

Автоматический перевод правил дорожного движения в теоретико-графовую формальную модель

Менькин М.И.

Данная работа относится к области семантического анализа юридических документов. В статье приводится один из возможных подходов по моделированию правил движения. Описаны основные этапы семантического разбора некоторого класса предложений.

Ключевые слова: семантический анализ нормативно-правовых актов, прагматический анализ, построение семантической модели текста.

1. Введение

Данная статья продолжает работу по семантическому анализу текста Постановления Правительства РФ «О правилах дорожного движения» (далее — ПДД) [1], начатую в [2].

Суть работы заключается в следующем. У нас есть текст на естественном языке (ЕЯ), и мы хотим понять его содержание («смысл»). Ограничимся классом текстов, «смысл» которых можно формализовать. Здесь наиболее удачным выбором являются юридические документы, в частности, Правила дорожного движения, т.к. текст этого документа фактически является алгоритмом (т.е. набором действий для всех участников движения в разных дорожных ситуациях), написанного текстом. И мы хотим перевести эти алгоритмы, выраженные текстом на ЕЯ, в формальную модель (т.е. извлечь «смысл»). В нашем случае такая формальная модель представлена моделью дорожных ситуаций, основанная на теоретико-графовом подходе. Модель правил представлена функцией на графах.

Таким образом, наша задача — определение формальной модели (что ожидаем получить на выходе алгоритма) и сам алгоритм, переводящий текст в эту формальную модель.

Для перехода к формальной модели необходимо делать синтаксический разбор предложений. Для этого использовался ABBYY InfoExtractor Demo [3], являющийся лучшим среди существующих аналогов. На рис. 1 приведён пример разбора предложения в данной программе.

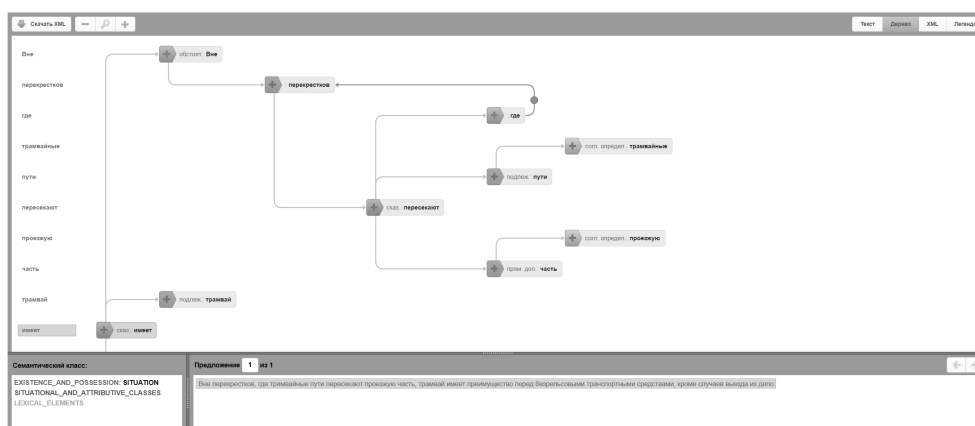


Рис. 1. Интерфейс ABBYY InfoExtractor Demo.

В данной статье описывается алгоритм перевода наиболее часто встречающихся конструкций с манёвром «Уступить дорогу» в формальную модель. В работе ограничиваемся только этим манёвром, поскольку «Уступить дорогу» — наиболее нетривиальный манёвр, т.к. условия его применимости зависят от типа и свойств участка дороги, на котором находится транспортное средство, типа и свойств соседних участков дороги, а также наличия и местоположения других участников дороги и их свойств.

Данная работа идейно продолжает исследования Е. М. Перпера по семантическому анализу юридических документов [4].

Автор выражает благодарность Э. Э. Гасанову за постановку задачи и оперативную помощь в процессе работы над статьёй.

Для проведения синтаксического разбора предложений в работе использовались технологии NLP компании ABBYY.

2. Основные определения и постановка задачи

Позаимствуем основные определения из [2] с небольшими изменениями.

Участок дороги (УД) — часть полосы движения, инвариантная относительно действующих на ней правил движения.

Граф дороги — раскрашенный граф $D = (V, E, P)$, где V — множество вершин, каждому элементу которого соответствует один УД, E — множество рёбер, каждому элементу которого соответствует одно отношение соседства «УД»–«УД» (т.е. ребром соединяем только примыкающие друг к другу УД), $P = P_v \times P_e$ — множество цветов, P_v — атрибуты вершин, P_e — атрибуты рёбер. Рёбра из E , которые соединяют УД, принадлежащие одной полосе, ориентированы по направлению движения на этой полосе.

Примеры цветов вершин: «тип перекрёстка» (если УД — перекрёсток), «главная дорога» (если УД является частью главной дороги), «жилая зона» (если УД — часть жилой зоны). Примеры цветов рёбер: «вид горизонтальной разметки» (если есть горизонтальная разметка), «наличие светофора», «сигнал светофора» (при наличии светофора).

Манёвр — потенциально возможное действие транспортного средства (ТС), в результате которого оно изменяет УД (например, меняет полосу движения, выезжает на перекрёсток или на УД с новым действующим знаком, и т.д.) или заметно изменяет свою скорость. В зависимости от дорожной ситуации, манёвр имеет статус «обязателен для исполнения», либо «допустим», либо «запрещён».

Перечислим манёвры, разбив их на четыре класса (в скобках — пример пункта, где манёвр встречается в ПДД):

- 1) Траекторные (относящиеся к изменению направления движения): перестроение (п. 8.1), разворот (п. 8.1), поворот (п. 8.1), обгон (п. 9.2), объезд (п. 9.2), опережение (п. 11.5);
- 2) Участковые (связанные с изменением участка дороги): съезд (п. 8.3), выезд (п. 9.4), въезд (п. 12.4);
- 3) Скоростные (относящиеся к изменению скорости): начало движения (п. 8.1), остановка (п. 8.1), повышение скорости (п. 10), снижение скорости (п. 10), резкое торможение (п. 10.5);
- 4) Уступить дорогу (не создавать помех) (п. 11.7) — сложный / условный манёвр; является комбинацией других манёвров, которая за-

висит от ДС; другими словами, этот манёвр — это совокупность других манёвров с условиями их применимости.

