

УДК: 51-72:65

Конструктивный метод формализации текстов-инструкций

Воронов М.В.*

Московский государственный психолого-педагогический университет
(ФГБОУ ВО МГППУ), г. Москва, Российская Федерация
ORCID: <https://orcid.org/0000-0001-7839-6250>
mivoronov@yandex.ru

Пичугина Е.И.**

Московский государственный психолого-педагогический университет
(ФГБОУ ВО МГППУ), г. Москва, Российская Федерация
pichugina070698@gmail.com

Рассматривается проблема формализации текстовых данных, лежащая на стыке лингвистики, теории систем и информационных технологий. Основное внимание уделяется процессу переноса содержания, представленного в текстах-инструкциях, в формальные структуры. Предлагается метод извлечения содержания из предложений текстов-инструкций, который позволяет эффективно преобразовывать вербальную информацию в формализованные данные.

Ключевые слова: регулятивный текст, инструкция, конструктивный алгоритм, формальная структура, фрейм.

Для цитаты:

Воронов М.В., Пичугина Е.И. Конструктивный метод формализации текстов-инструкций // Моделирование и анализ данных. 2024. Том 14. № 4. С. 63–77.
DOI: <https://doi.org/10.17759/mda.2024140404>

***Воронов Михаил Владимирович**, доктор технических наук, профессор, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-7839-6250>, e-mail: mivoronov@yandex.ru

****Пичугина Елена Игоревна**, аспирант, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, e-mail: pichugina070698@gmail.com.

1. ВВЕДЕНИЕ

Основные результаты своей познавательной деятельности люди фиксируют в текстах, представляющих собой организованную определенным образом последовательность знаков, взятых из некоторого множества, например, алфавита. Основные функции текста: сохранять и передавать реципиенту зафиксированные в нем сведения. Следовательно, при манипуляциях с текстом, в том числе с использованием вычислительной техники, необходимо стремиться к сохранению его содержания, с помощью которого автор имел цель передать свои мысли. Однако, текст является лишь отображением мыслей автора, т.е. их моделью. От того, как построена эта модель и зависит уровень соответствия между мыслями автора и текстом, их отражающим. Несомненно, различные индивидуумы, после ознакомления с одним и тот же текстом (знакомясь с содержанием, представляющим собой лишь объективное отражение мыслей автора текста) будут воспринимать его не одинаково (Рис. 1).

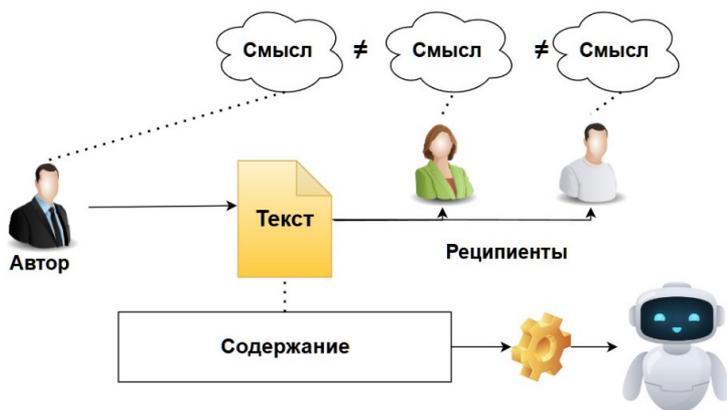


Рис. 1 Восприятие смысла текста

В отличие же от человека программный продукт реализует операции строго в соответствии с заложенным в нем алгоритмом, в основе которого лежит определенное выраженное в формальном виде содержание. В этой связи только достаточно высокий уровень адекватности содержания формализованного варианта текста его вербальному источнику обеспечивает, по крайней мере потенциально, эффективность оперирования им (содержанием) при помощи информационных технологий.

В общем случае вербально представленный текст есть упорядоченное множество слов, каждое из которых есть определенная последовательность букв. С точки зрения лингвистики буква в рамках конкретного языка представляет собой графический символ, чаще всего соответствующий определённому звуку, и, как правило, локализованная буква не несет никакого содержания. Отдельные слова, т.е. обособленно



определенное упорядоченное множество букв данного алфавита, обычно выполняют номинативную функцию, обозначая имена объектов или действий, их свойств и отношений между ними. При этом, по крайней мере в определенной предметной области, за конкретным словом закреплено определенное содержание, правда часто встречаются и многозначные слова, когда одним и тем же словом могут обозначаться разные сущности. В этой связи сформированы различного рода словари (толковые, энциклопедические, фразеологические, двуязычные и др.), которые широко используются на практике, в том числе в компьютерных системах анализа содержания текстов.

