

# Пространственные отношения в семантическом языке SL и его прототипе универсальном языке Диал

Куликов В. В., Елкин С. В., Мансурова О. Ю.

## Введение

Для задач теории искусственного интеллекта разработан широкий спектр языков представления знаний. Как правило, все эти языки в математическом отношении основываются на том или ином комплекте теоретико-множественных операций и служат для представления знаний в одной или нескольких предметных областях. И хотя теоретико-множественный подход обладает значительной степенью общности, однако, он не предназначен для решения такой, совсем недавно поставленной задачи, как трансляция знаний из одной предметной области в другую. Данная работа выросла из разработок по принципам и основам грамматики универсального языка Диал [1]. Этот язык изначально создавался для наиболее абстрактного описания процессов, структура которых остается инвариантной в любых областях. Сразу же после появления Диала была поставлена задача его математической формализации, которая была успешно решена в 1992 г. Первое сообщение о языке Диал было сделано на Циолковских чтениях в Калуге [1], а первая публикация вышла в 1994 г. [2]. В дальнейшем математическая версия языка развивалась для задачи машинного перевода в качестве семантического языка-посредника SL [3–5]. На основе Диала был разработан семантический словарь и анализатор для машинного переводчика Кросслейтор 1.0 [6], системы семантического поиска и реферирования документов [7, 8]. Благодаря поддержке Российского фонда фундаментальных исследований<sup>1</sup>, работы над SL приобрели новый импульс. Авторы пришли к выводу, что высокая степень абстрактности и универсальности языка SL (Диал) может позволить осуществить трансляцию знаний из одной предметной области в другую [9]. Идея трансляции знаний открывает новое направление в теории систем, ориентированных на знания.

В этой работе мы изложим часть грамматики языка SL, касающуюся пространственных отношений (ПрО).

Язык образует свое пространство–время, точнее - множество пространств–времен, фрактально «вложенных» друг в друга, словно матрешки. Язык всегда

---

<sup>1</sup> Работа поддержана грантом РФФИ № 06-01-00538.

отображает мир нашей Вселенной таким, каким он видится нам сегодня, и SL (Диал) представляет собой лишь более развитую, но все же модель нашего мира. В этой модели существительные отображают относительно устойчивые события, которые принято называть объектами или субъектами, а глаголы — менее устойчивые, текучие процессы нашего мира. Прилагательные и наречия отображают свойства тех и других, т. е. “силы” и “поля”, которые эти объекты-процессы “распространяют” вокруг себя. И те, и другие, и третьи суть одно и то же, в разных условиях проявляющее себя по-разному.

Для любых систем, описывающих пространственные отношения в задачах машинной лингвистики, необходимо однозначное описание и точная передача сложной пространственной ситуации, выражаемой в естественных языках предложениями, наречиями, словосочетаниями и другими лексическими средствами. В данной работе мы развиваем средства языка SL, позволяющие успешно решать эту задачу, как для машинного перевода, так и для трансляции знаний.

### **Пространственные отношения. Локальная ситуация и ее участники**

Пространственные отношения определяются разными исследователями по-разному. М. В. Всеволодова и О. Ю. Дементьева в самом общем виде определяют ПрО как «соположение в пространстве какого-либо предмета, действия (события), признака и некоторого пространственного ориентира — локума» [10]. Иштван Пете пишет, что ПрО выражают локализацию предмета, действия или признака в определенном фрагменте действительности, то есть в положении вещей [11]. В. Г. Гак определяет пространственное отношение в чистом виде формулой

$$S(O) + V1 + r + L,$$

где **S(O)** — локализуемый объект, **V1** — глагол, в значение которого может входить пространственная сема (это, прежде всего, глаголы местоприбывания и движения), **r** — конкретный тип локализации, выражаемый, например, предлогом, а **L** — локализатор, то есть двухмерное или трехмерное пространство, в котором способен помещаться локализуемый объект [12].

Рассмотрим теперь, что представляет собой локальная ситуация.

Центральной частью локальной ситуации является процесс. В тексте он обычно представлен глаголом или его формами: причастием, деепричастием или

отглагольным существительным. Иногда действие может быть выражено в предложении неявно: *Цветы в вазе были очень красивыми. Мое внимание привлекла огромная картина на стене.* На синтаксическом уровне никакой глагольной формы в таких предложениях нет, но семантически всегда можно восстановить предикативное слово: «находившиеся», «висящая».