Пусть $D = (V, E, P)$ — граф дороги, T — множество ТС, M — множество манёвров. Тогда *дорожная ситуация* (ДС) — это пятёрка $S = (D, M, T, f, g)$, где $f: T \rightarrow V$ — функция позиционирования, которая каждому ТС из T ставит в соответствие один УД из V , $g: T \rightarrow M \cup \{*\}$ — функция назначения манёвра, которая каждому ТС из T ставит в соответствие один манёвр из M . Если функция назначения манёвра принимает значение $*$, то это понимается как «манёвр не определён». Множество T не обязательно должно содержать все ТС, которые находятся на дороге. Множество M не обязательно должно содержать все возможные манёвры.

Разрешение дорожной ситуации для данного транспортного средства — определение всех допустимых манёвров для данного ТС в данной ДС.

Разрешение дорожной ситуации — разрешение ДС для тех ТС, у которых значение манёвра неопределенно.

Уступить дорогу (не создавать помех) — сложный / условный манёвр; является комбинацией других манёвров, которая зависит от ДС; другими словами, этот манёвр — это совокупность других манёвров с условиями их применимости [2].

Текущая задача — моделирование правил, касающихся конкретного манёвра, для разрешения дорожной ситуации. Будем рассматривать предложения из ПДД, которые содержат манёвр «Уступить дорогу» (и двойственный ему «Имеет преимущество»). Этот манёвр представляется наиболее сложным и интересным для изучения.

3. Примеры предложений с манёвром «Уступить дорогу»

Всего в ПДД есть 48 предложений, содержащих ключевые слова «уступить» или «преимущество» (из которых 28 предложений соответствуют приводимому ниже синтаксическому шаблону). Поиск осуществлялся по ключевым словам «Уступить дорогу» и «Имеет преимущество». Приведём несколько таких предложений в качестве примера (здесь и далее нумерация предложений сквозная, по всему тексту ПДД):

520) При повороте налево или развороте вне перекрестка водитель безрельсового транспортного средства обязан уступить дорогу встречным транспортным средствам и трамваю попутного направления.

719) При повороте направо или налево водитель обязан уступить дорогу пешеходам и велосипедистам, пересекающим проезжую часть дороги, на которую он поворачивает.

735) Если сигналы светофора или регулировщика разрешают движение одновременно трамваю и безрельсовым транспортным средствам, то трамвай имеет преимущество независимо от направления его движения.

4. Шаблон дороги

Центральное место в нашей модели занимает «Шаблон дороги».

Правила движения ТС иногда зависят не только от того, какого типа участок дороги, на котором находится данное ТС (например, «разворот на мосту запрещён»), но и от того, какие участки дороги являются соседними для данного ТС (например, «на нерегулируемом перекрёстке при движении по неглавной дороге водитель должен уступить дорогу другим ТС, которые движутся по главной»). Для семантического разбора предложений и формализации понятия «правило движения (ТС)» необходимо ввести понятие «шаблон дороги».

Поясним, что будем понимать под шаблоном дороги. Для этого рассмотрим ТС, находящееся на дороге (рис. 2а). Можно считать, что дорога делится на 8 частей относительно ТС (рис. 2б). Эти части дороги можно схематично изобразить как показано на рис. 2в. Схему можно упростить, опустив цифры, и далее будем использовать упрощённый вид шаблона дорог, изображённый на рис. 2г. Чёрный квадрат в центре — ТС (для которого определяется правило движения); если скорость ТС отличная от нуля, предполагаем, что ТС движется в квадрат «вверх». Белые недиагональные квадраты — это части УД, ограниченные по размеру ТС (или шириной полосы движения). Белые диагональные квадраты — другие УД (соседние полосы; участки, принадлежащие перекрёстку или съезду с дороги, и т.д.). Некоторые белые квадраты могут отсутствовать (из возможных 8) — значит, правило ничего не требует от типа УД в этом месте (в т.ч. нахождение какого-либо ТС на нём).

Под шаблоном дороги (для каждого правила) будем понимать один из объектов, изображенных на рис. 3. Тогда «правилом движения» будем называть множество шаблонов дороги с указанием свойств для каждого «квадрата» шаблона дороги (тип УД, наличие ТС и их свойств, и т.д.).

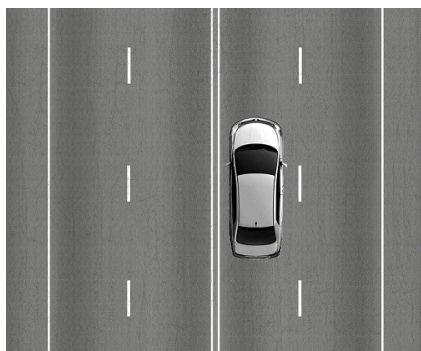


Рис. 2а.

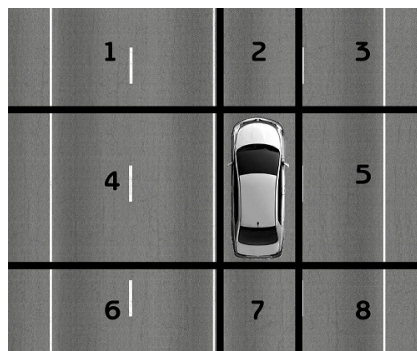


Рис. 2б.

1	2	3
4	ТС	5
6	7	8

Рис. 2в.

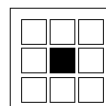


Рис. 2г.

Шаблон дороги для каждого правила определяется манёвром, о котором говорится в правиле. Исключение — манёвр «Уступить дорогу», т.к. этот манёвр состоит из нескольких других простых манёвров, поэтому может содержать более одного шаблона дороги.

«Чёрным квадратом» шаблона дороги для манёвра «Уступить дорогу» будет ТС, относительно которого принимаем решение о допустимости манёвра, «белые квадраты» могут содержать других участников движения. Типы ТС в шаблоне дороги будут множественными типами, каждое ТС будет иметь траекторию в виде упорядоченного множества УД с указанием начала и конца траектории. В качестве начала и конца траектории могут быть выбраны УД или другие участники движения.

Таким образом, «чёрный квадрат» содержит (множественный) тип и свойства (при наличии) ТС, которое уступает дорогу. «Белый квадрат» содержит: тип и свойства (при наличии) участка или участков дорог, которым соответствует «квадрат», и тип и свойства (при наличии) ТС, которые находятся на данном «квадрате».

Правило «Уступить дорогу» будет определяться совокупностью всех шаблонов дорог, относящимися к этому манёвру.

Если говорить о произвольных манёврах, то шаблон дороги — модель дорожной ситуации, где для ТС (изображенного чёрным квадратом) определяется допустимость того или иного манёвра в зависимости от типа текущего участка дороги, типов соседних участков дорог и дру-

гих участников движения, которые могут находиться на соседних участках дороги.