В зависимости от преследуемых целей слова в тексте определенным образом (обычно для фиксации законченной мысли, а также конкретизации и уточнения передаваемой отдельными словами информации) объединены в группы, называемыми предложениями. Для данного исследования имеет принципиальное значение следующий факт: передаваемое предложением содержание в существенной мере определяется не только входящими в него словами, но и самой структурой его построения (налицо так называемая семантическая эмерджентность). Это обстоятельство приводит к важному выводу: заключенное в предложении содержание шире, чем суммарное содержание слов предложения взятых в отдельности. Заметим, подобный эффект, правда менее выраженный, наблюдается и при формировании из совокупности предложений конкретного текста: данная совокупность предложений в зависимости от схемы ее построения может передавать разный смысл. Налицо подтверждение ставшего классическим утверждения теории систем: переход на более высокий уровень сложности приводит к появлению у объекта новых свойств, которыми не обладают его более простые версии [1]. В ряде работ лингвистов приводятся близкие по сути результаты: «понимание отдельных пропозиций детерминировано их местом, ролью и функциями в текстовом пространстве» [2, с. 50].

Можно сделать вывод, что в процессе формализации текста в интересах сохранения находящегося в нем содержания необходимо выявлять, фиксировать и сохранять структуру составляющих его предложений, а также последовательность последних и связи между ними.

2. ПОСТАНОВКА ЗАДАЧИ

Задача представления текстов в формальном виде, в максимальной степени сохраняющем содержание источника, столь сложна, что большинство авторов не без основания считают, что разработать универсальные средства формализации произвольного представленного в вербальной форме текста, при сохранении его содержания на достаточно высоком уровне, в обозримом будущем вряд ли возможно [3]. Вместе с тем представляется, что для некоторых типов текстов имеются исключения. К таковым, по нашему мнению, относятся так называемые «регулятивные тексты» – тексты выполняющие координационные (регулятивные) функции. Это распорядительные документы, инструкции, описание технологий, учебные тексты и т.п. Назначение такого рода текстов заключается в том, чтобы обеспечить реципиентам



возможность в достаточно адекватной мере воспроизводить описываемые в текстах действия. Исходя из их назначения, регулятивные тексты стремятся строить соответствующим образом: использовать принятые в данной предметной области слова и обороты речи, строить однозначно интерпретируемые фразы, строго учитывать последовательность отдельных действий и т.п.). Ряд такого рода особенностей открывает, по нашему мнению, возможность успешной формализации заданных в вербальной форме регулятивных текстов [4]. Наиболее ярко признаки регулятивности проявляются у текстов-инструкций, которые являются объектами данного исследования.

Под инструкцией обычно понимают текст, в котором излагается регламент определенной целевой деятельности [5]. Иными словами, это изложение последовательности действий, которые следует произвести.

Лингвисты рассматривают широкий спектр типов инструкций: формально-семантический, собственно прескриптивный, ограничительно-прескриптивный, наглядно-прескриптивный, вопросительно-прескриптивный и др.). Вместе с тем с позиций читающего, все инструкции это, вообще говоря, упорядоченный набор предписаний (прескрипций), т.е. побудительных высказываний об определенных действиях (или их отсутствии). По существу, в тексте-инструкции в словесной форме излагается алгоритм действий, которые реципиенту необходимо неукоснительно осуществить. Заметим, автор текста-инструкции обычно полагает, что читателю понятно, как исполнять данное конкретное (элементарное для данного текста) действие. Этот факт имеет существенное значение при решении вопросов формализации побудительных предложений исходного текста-инструкции. Если же в исходном тексте присутствуют предложения дескриптивного характера, то они носят вспомогательный характер и обычно предназначены для «неквалифицированных» субъектов.

Пусть поставлена задача: в вербальной форме задан текст-инструкция, требуется построить формальную структуру, адекватно передающую содержание исходного текста.

Примечание. Будем считать, что при построении формальной модели данного текста-инструкции располагаем лишь представленной явно словесной структурой, и только! В ряде работ отмечается, что, например, словосочетание «прилагательное + существительное» содержит скрытую предикацию [6, с. 94]. Это влечет за собой за собой предположение о наличии в тексте скрытой предикации. Заметим, именно этот факт в существенной мере определяет (в том числе) принципиальную сложность разработки универсальных средств формализации текстов. В данном же случае, при рассмотрении текстов-инструкций, предполагается отсутствие (или неучет) скрытой предикации. Иначе говоря, формализации подлежит только содержание исходного текста, а не восстановленного смысла, который пытался заложить автор в текст инструкции (иначе говоря, не следует путать понятия смысл и содержание). Содержание это означаемое собственно текстом, а термин «смысл» используют при рассмотрении результатов осознания субъектом в данном случае содержания текста [7]. В конечном счете, в рассматриваемой задаче «домысливание», трактовка представленных в предложении сведений не допустима, кроме разве что идентификация типа отношений между словами текста.