Далее необходимо различать двух основных участников локальной ситуации: локализуемый объект (в дальнейшем Ob) — предмет, перемещение или местонахождение которого включает описываемая пространственная ситуация, и ориентир — предмет, относительно которого определяются пространственные координаты Ob или ситуация в целом: то место, где находится Ob, или куда/ откуда/ по которому и т. д. он перемещается.

В естественных языках локализация чаще всего приблизительна и размыта, поэтому использование понятия ориентир вполне оправдано, в то время как в искусственном языке, задачей которого является трансляция знаний, лучше использовать более точное понятие. Так в физике вместо ориентира используется термин система отсчёта (СО) — тело отсчета или совокупность тел, неподвижных относительно друг друга, по отношению к которым рассматривается движение и совокупность часов, отсчитывающих время. Для определения положения тела к СО добавляют систему координат — набор линеек. Так, например, в предложении *Книга упала на пол* роль Ob играет участник, обозначенный именем *книга*, а роль тела отсчета СО играет участник, названный существительным *пол*. Система отсчета при этом может указывать:

- на *местонахождение* Ob — место, в котором локализован объект или пространственная ситуация перемещения в целом (*книга лежит на столе; мальчик гулял в парке/по парку*);
- на *исходный пункт* движения/перемещения — точку, в которой объект находился до начала движения (*мальчик вышел из комнаты, книга упала со стола*);
- на *конечный пункт* движения/перемещения — точку, в которой объект оказался после окончания перемещения (*мальчик вошел в комнату, книга упала на пол*);
- на *траекторию* движения/перемещения — координаты, которые последовательно занимал объект, перемещаясь от исходного пункта к конечному (*мальчик шел по лесу*).

В языке SL местоположение принято обозначать оператором  $\hat{x}\hat{t}$ , исходный пункт —  $(\hat{x}\hat{t}) = x^+$ , конечный пункт —  $(\hat{t}\hat{x}) = x^-$ , перемещение (скачек) —  $\hat{x}\hat{t}\hat{x}\hat{t}\hat{x} = \hat{x}^+\hat{x}\hat{x}^-$ , траекторию —  $\hat{x}(t)$ , (здесь  $\hat{x}$  — оператор пространства,  $\hat{t}$  — оператор времени).

Наконец необходимо определить тот конкретный участок пространства, который непосредственно принимает участие в ситуации перемещения или местонахождения. Относительно ориентира (тела отсчета) пространство членится на ряд топологических зон: *поверхность ориентира, внутренняя область ориентира, окрестность (пространство около) ориентира, пространство перед ориентиром* и т. д.

Семантически ПрО связывают Об и СО, т. е., как об этом пишет М. В. Всеволодова, они обозначают «соположение в пространстве какого-либо предмета, действия (события), признака и некоторого пространственного ориентира». На синтаксическом уровне ПрО устанавливаются между глаголом (или причастием, деепричастием, отглагольным существительным или обычным существительным) и именной группой слова — СО: *положить* → *на стол*, *взгляд* → *из-под ресниц*, *полет* → *в космос*.

Классический вариант локальной конструкции — это предикат, от которого, с одной стороны, зависит подлежащее, а с другой, предложно/послеложная группа:

**идти** → мальчик  
 ↳ **в** → школа

Вместо глагола могут стоять причастие, деепричастие, отглагольное существительное или придаточное предложение:

2) мальчик → **идуший** → **в** → школу  
 мальчик → **кладущий** → книгу  
 ↳ **на** → стол  
 мальчик → который → **идет** → **в** → школу  
 мальчик → который → **кладет** → книгу  
 ↳ **на** → стол  
 мальчик → **летя** → **на** → самолете  
 мальчик → **кладя** → книгу  
 ↳ **на** → стол

погрузка → товаров  
└ → на → корабль  
приезд → в → Стамбул

При неявно выраженном действии именной группой слова (СО) управляет существительное. Например, *Цветы в вазе, стоящей на столе посреди комнаты, были просто изумительны: цветы → в вазе...*

### Основы классификации ПрО в языке SL

О. Ю. Мансурова в работе [13] выделила те элементы существующих классификаций, которые позволили создать список ПрО, оптимальный для целей машинного перевода. Грамматика, построенная на базе такой классификации, с одной стороны, охватывает все выделенные и описанные в литературе ПрО (хотя бы для некоторого набора языков), а с другой стороны, достаточно стройна и максимально обобщена, чтобы ее легко было обрабатывать алгоритмически и подвергнуть универсализации. В её работе используется в основном терминология И. А. Мельчука и В. А. Плунгяна: *взаиморасположение предметов* отображает параметр локализации, а *характер ситуации* (статичная или динамичная, если динамичная, то какой именно тип движения присутствует) — параметр ориентации. Метаязык описания ПрО максимально прост, логичен и хорошо запоминается, т.к. для обозначения категорий выбраны предлоги русского языка как максимально мнемоничные для русскоязычных читателей.

