллолексемы»⁷.

То же соображение касается и грамматических значений: по достижении некоторого разумного объёма семантически размеченный корпус текстов может послужить основой для создания такого частотного «грамматикона» («грамматикария»?), входом в который будут не только граммы или их сочетания (грамматические формы), как это было принято доселе в практике частотной лексикографии (точнее, «грамматикографии»), но и отдельные ЧЗ граммем.

Создание и перманентное пополнение лексико-грамматической базы данных на основе КИМП было бы весьма полезным экспериментом не только для дальнейшего использования созданного корпуса в качестве источника сведений об употребительности ЧЗ ЯЕ или источника примеров при семантически сформулированном поисковом запросе, но и для решения некоторых других задач.

Среди таких задач можно выделить следующие:

(1) **Дидактическая функция.** Упражнения по КИМ-разбору РО полезны тем, что формируют у учащегося одновременно несколько умений:

(а) языковую рефлексию сознательного носителя, то есть умение осознать то, в каком из своих значений употреблена в данном контексте некоторая многозначная ЯЕ, и выделить именно это значение;

(б) профессиональную компетенцию сознательного читателя лингвистических описаний, то есть умение найти в тексте описания (словарного или грамматического) информацию об интересующей нас в данный момент ЯЕ и об интересующем нас в данный момент ЧЗ из числа тех, что выделены автором описания;

(в) металингвистическую компетенцию, то есть умение соотнести первое («контекстное», «речевое» значение) со вторым («инвентарным» значением), в частности, умение при несовпадении первого со вторым ощутить контекстный семантический сдвиг (речевой «троп») и сформулировать, в чём он состоит.

В силу принципа естественно-языковой интегральности учащийся, выполняющий упражнение по УС-интерпретации в рамках КИМП, приобретает полезный навык: не упускать из виду ни один важный аспект ПС РО. Т. о., КИМП противостоит подходу, при котором усиленное внимание к одному аспекту отрицательно сказывается на общем результате анализа, так как сопровождается ослаблением внимания к другим, не менее важным аспектам.

(2) **Мелиоративная (оптимизирующая) металингвистическая функция.** Каждый случай несовпадения контекстного значения с набором зафиксированных в описании (в словарях и грамматиках) инвентарных значений может объясняться не только контекстным семантическим сдвигом,

⁴ В смысле И. А. Мельчука.

⁵ В смысле Л. В. Щербы.

⁶ В смысле А. И. Смирницкого.

⁷ В смысле Ю. С. Маслова.

но и другими факторами - в частности, неполнотой имеющихся описаний. В таком случае КИМП может служить эффективным экспериментальным методом, с помощью которого проверяется относительная полнота лингвистических описаний (как словарных, так и грамматических) и выясняется то, каких ЧЗ в данном описании «не хватает». Очевидна польза такого эксперимента для работы по дальнейшему усовершенствованию словарных и грамматических описаний языка.

(3) **Компаративная металингвистическая функция.** Для описаний, входящих в КЯ ПОМФ, КИМП позволяет дать их сравнительную оценку по степени полноты. Для многих описаний окажется возможным ответить на вопросы: какое из этих описаний покрывает более значительную долю текста, какое - менее значительную; какие типы ЧЗ более полно описываются одной теорией, а какие - другой теорией. И абсолютная (по доле покрываемости текста данной теорией), и относительная (по соотношению долей покрываемости текста двумя разными теориями) может способствовать более выработке более объективных представлений о достоинствах и недостатках разных лингвистических теорий.

В одних случаях эффектом такого сравнения будет улучшение сложившейся ранее репутации некоторой теории, в другом - напротив, ухудшение такой репутации. И то, и другое приблизит нас к объективной оценке этих теорий.