Шаблоны определяются по ключевым словам, соответствующим названиям шаблонов. Расскажем подробно про каждый шаблон дороги (рис. 3), при необходимости используя нумерацию «белых квадратов» согласно рис. 5в.

Шаблон 1. «При приближении...». Данный шаблон работает в ситуации, когда речь идёт о преимуществе другого ТС, независимо от его местоположения (по отношению к нашему ТС). В зависимости от контекста речь может идти, например, о приближении справа (квадраты 3, 5 и 8).

Шаблон 2. «Обгон». Шаблон, помимо обгона, в некоторых ситуациях соответствует объезду (слева). Квадраты 1 и 4 говорят о необходимости уступить дорогу встречным ТС (на этих квадратах). Квадраты 4, 6 и 7 — о необходимости уступить дорогу тем ТС, которые уже начали обгон (опережение, объезд).

Шаблон 3. «Объезд справа». Этот шаблон также соответствует опережению (справа). Комментарии аналогичны шаблону №2 (с поправкой на то, что обгона справа не существует).

Шаблоны 4.1, 4.2. «Перестроение» (левое и правое). Также соответствует опережению (слева и справа). Комментарии аналогичны шаблону №2 и №3, с той разницей, что нет встречных ТС.

Шаблон 5. «Разворот» и «Поворот налево». При совершении данных манёвров необходимо уступить дорогу ТС, движущимся с квадратов 2 и 3 в сторону 1 квадрата, встречным ТС с квадрата 1, а также трамваю и попутным ТС с квадратов 4 и 6.

Шаблон 6. «Поворот направо». Также используется в ситуации выезда на круговой перекрёсток и при выезде на перекрёсток неравнозначных дорог со стороны второстепенной дороги. Дорога уступается ТС, движущимся с квадратов 1 и 2 (в сторону 3 квадрата) — если поворот направо или круговой перекрёсток, при этом при повороте направо в некоторых ситуациях необходимо уступать дорогу квадрату 3 (например, пешеходам, которые переходят дорогу во время совершения манёвра). При выезде на перекрёсток неравнозначных дорог со стороны второстепенной дороги нужно уступать дорогу квадратам 1 и 3, а также тем ТС, которые двигаются по 2 квадрату с направлений 1-го и 3-го.

Шаблон 7. «Выезд» (кроме выездов на перекрёсток). Также соответствует съезду и въезду. Используется в случае, если ТС меняет участок дороги, и ТС нового УД имеют преимущество. Шаблон имеет

несколько разновидностей, белым квадратом может быть 2, 4, 5 или 7, но только один (на рисунке белый квадрат — 2).

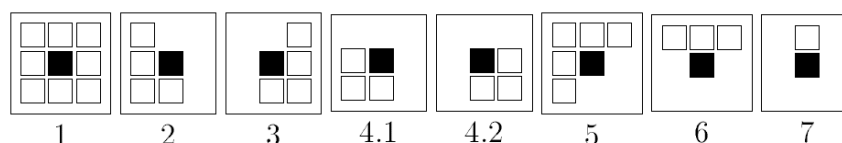


Рис. 3. Шаблоны дорог.

5. Синтаксические шаблоны предложений

Для конструирования «правил движения» необходимо извлечь всю информацию, которая выражена текстом в документе для рассматриваемого манёвра (под конструированием подразумеваем размещение ТС, перечисленных в тексте, на «квадратах» шаблона дороги, с указанием их свойств и свойств УД, и т.д.). Каждое «правило движения» обязательно содержит объект, который выполняет манёвр — «чёрный квадрат» (КТО_УСТУПАЕТ) и (возможно) объекты, которым надо уступить дорогу, находящихся на «белых квадратах» (КОМУ_УСТУПИТЬ). Также правила могут содержать в себе условия на особенности данной дорожной ситуации (СИТУАЦИЯ).

Общий вид предложений, для которых ниже перечислены синтаксические шаблоны, приведён на рис. 4 (на примере шаблона «должен уступить»). Выбор именно таких предложений обусловлен тем, что таких предложений больше (поэтому другие можно считать «исключениями»). Предложения, которые синтаксически устроены иначе (другое устройство синтаксического дерева и другие синтаксические шаблоны), в данной работе рассматриваться не будут.

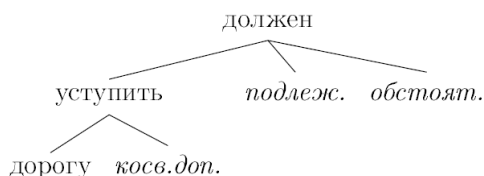


Рис. 4. Шаблон синтаксического дерева для манёвра «Уступить дорогу».

Помимо прочего, в правилах иногда встречается условие про «пересечение траекторий ТС», но оно подразумевается всегда (т.е. и в том случае, когда не говорится об этом), поэтому при разборе предложений эти слова будем игнорировать.

СИТУАЦИЯ может содержать в себе:

- условия на движение по текущему УД или планируемому для проезда;
- простые манёвры, совершаемые КТО_УСТУПАЕТ (например, «разворот»);
- место движения (тип УД).

КТО_УСТУПАЕТ и КОМУ_УСТУПИТЬ содержат списки ТС с указанием их типов и свойств.

Ниже приведём детальную характеристику возможных синтаксических шаблонов предложений, которые соответствуют значениям полей для каждого объекта правила движения, и которые в дальнейшем будут использоваться для алгоритма разбора предложений. Указывать значения полей будем четвёрками (слово, часть речи, семантический класс, ярус синтаксического дерева). Под семантическим классом будем понимать названия множества слов с эквивалентной семантикой, при этом применяя онтологию Linguistic [5]. Семантические классы понятны по их названию.

Используемые сокращения для синтаксических ролей слов предложения:

- подлежаж. — подлежащее
- сказ. — сказуемое
- согл.определ. — согласованное определение
- несогл.определ. — несогласованное определение
- прям.доп. — прямое дополнение
- косв.доп. — косвенное дополнение
- обстоят. — обстоятельство

Смежные (по синтаксическому дереву) слова обозначим знаком «+». Словом *обстоятельство* обозначаем произвольное слово, являющееся одноимённой частью речи.

Отметим важную деталь относительного синтаксического разбора предложений — для автоматического разбора предложения по синтаксическому шаблону с фиксированным множеством слов для каждой части шаблона (т.е. в каждом шаблоне множество возможных используемых слов полностью определено), для каждого слова предложения требуются наличие следующих свойств:

- слово-родитель — требуется всегда;
- синтаксическая роль — требуется всегда;
- порядок слова в предложении (или иначе, слова-соседи) — требуется при наличии у слова зависимых служебных частей речи или однородных с ним членов.