Формально два ключевых параметра классификации можно определить следующим образом:

— **параметр локализации** указывает на область пространства относительно СО, в которой находится (в которую или из которой перемещается) участник ситуации;

— **параметр ориентации** характеризует отношение между описываемым объектом и СО в терминах приближения, удаления или покоя (отсутствия движения).

Возьмем, например, словосочетание «*положить книгу на стол*»: локализация — «на чем-л.», а ориентация — перемещение «*на стол*». Ещё пример: «*пройти мимо здания*» — локализация «перед или около» (здания), ориентация — «трасса».

Метаязык в работе О. Ю. Мансуровой для описания семантики и инвентаризации средств выражения пространства представляет собой комбинацию парамет-

ров «локализация + ориентация», то есть для описания ПрО используется следующее обозначение: ЛОК\_‘*локализация*’, ‘*ориентация*’, при этом параметры «локализация» и «ориентация» принимают конкретные значения, например, ЛОК\_около, старт, ЛОК\_перед, трасса.

В целом соглашаясь с необходимостью введения двух параметров для классификации ПрО, мы считаем термин *ориентации* не достаточно точным. Так Русский энциклопедический словарь дает следующее определение: «**ориентация** — обобщение понятия направления на прямой на геометрические фигуры более общей структуры». В данном же случае в работе О. Ю. Мансуровой этот параметр имеет следующие значения.

Таблица № 1. Значения категории ориентация

<i>Ориентации</i>	<i>Значение ориентации</i>	<i>Примеры</i>
<b>стат.</b>	нахождение л-объекта в покое по отношению к ориентиру	<i>жить в городе; лежать на столе; работать в школе</i>
<b>ненапр.</b>	перемещение объекта по области ориентира или расположение л-объекта в этой области	<i>бродить по городу; ездить по России</i>
<b>старт</b>	перемещение л-объекта из области ориентира	<i>уехать из города; взять со стола</i>
<b>финиш</b>	перемещение л-объекта к области ориентира	<i>приехать в город; положить на стол</i>
<b>трасса</b>	перемещение л-объекта через область ориентира (путь, траектория, трасса)	<i>идти по улице; пройти через мост; пронести через проходную</i>

В место него, как нам кажется, лучше ввести термин *динамика* — Д(*i*). Для нашей задачи в качестве значений *i* параметра динамика будем использовать следующие:

- o** — покой в данной СО,
- a** — движение в данной СО,

**у** — приближение в данной СО,

**и** — удаление в данной СО,

**ы** — неустойчивое движение в данной СО (например: «Merrill Lynch считает, что неустойчивые колебательные движения доллар/йены сохранятся до середины сентября», «срыв потока описывается уравнением с сингулярностью»),

**э** — ненаправленное (случайное, хаотическое) движение в данной СО (например, «блуждать по лесу», «бесцельно бродить» и т. д.),

**а<sup>+</sup>** — начало движения (старт),

**а<sup>-</sup>** — завершение движения.

Значения выбирались в соответствие с общей логикой SL (Диала). Комбинации значений могут давать новые значения, например, **оа** — ускорение, **ао** — торможение, **и<sup>+</sup>** — начало отталкивания, **э<sup>-</sup>** — завершение случайного движения, **иу** — колебательное движение, и т. д. Список значений параметра принципиально не фиксирован и может быть в любой момент расширен по принципам языка SL.

### **Параметр «топологическая зона», его значения**

Пространство СО делится на ряд топологических зон. Обычно в качестве исходной точки для теоретических построений берется список логически возможных топологических зон тела отсчета (ориентира): «в X», «на X», «под X», «перед X», «сзади X», «на боковой стороне X», «у/около X», «вокруг X», «между X / среди X» [14]. Языковое членение пространства, однако, далеко не всегда совпадает с логическим. Например, для обозначения одной и той же топологической зоны в языке может существовать более одного показателя. Поэтому обычно различаются понятия *топологической зоны ориентира* и *локализации*. Понятие *топологической зоны* применяется к логически мыслимым фрагментам пространства в СО, тогда как понятие *локализации* указывает на членение пространства относительно СО «по данным языка». Тем самым инвентарь возможных локализаций представляет собой максимально дробное разбиение пространства по данным представительной выборки языков. Другими словами, элементы «категории локализации указывают на определенную область пространства по отношению к некоторому ориентиру» [15], то есть эта категория описывает собственно местонахождение объекта или взаиморасположение объекта и СО (ориентира).

Для языка трансляции знаний более приемлемо использовать понятие топологической зоны  $To3(j)$ . Значения параметра  $To3$  так же выберем в соответствии с категориальной системой  $SL$  (диал).

Исходными значениями  $j$  будут категории оппозиционной пары «определённый»/«неопределённый» ( $x - \bar{x}$ ). Например, «попугай сидит на жердочке» — определённая  $To3$ , «все ушли на фронт» — неопределённая  $To3$ .

Переход от неопределённой  $To3$  к определённой и обратно составляет вторую оппозиционную пару ( $x\bar{x} = x^- \Leftrightarrow \bar{x}x = x^+$ ). Например, «он вышел из дома» —  $To3(x^-)$ , «он вошел в дом» —  $To3(x^+)$ .

Следующий шаг синтеза значений параметров  $To3$  должен состоять из объединения операторов — значений  $To3$   $x\bar{x} = x^-$  и  $\bar{x}x = x^+$  — «уменьшение степени определённости  $To3$ » и «увеличение степени определённости  $To3$ ». Таким образом, получается, что  $x\bar{x}x$  и  $\bar{x}x\bar{x}$  — это «перенос (скачек) определённости» и «перенос неопределённости». Здесь так же возможны операторы перехода от уменьшения определённости  $To3$  к увеличению —  $x^+x^-$ . Например, «целый день Варвара ходила из дома в дом» —  $To3(x\bar{x}x)$ , «он вошёл в её положение и вышел из него и оставил её в её положении» (М. Жванецкий) —  $To3(x^+x^-)$ .

Операторы количества ( $K^+$ ,  $K$ ,  $K^-$ ) в  $SL$  позволяют ввести различные качественные градации (дискретный набор) определённости и неопределённости  $To3$ : *далеко, близко, рядом, непосредственно, поблизости, невдалеке, около, вне, за пределами* и т.д.. Так же как, например, можно ввести градации холодного и горячего.

Операторы фаз и их превращений позволяют различить состояния  $To3$ : *сверху, снизу, слева, справа, за, перед, под, над* и т.д.

Так как в задаче трансляции знаний могут возникнуть совершенно неожиданные пространственные отношения, то список значений параметра  $To3$  принципиально не фиксирован и может быть в любой момент расширен по принципам  $SL$ .

Таким образом, описание  $PrO$  с использованием вышеизложенного должно выглядеть так:  **$PrO(D(i), To3(j))$** . В качестве самого  $PrO$  выбираем предлог  $SL$  (Диала) из списка предлогов.

Предлоги  $SL$  (Диала), отвечающие за пространственные отношения, разбиты на три больших класса. Предлоги  $SL$  могут выражаться в естественных языках разными лексическими средствами: предлогами, наречиями, словосочетаниями.

А — класс вертикальных отношений:

а — на (чем-либо), посередине, между (чем-либо), и т. д.;

а/ — над (чем-либо), сверху (чего-либо), вверх (к чему-либо, от чего-либо) и т. д.;

а\ — под (чем-либо), вниз (к чему-либо, от чего-либо), снизу (чего-либо) и т. д.

В случае наличия границы в виде линии или поверхности:

ля — на, по (поверхности);

аль — под (поверхностью);

ля/ — над (поверхностью);

ля/ль — между (поверхностями);

алья/ — сквозь, через (поверхность).

О — эта фонема определяет класс горизонтальных отношений:

о — посередине, в центре (чего-либо);

о/ — справа от (чего-либо);

о\ — слева от (чего-либо).

В случае наличия границы в виде линии или поверхности (в случае четырехмерного пространства границей будет трехмерный объем):

льо — справа от (поверхности); оль — слева от (поверхности);

льо/ — направо от (границы); льо\ — налево от (границы);

оль/ — справа к (границе); оль\ — слева к (границе);

льоль — внутри (замкнутой поверхности); ольо — вокруг (поверхности).