5.1. Начало разбора

Среди предложений, которые относятся к манёвру «Уступить дорогу» в данной статье мы будем рассматривать только следующие базовые конструкции. Остальные типы предложений в рамках данной статьи не рассматриваются. В предложениях, связанных с манёвром «Уступить дорогу», встречаются следующие базовые конструкции:

- (должен / должны, сказ., MODAL, 0) + (уступить, прям.доп., TO_CEDE, 1) + (дорогу, прям.доп., ROAD, 2);
- (обязан, сказ., TO_OBLIGE, 0) + (уступить, прям.доп., TO_CEDE, 1) + (дорогу, прям.доп., ROAD, 2);
- (имеет, сказ., EXISTENCE_AND_POSSESSION, 0) + (преимущество, прям.доп., CH_BENEFIT, 1)

5.2. Объект КТО_УСТУПАЕТ

В данном пункте перечислим существующие типы КТО_УСТУПАЕТ, встречающихся в ПДД:

- (водитель / водители, подлежа., TRANSPORT_DRIVER, 1);
- (трамвай, подлежа., TRAM, 1);
- (пешеходы, подлежа., ON_FOOT, 1);
- (велосипедисты, подлежа., CYCLIST, 1).

Для «водитель» возможно наличие такого свойства:

- (безрельсового, согл.определ., NO_PREFIXOID + RAIL, 4-3) + (транспортного, согл.определ., TRANSPORT, 3) + (средства, несогл.определ., TRANSPORT, 2).

5.3. Объект СИТУАЦИЯ

В рассматриваемых предложениях может быть несколько типов возможных синтаксических конструкций, которые могут встречаться в СИТУАЦИЯ.

«Условие на событие» характеризуется наличием следующих синтаксических конструкций:

- ((в) случае / (в) случаях / (кроме) случаев, обстоят., CASE, 1)
- ((при) наличии, обстоят., EXISTENCE_AND_POSSESSION, 2) + (препятствия, несогл.определ., OBSTACLE_AS_OBJECT_BY_FUNCTION, 3)
- ((при) включении, обстоят., TO_ACTIVATE, 1) + (сигнала, несогл.определ., SIGNALING_ACTION, 2) + (светофора, несогл.определ., TRAFFIC_LIGHTS, 3)

«Условие на движение» характеризуется наличием следующих синтаксических конструкций:

- ((при) движении, обстоят., GENERAL_MOTION, 1) + *обстоятельство*;
- ((при) выезде, обстоят., MOTION_ON_DEVICES, 1) + *обстоятельство*;
- ((при) съезде, обстоят., MOTION_ON_DEVICES, 1) + *обстоятельство*;
- ((при) перестроении, обстоят., TO_PLACE_IN_ORDER, 1) + ((транспортных) средств, несогл.определ., TRANSPORT, 2) + (движущихся, согл.определ., GENERAL_MOTION, 3) + *обстоятельство*;
- ((при) повороте, обстоят., TO_TURN, 1) + *обстоятельство*;
- ((при) развороте, обстоят., TO_TURN, 1) + *обстоятельство*;

«Тип УД» характеризуется наличием следующих синтаксических конструкций:

- ((на) перекрёстке, обстоят., CROSS_PLACE, 1) + (дорог, несогл.определ., ROAD, 2) + (равнозначных / неравнозначных, согл.определ., EQUAL_SAME, 3)
- ((в) пунктах, обстоят., INHABITED_LOCALITY, 1) + (населённых, согл.определ., POPULATION_DENSITY, 2)

- ((в) зоне, обстоят., ZONE, 1) + (жилой, согл.определ., RESIDENTIAL, 2)
- ((в) зоне, обстоят., AREA_OF_ACTIVITY_AS_WHOLE, 1) + (велосипедной, согл.определ., BICYCLE, 2)
- (вне, обстоят., ZONE, 1) + (перекрёстков, косв.доп., CROSS_PLACE, 2)

5.4. Объект КОМУ УСТУПИТЬ

В данном пункте перечислим существующие типы КОМУ УСТУПИТЬ:

- ((транспортному) средству / (транспортным) средствам, косв.доп., TRANSPORT, 2);
- (пешеходам, косв.доп., ON_FOOT, 2);
- (велосипедистам, косв.доп., CYCLIST, 2);
- (трамваю, косв.доп., TRAM, 2).

Перечислим возможные свойства типов КОМУ УСТУПИТЬ:

- (движущимся / движущегося, согл.определ., GENERAL_MOTION, 3) + *обстоятельство*;
- (находящемуся, согл.определ., POSITION_IN_SPACE, 3) + *обстоятельство*;
- (приближающимся / приближающегося, согл.определ., TO_GO_AND_TRANSFER, 3)
- (встречным, согл.определ., OPPOSITE, 3);
- (попутного, согл.определ., FOLLOWING, 4) + (направления, несогл.определ., REFERRAL, 3);
- (начинающим / завершающим, согл.определ., BEGIN / END_TO_TAKE_PLACE, 3) + (движение, прям.доп., GENERAL_MOTION, 4);
- для пешеходов: (переходящим, согл.определ., TO_WALK, 3)
- для пешеходов: (вступившим, согл.определ., TO_STEP, 3)

6. Основные этапы разбора

Как было сказано в предыдущем разделе, несмотря на то, что семантически предложения о манёвре «Уступить дорогу» выглядят одинаково

(можно однозначно выделить части предложений, определяющие тип манёвра, КТО_УСТУПАЕТ, КОМУ_УСТУПИТЬ, СИТУАЦИЯ), синтаксически они устроены по-разному. Разнообразие синтаксических конструкций делает юридический документ более читабельным, но усложняет его автоматический разбор.

Поэтому перед проведением разбора предложений необходимо отсеять от всего текста только те, которые, во-первых, относятся к рассматриваемому нами манёвру и, во-вторых, совпадают с синтаксическими шаблонами «Начало разбора», приведённым в предыдущем разделе. Первые три пункта — поиск именно таких предложений. Последующие пункты — семантический разбор (т.е. построение соответствия «текст»–«правило движения»).

1. Ищем ключевые слова «уступить» или «преимущество» (найдено: 48).

2. Среди найденных предложений ищем те, у которых корневое слово синтаксического графа — сказуемое «должен», «обязан», «имеет» (найдено: 33).

3. Среди найденных предложений ищем один из шаблонов «должен уступить», «обязан уступить», «имеет преимущество» (найдено: 28).

Перечислим этапы разбора для предложений, найденных после третьего пункта. При этом используя синтаксические шаблоны из предыдущего раздела.

4. Определяем тип манёвра — «Уступить дорогу» или «Имеет преимущество».