У — класс продольных отношений:

у — около, вблизи (чем-либо) и т. д.;

у/ — к (чему-либо), вперед (к чему-либо) и т. д.;

у\ — от (чего-либо), обратно, назад (от чего-либо) и т. д.;

лю — после, за (границей);

уль — перед; до (границы);

люль — внутри, под (поверхностью);

улю — сквозь, через (границу).

При употреблении отрицательных частиц, скажем "ха", с соответствующими предлогами смысл преобразуется:

ха у — не близко, вдали (от чего-либо)

Таким образом, ПрО в записи на SL выглядит следующим образом:

«лодка плыла вдоль берега» — [а (Д(а), ТоЗ(х)) = вдоль],

«дети бегали вокруг беседки» — [ольо (Д(а), ТоЗ(х)) = вокруг],

«бродить по городу» — [у/ (Д(э), ТоЗ(х̄)) = по],

«пронести через проходную» — [улю (Д(а), ТоЗ(х)) = через].

### Список условных обозначений

Об — локализуемый объект

Д — динамика

ПрО — пространственные отношения

СО — система отсчёта

ТоЗ — топологическая зона

### Список литературы

1. Куликов В. В., Елкин С. В., Гаврилов Д. А. Универсальный язык ДИАЛ как средство научного общения и производства открытий // Материалы XXV-х Циолковских чтений. Секция №6. Философские проблемы освоения Космоса. — Калуга: ФК СССР. 12. 09.1990.

2. Куликов В. В., Гаврилов Д. А., Елкин С. В., Универсальный искусственный язык — “hOOM-Диал”. Методические указания для изучающих язык, — М.: Гэлэкси Нэйшн, 1994. — 113 С.

3. Ёлкин С.В., Ёлкин С.С. Информационное исчисление и семантический язык SL как альтернатива универсальному сетевому языку UNL // Сб. трудов 5-го научно-практического семинара "Новые информационные технологии". — М.: МГИЭМ, 2002. — С. 97–132.

4. Ёлкин С.В. Ёлкин С.С. Информационное исчисление // НТИ. 2002. сер 2, № 11. — С. 17–24.

5. Ёлкин С. В. Открытый семантический язык SL // НТИ. 2003. сер 2, № 4. — С. 5–15.

6. Бетин В. Н., Ёлкин С. В., Хачукаев Э. М. Принципы построения семантического словаря для решения задачи устранения омонимии // НТИ. 2001. сер 2, № 1. — С. 34–37.

7. Ёлкин С. В., Бетин В. Н., Жигарев А. Е, Простаков О.В, Хачукаев Э. М. Разработка семантического анализа текстов при автореферировании // НТИ. 2001. сер 2, № 12.

8. Бетин В.Н., Ёлкин С.В. Применения семантического словаря для поиска информации и реферирования документов // НТИ. 2001. сер 2, № 6.

9. Ёлкин С. В., Куликов В. В., Клышинский Э. С., Максимов В. Ю., Аминева С. Н. Основы методологии трансляции знаний между различными предметными областями //Препринт № 73 за 2006г. ИПМ им. Келдыша М.В. РАН. Москва.

10. Всеволодова М. В. К вопросу о методологиях и методиках лингвистического анализа (на примере категорий пространственных, временных и причинных) (Статья вторая) // Вестник Московского университета. Сер. 9. Филология. 2005 № 2. — М.: Издательство Московского университета, 2005. — С. 9–34.

11. Пете И. Пространственность, предлоги, локальные отношения, картины мира и явления асимметричности // Вестник Московского университета. Сер. 9. Филология. 2004 № 3. — М.: Издательство Московского университета, 2004. — С. 61–74.

12. Гак В. Г. Пространство вне пространства // Логический анализ языка. Языки пространств. — М.: Языки русской культуры, 2000. — С. 127–134.

13. Мансурова О. Ю. Представление и передача пространственных отношений в многоязычной системе машинного перевода (на материале английского, русского и турецкого языков) / дис. канд. фил. наук. — М., 2005. — 160 С.

14. Мельчук И. А. Курс общей морфологии. Том II (Часть вторая: Морфологические значения) / Пер. с франц. — М.: ЯРК; Вена: WSA. — 405 С.

15. Плунгян В. А. К типологии глагольной ориентации/ Ред. Арутюнова, Ша-туновский, М.: 1999. — 110 С.