5. Определяем тип предложения — простое или сложное. Предложение будет считаться сложным, если сказуемых, которые относятся к манёвру больше одного; либо есть фразы «кроме случаев», «однако»; либо в предложении (а точнее, в обстоятельстве предложения) указано более одного манёвра (через запятую или слово «или»). Т.е., в нашем понимании, сложное предложение — это то, которое содержит более одного правила (но не всегда разным правилам соответствуют разные шаблоны дорог, см. примеры ниже). Например, синтаксические шаблоны «обязан уступить при съезде ..., а также при выезде с ...» и «должен уступить ..., кроме случаев ...» соответствуют сложным предложениям, которые содержат в себе по два правила сразу. Если предложение окажется сложным, то следующие пункты (6–9) применяем поочерёдно к каждой части предложения, содержащей своё правило.

6. Подлежащее (зависящее от «должен» / «обязан» / «имеет») и его зависимые слова проверяем на наличие шаблонов КТО_УСТУПАЕТ.

7. Обстоятельство (зависящее от «должен» / «обязан» / «имеет») и его зависимые слова проверяем на наличие шаблонов СИТУАЦИЯ. Выбираем шаблоны дороги, которые соответствуют СИТУАЦИЯ. Определяем типы «белых квадратов».

8. Косвенные дополнения (зависящие от «уступить» / «преимущество») и их зависимые слова проверяем на наличие шаблонов КОМУ_УСТУПИТЬ. Определяем перечень других ТС и их свойства.

9. Определяем местоположение ТС на шаблоне дороги. Если манёвр «Уступить дорогу», то «чёрный квадрат» — КТО_УСТУПАЕТ, «белые квадраты» — КОМУ_УСТУПИТЬ. Если манёвр «Имеет преимущество», то «чёрный квадрат» — КОМУ_УСТУПИТЬ, «белые квадраты» — КТО_УСТУПАЕТ (т.е. меняем значение таким образом, чтобы правило всегда означало «чёрный квадрат уступает»).

7. Примеры разбора предложения

Приведём примеры разборов некоторых предложений, содержащих манёвр «Уступить дорогу». Синтаксический анализ предложений осуществлялся при помощи АБВУУ InfoExtractor Demo. Результат работы продукта — XML-файл с разбором всех предложений документа. В первом примере в квадратных скобках указаны элементы и атрибуты элементов XML-файла, которые используются для анализа предложения. Также для первого примера приведено полное синтаксическое дерево предложения.



Рис. 5а. Синтаксическое дерево (пример 1).

Пример 1. «При въезде на перекресток, на котором организовано круговое движение и который обозначен знаком 4.3, водитель транспортного средства обязан уступить дорогу транспортным средствам, дви-

жущимся по такому перекрестку». На рис. 5а приведён синтаксический граф предложения.

1. Манёвр один, сказуемое манёвра одно, ключевые слова «однако» и «кроме» отсутствуют, значит предложение простое.

2. Шаблон «начала разбора» — «обязан уступить», значит тип манёвра — Уступить дорогу.

3. Определяем КТО_УСТУПАЕТ. Это должно быть подлежащее [`<syntactic_role>=Subject`], зависящее [`<syntactic_parent>`] от сказуемого «обязан» [`<syntactic_role>=Predicate`]. В данном случае — «водитель». От «водитель» идут зависимые [`<syntactic_parent>`] слова «транспортного» [`<syntactic_role>=AgreedAttribute`] средства [`<syntactic_role>=NonAgreedAttribute`]. Дополнительных свойств нет. Значит, КТО_УСТУПАЕТ — ТС (без каких-либо свойств, т.е. любое).

4. Определяем СИТУАЦИЯ. Обстоятельство, зависящее от сказуемого — «(при) выезде» [`<syntactic_role>=AdverbialModifier`]. Предварительный шаблон для правила — номер 7 («выезд»). Продолжаем разбор, смотрим на зависимые [`<syntactic_parent>`] от «при выезде» слова. Это обстоятельство «(на) перекрёсток» [`<syntactic_role>=AdverbialModifier`]. Нужно определить свойства перекрёстка, продолжаем разбор от слов «на перекрёсток». Определяем, что это перекрёсток с круговым движением. Изменяем шаблон на номер 6 («поворот направо»), т.к. это выезд на перекрёсток с круговым движением. Определяем свойства «белых квадратов» перед ТС — «круговой перекрёсток».

5. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «(транспортным) средствам» [`<syntactic_role>=ObjectIndirect`], значит множественный тип КОМУ_УСТУПИТЬ — ТС. Определяем возможные свойства, продолжаем разбор от слова «средствам». Единственное зависимое [`<syntactic_parent>`] слово — «движущимся» [`<syntactic_role>=AgreedAttribute`] и от него «по перекрестку такому». Это полностью соответствует «белым квадратам» шаблона, дополнительных свойств нет.

Разбор предложения завершён (рис. 5б).

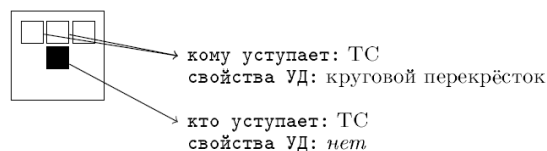


Рис. 5б. Разбор предложения (пример 1).

Пример 2. «При выезде на дорогу с прилегающей территории водитель должен уступить дорогу транспортным средствам и пешеходам, движущимся по ней, а при съезде с дороги — пешеходам и велосипедистам, путь движения которых он пересекает».

1. Присутствует эллипсис сказуемого «должен» и два манёвра — «при выезде» и «при съезде», значит предложение сложное. Предложение содержит два правила — «должен уступить при выезде...» и «должен уступить при съезде...».

Разделим дальнейший предложения разбор на 2 части, начнём с «должен уступить при выезде...».

2.1. Шаблон «начала разбора» — «должен уступить», значит тип манёвра — Уступить дорогу.

2.2. Определяем КТО_УСТУПАЕТ, это подлежащее, зависящее от сказуемого «должен». В данном случае — «водитель». От «водитель» нет зависимых слов. Дополнительных свойств нет. Значит, КТО_УСТУПАЕТ — ТС (без каких-либо свойств, т.е. любое).

2.3. Определяем СИТУАЦИЯ. Обстоятельство, зависящее от сказуемого — «при выезде». Предварительный шаблон для правила — номер 7 («выезд»). Продолжаем разбор, смотрим на зависимые от «при выезде» слова. Это обстоятельства «на дорогу» и «с территории». Нужно определить свойства «на дорогу» и «с территории». Продолжаем разбор от слов «на дорогу», зависимых слов нет. Значит, дополнительных свойств нет. Продолжаем разбор от слов «с территории», есть зависимое слово «прилегающей». Определяем, что это прилегающая территория, дополнительных свойств нет. Шаблон остаётся таким же (номер 7, «выезд»). Определяем свойства «белого квадрата» — «дорога». Определяем свойство «чёрного квадрата» — «прилегающая территория».

2.4. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенные дополнения, зависящее от прямого дополнения «уступить». Это «транспортным средствам» и «пешеходам», значит КОМУ_УСТУПИТЬ соответствуют два множественных типа КОМУ_УСТУПИТЬ — ТС и пешеход. Определяем возможные свойства, продолжаем разбор от слов «средствам» и «пешеходам». Единственное зависимое слово — «движущимся», а от него «по ней» в значении «по дороге» (анафора). Это соответствует «белому квадрату» шаблона, дополнительных свойств нет.

Разбор первой части предложения завершён.

Разбираем «должен уступить при съезде...».

3.1. Аналогичен пункту 2.1.

3.2. Аналогичен пункту 2.2.

3.3. Определяем СИТУАЦИЯ. Обстоятельство, зависящее от сказуемого — «при съезде». Предварительный шаблон для правила — номер 7 («выезд»). Продолжаем разбор, смотрим на зависимые от «при съезде» слова. Это обстоятельство «с дороги». Нужно определить свойства дороги, продолжаем разбор от слов «с дороги». Зависимых слов нет, значит нет и дополнительных свойств. Шаблон остаётся таким же (номер 7, «выезд»). Определяем свойство «чёрного квадрата» — «дорога». Т.к. манёвр — съезд, то определяем свойство «белого квадрата» как «НЕ дорога».

3.4. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «пешеходам» и «велосипедистам», значит КОМУ_УСТУПИТЬ соответствуют два множественных типа КОМУ_УСТУПИТЬ — пешеход и велосипедист. Определяем возможные свойства, продолжаем разбор от слов «пешеходам» и «велосипедистам». Единственное зависимое слово — «пересекает» в составе конструкции «путь которых он пересекает». Поскольку манёвр «Уступить дорогу» всегда подразумевает пересечение траекторий, то данную конструкцию игнорируем. Других свойств нет.

Разбор второй части, а значит всего предложения завершён (рис. 6).

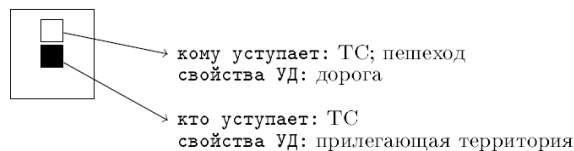


Рис. 6а. Разбор первой части (пример 2).

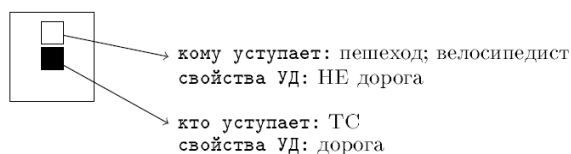


Рис. 6б. Разбор второй части (пример 2).

Пример 3. «При перестроении водитель должен уступить дорогу транспортным средствам, движущимся попутно без изменения направления движения».

1. Манёвр один, сказуемое манёвра одно, ключевые слова «однако» и «кроме» отсутствуют, значит предложение простое.

2. Шаблон «начала разбора» — «должен уступить», значит тип манёвра — Уступить дорогу.

3. Определяем **КТО_УСТУПАЕТ**. Это должно быть подлежащее, зависящее от сказуемого «должен». В данном случае — «водитель». От «водитель» нет зависимых слов, дополнительных свойств нет. Значит, **КТО_УСТУПАЕТ** — ТС (без каких-либо свойств, т.е. любое).

4. Определяем **СИТУАЦИЯ**. Обстоятельство, зависящее от сказуемого — «при перестроении». Предварительный шаблон для правила — номер 4 («перестроение»). Шаблонов, соответствующих перестроению, два (левое и правое перестроение). Поэтому предложению будет соответствовать два шаблона. Продолжаем разбор, зависимых от «при перестроении» слов нет, значит других свойств **СИТУАЦИЯ** нет. Определяем свойства «белых квадратов» — «односторонняя полоса».

5. Определяем тип **КОМУ_УСТУПИТЬ**. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «транспортным средствам», значит множественный тип **КОМУ_УСТУПИТЬ** — ТС. Определяем возможные свойства, продолжаем разбор от слова «средствам». Единственное зависимое слово — «движущимся» и от него две зависимости — «попутно» и «без изменения направления движения». Свойство «попутно» соответствует «белым квадратам» шаблона, дополнительное свойство — «без изменения направления движения».

Разбор предложения завершён (рис. 7).

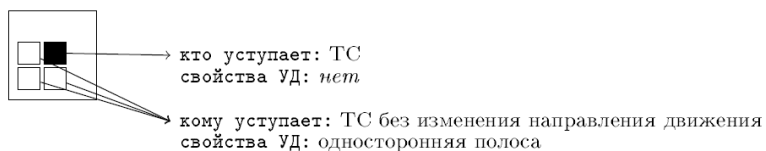


Рис. 7а. Первый шаблон (пример 3).

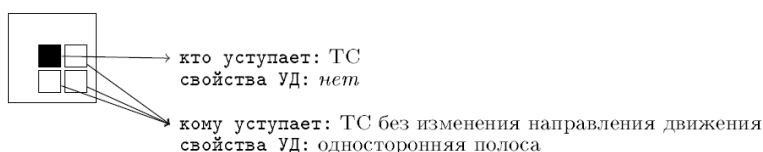


Рис. 7б. Второй шаблон (пример 3).

Пример 4. «При повороте налево или развороте вне перекрестка водитель безрельсового транспортного средства обязан уступить дорогу встречным транспортным средствам и трамваю попутного направления».

1. Присутствуют два манёвра (поворот налево и разворот), значит предложение сложное.

2. Шаблон «начала разбора» — «обязан уступить», значит тип манёвра — Уступить дорогу.

3. Определяем КТО_УСТУПАЕТ. Это должно быть подлежащее, зависящее от сказуемого «обязан». В данном случае — «водитель». От «водитель» идут зависимые слова «транспортного средства» и «безрельсового». Других зависимых слов нет. Значит, КТО_УСТУПАЕТ — ТС с дополнительным свойством «безрельсовое».

4. Определяем СИТУАЦИЯ. Обстоятельства, зависящие от сказуемого — «при повороте налево» и «развороте». И повороту налево, и развороту соответствует один шаблон — номер 5. Продолжаем разбор, смотрим на зависимые от «при повороте налево» и «развороте» слова. Это обстоятельство «вне» с зависимым словом «перекрёстка». Других зависимых слов нет. Определяем свойства «белых квадратов» и «чёрного квадрата» — «НЕ перекрёсток».

5. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «транспортным средствам» и «трамваю», значит множественный тип КОМУ_УСТУПИТЬ — ТС и трамвай. Определяем возможные свойства. Продолжаем разбор от слова «средствам». Получаем дополнительное свойство — «встречным». Других зависимых слов от «средствам» нет. Продолжаем разбор от слова «трамваю». Получаем дополнительное свойство — «направления» с зависимым словом «попутного». Других зависимых слов от «трамваю» нет. Значит, КОМУ_УСТУПИТЬ — встречные ТС и попутный трамвай. Т.к. движение правостороннее, а трамвайные пути находятся посередине дороги, то встречные ТС и трамваи могут находиться только на левых квадратах шаблона (это замечание не обязательно нужно учитывать и можно было доопределить и другие квадраты шаблона, т.к. на практике такая ситуация всё равно не возникнет).

Разбор предложения завершён (рис. 8).

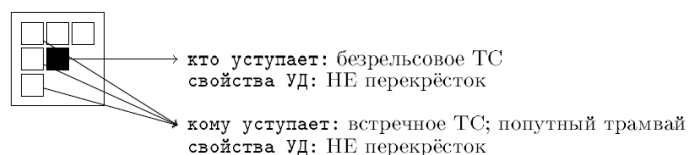


Рис. 8. Итог разбора предложения (пример 4).

Пример 5. «При повороте направо или налево водитель обязан уступить дорогу пешеходам и велосипедистам, пересекающим проезжую часть дороги, на которую он поворачивает».

1. Присутствуют два манёвра (поворот направо и поворот налево), значит предложение сложное.

2. Шаблон «начала разбора» — «обязан уступить», значит тип манёвра — Уступить дорогу.

3. Определяем КТО_УСТУПАЕТ. Это должно быть подлежащее, зависящее от сказуемого «обязан». В данном случае — «водитель». От «водитель» нет зависимых слов, дополнительных свойств нет. Значит, КТО_УСТУПАЕТ — ТС (без каких-либо свойств, т.е. любое).

4. Определяем СИТУАЦИЯ. Обстоятельства, зависящие от сказуемого — «при повороте направо или налево». Повороту налево соответствует шаблон номер 5, повороту направо — шаблон номер 6. Зависимых от обстоятельств слов нет. Значит, у «белых квадратов» и «чёрного квадрата» нет свойств УД.

5. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «пешеходам» и «велосипедистам», значит множественный тип КОМУ_УСТУПИТЬ — пешеход и велосипедист. Определяем возможные свойства. Продолжаем разбор, единственное зависимое слово — «пересекающим» и от него конструкция «проезжую часть дороги, на которую он поворачивает». Это соответствует «белому квадрату» №1 шаблона №5 и «белому квадрату» №3 шаблона №6. Других дополнительных свойств нет.

Разбор предложения завершён (рис. 9).

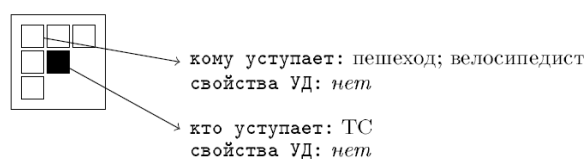


Рис. 9а. Поворот налево (пример 5).

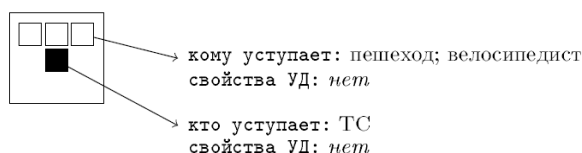


Рис. 9б. Поворот направо (пример 5).

Пример 6. «При повороте налево или развороте по зеленому сигналу светофора водитель безрельсового транспортного средства обязан уступить дорогу транспортным средствам, движущимся со встречного направления прямо или направо.»

1. Присутствуют два манёвра (поворот налево и разворот), значит предложение сложное.

2. Шаблон «начала разбора» — «обязан уступить», значит тип манёвра — Уступить дорогу.

3. Определяем КТО_УСТУПАЕТ. Это должно быть подлежащее, зависящее от сказуемого «обязан». В данном случае — «водитель». От «водитель» идут зависимые слова «транспортного средства» и «безрельсового». Других зависимых слов нет. Значит, КТО_УСТУПАЕТ — ТС с дополнительным свойством «безрельсовое».

4. Определяем СИТУАЦИЯ. Обстоятельства, зависящие от сказуемого — «при повороте налево» и «развороте». И повороту налево, и развороту соответствует один шаблон — номер 5. Других зависимых от обстоятельства слов нет. Однако присутствует условие на событие, у сказуемого «обязан» есть зависимое косвенное дополнение «по сигналу». Далее, у «по сигналу» есть зависимые слова «зелёному» и «светофора». Других зависимых слов нет. Определяем свойство передних «белых квадратов» (№1, №2, №3) — «перекрёсток» и остальных «белых квадратов» и «чёрного квадрата» — «НЕ перекрёсток».

5. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «транспортным средствам», значит множественный тип КОМУ_УСТУПИТЬ — ТС. Определяем возможные свойства, продолжаем разбор от слова «средствам». Единственное зависимое слово — «движущимся» и от него фразы «со встречного направления» и «прямо или направо». Значит, КОМУ_УСТУПИТЬ — встречное ТС с направлением движения прямо или направо.

Разбор предложения завершён (рис. 10).

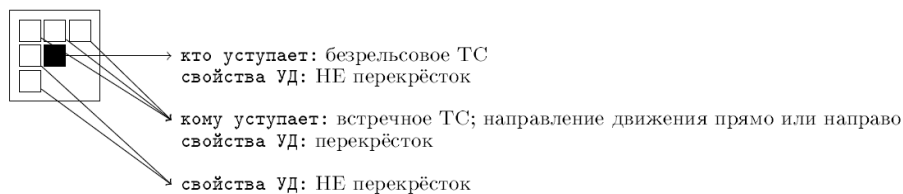


Рис. 10. Итог разбора предложения (пример 6).

Пример 7. «На перекрестке неравнозначных дорог водитель транспортного средства, движущегося по второстепенной дороге, должен уступить дорогу транспортным средствам, приближающимся по главной, независимо от направления их дальнейшего движения».

1. Манёвр один, сказуемое манёвра одно, ключевые слова «однако» и «кроме» отсутствуют, значит предложение простое.

2. Шаблон «начала разбора» — «должен уступить», значит тип манёвра — Уступить дорогу.

3. Определяем КТО_УСТУПАЕТ. Это должно быть подлежащее, зависящее от сказуемого «должен». В данном случае — «водитель». От «водитель» идут зависимые слова «транспортного средства» и «движущегося». От слова «движущегося» идут зависимые слова «по второстепенной дороге». Эта фраза накладывает ограничение на УД. Добавляем свойство «второстепенная дорога» для «чёрного квадрата». У «водитель» других дополнительных свойств нет. Значит, КТО_УСТУПАЕТ — ТС (без каких-либо свойств, т.е. любое).

4. Определяем СИТУАЦИЯ. Обстоятельство, зависящее от сказуемого — «на перекрёстке». Продолжаем разбор, смотрим на зависимые от «на перекрёстке» слова. Это фраза «неравнозначных дорог». Пользуясь фразой из предыдущего пункта о движении по второстепенной дороге, определяем, что это выезд на перекрёсток неравнозначных дорог со стороны второстепенной дороги. Поэтому выбираем шаблон номер 6 («поворот направо»). Определяем свойства «белого квадрата» номер 2 — «перекрёсток неравнозначных дорог», и для других «белых квадратов» — «главная дорога».

5. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «транспортным средствам», значит множественный тип КОМУ_УСТУПИТЬ — ТС. Определяем возможные свойства, продолжаем разбор от слова «средствам». Единственное зависимое слово — «приближающимся» и от него «по главной дороге». Это полностью соответствует 1-му и 3-му «белым квадратам» шаблона, дополнительных свойств нет.

6. Также в предложении есть фраза «независимо от направления их дальнейшего движения», но именно это и предполагается по умолчанию во всех шаблонах правил и данная фраза не вносит никаких дополнительных ограничений на свойства ТС.

Разбор предложения завершён (рис. 11).

Пример 8. «При выезде из жилой зоны водители должны уступить дорогу другим участникам движения».

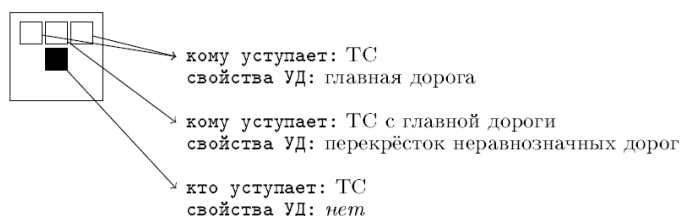


Рис. 11. Итог разбора предложения (пример 7).

1. Манёвр один, сказуемое манёвра одно, ключевые слова «однако» и «кроме» отсутствуют, значит предложение простое.

2. Шаблон «начала разбора» — «должны уступить», значит тип манёвра — Уступить дорогу.

3. Определяем КТО_УСТУПАЕТ. Это должно быть подлежащее, зависящее от сказуемого «должны». В данном случае — «водители». От «водители» нет других зависимых слов. Значит, дополнительных свойств нет. Определяем КТО_УСТУПАЕТ — ТС (без каких-либо свойств, т.е. любое).

4. Определяем СИТУАЦИЯ. Обстоятельство, зависящее от сказуемого — «при выезде». Предварительный шаблон для правила — номер 7 («выезд»). Продолжаем разбор, смотрим на зависимые от «при выезде» слова. Это фраза «из жилой зоны». Шаблон дороги остаётся таким же. Определяем свойство «чёрного квадрата» — «жилая зона». Определяем свойства «белого квадрата» — «НЕ жилая зона».

5. Определяем тип КОМУ_УСТУПИТЬ. Смотрим на косвенное дополнение, зависящее от прямого дополнения «уступить». Это «другим участникам движения», значит множественный тип КОМУ_УСТУПИТЬ — участник движения (подразумеваемая любого участника движения). Дополнительных свойств нет.

Разбор предложения завершён (рис. 12).

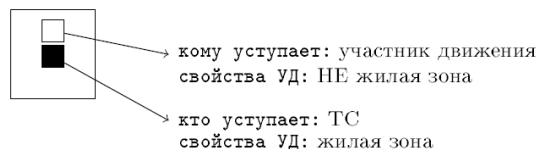


Рис. 12. Итог разбора предложения (пример 8).

8. Заключение

Данная статья является началом работы по автоматизации извлечения знаний из текста предложений (в нашем случае — построение модели правил ПДД). Разобраны предложения, содержащие манёвр «Уступить дорогу» / «Имеет преимущество» и имеющих эквивалентные синтаксические деревья (см. рис 4). Однако, по приведённому выше подходу не может быть осуществлён разбор некоторых типов предложений. А именно:

- предложений, в которых зависимые слова находятся в разных предложениях. Пример (предложения 521-522 из [1]): «Если при развороте вне перекрестка ширина проезжей части недостаточна для выполнения маневра из крайнего левого положения, его допускается производить от правого края проезжей части (с правой обочины). При этом водитель должен уступить дорогу попутным и встречным транспортным средствам». Предложений такого типа — 7.
- предложений с нестандартным (нешаблонным) порядком слов. Например, вместо синтаксического шаблона «должен уступить дорогу» используется деепричастный оборот с фразой «уступая дорогу» (предложение 796).

Дальнейшая работа будет акцентирована на разработке методов разбора этих типов предложений. Описанные в текущей статье процедуры позволяют разбирать примерно 60% предложений.

Список литературы

- [1] «Постановление Правительства РФ от 23.10.1993 N 1090 (ред. от 04.12.2018) «О правилах дорожного движения»», *КонсультантПлюс*.
- [2] Менькин М.И., «Объектная модель правил дорожного движения», *Интеллектуальные системы. Теория и приложения*, **23**:1 (2019), 39–53.
- [3] «ABBYU InfoExtractor Demo», <http://parsertest.abbyy.com/ru/>.
- [4] Перпер Е.М., Гасанов Э.Э., Кудрявцев В.Б., «О семантическом анализе юридических текстов», *Интеллектуальные системы. Теория и приложения*, **22**:3 (2018), 45–87.
- [5] «Онтология Linguistic (ABBYU)», http://comprenoservice.abbyy.com/help/ru/ontologies_linguistic_linguistic.html.

**Auto-translation of rules of the road to formal graph-theoretic
model
Menkin M.**

The present article deals with semantic analysis of legal documents. In this paper, we introduce one possible approach to formal modeling of rules of the road. We describe the basic procedures for semantic analysis of one type of sentences.

Keywords: semantic analysis of legal documents, pragmatic analysis, modeling of text semantic.