3. АНАЛИЗ СТРУКТУРЫ ТЕКСТОВ-ИНСТРУКЦИЙ

В текстах-инструкциях основное место занимают так называемые глагольные предложения, содержащие в своем составе помимо собственно глагола, показатели времени, лица и наклонения, что дает ему возможность выражать основное значение предложения – предикативность [8]. При этом предикат предложения обычно определяет и участников действия (партиципантов, акторов). Именно в этом аспекте часто утверждается, что с предикатом ассоциируется содержательный аспект [9]. Помимо партиципантов при описании ситуации часто используются важные для передачи смысла действия дополнительные признаки (время, место, темп протекания и т. д.), так называемые обстоятельственные слова (сирконстанты), используются различные формы времени и вида [10].

В рамках структурной лингвистики разрабатывается вербоцентрическая теория структурного синтаксиса [11]. Согласно этой теории, глагол в высказывании является структурно-семантическим центром предложения, чье строение определяется способностью глагола создавать лакуны, подлежащие заполнению [12]. При этом все члены предложения, дополняющие семантику глагола, одинаково важны, вследствие чего подлежащее перестает быть главным членом и переходит в разряд дополнений [13, с. 70].

Вместе с тем, еще XIX века А.А. Дмитриевский по этому поводу писал: «Не два главных члена в предложении, а только один. Сказуемое есть неограниченный властелин, царь предложения, если есть в предложении кроме него другие члены, они строго ему подчинены и от него только получают свой смысл и значение, если нет их, даже подлежащего, сказуемое само собой достаточно выражает мысль и составляет целое предложение. Иначе сказать: и само предложение есть не что иное, как сказуемое, или одно, или с приданными ему другими членами» [14, с. 22]. Будем исходить из того, что носителем предикативности всего предложения выступает сказуемое (или сочетание соответствующих слов), т.е. является ключевым признаком предложения, делающим его единицей сообщения. Все остальные члены предложения лишь конкретизируют описываемое в нем действие.

Поскольку основу текстов-инструкций составляют побудительные предложения, занятие такой позиции существенно упрощает решение поставленной задачи. Действительно, в такого рода предложениях содержательную суть обычно несет императивная глагольная словоформа (подключить, снять, перевести и т.д.), остальные члены предложения лишь конкретизируют данное действие и условия его осуществления. Это содержательное ядро такого предложения, отталкиваясь от которого можно сформировать его формальное представление. Последнее удобно представлять в виде специфический граф, типа «гроздь» [15].

Предметом нашего рассмотрения является процесс формализации представленного в тексте-инструкции содержания. В процессе формализации должен осуществляться перенос имеющегося в вербальном тексте содержания в содержание формируемых формальных структур.



Содержание вербального текста передается и словами, и их взаимным расположением, т.е. его структурой. Особенности именно текстов-инструкций позволяют принять следующее условие: слова текста рассматриваются в качестве данных элементов (ими оперирую, как целым). В этом случае основные усилия сосредотачиваются в отображении отношений, в которых находятся сущности, обозначаемые словами текста.

Итак, пусть задан представленный в вербальной форме текст-инструкция. Требуется разработать метод построения его формализованного аналога, обеспечивающего возможность оперировать содержанием исходного текста как математической структурой.

Примечание. В такой постановке термин «формализация» использован скорее по традиции. В данном контексте он означает процесс перестроения структуры данного текста, но так, чтобы новое представление, сохраняя всю информацию, заложенную в содержание исходного варианта, отвечало возможности адекватного оперирования ею при разработке соответствующих программных продуктов.

4. АЛГОРИТМ ФОРМАЛИЗАЦИИ ПРЕДЛОЖЕНИЙ ТЕКСТА-ИНСТРУКЦИИ

Представляется целесообразным решение поставленной задачи осуществить в виде разработки конструктивного процесса построения соответствующей тексту семантической сети.

Напомним, процесс является конструктивным (в математическом аспекте), когда:

- Задан алфавит исходных объектов, рассматриваемых как нерасчленимые;
- Задан список правил образования новых объектов из ранее построенных;
- Процессы построения новых объектов осуществляются имеющими точную характеристику отдельными шагами, причем такими, что обеспечена их «нетрудная» реализация.» [16, с. 1058].

Характерными примерами конструктивных процессов является процесс построение слов из букв данного алфавита. Действительно, автор имеет некоторое множество исходных элементов (букв и ряд знаков препинания). Из них он формирует слова, которые образуются путем приписывания очередного подходящего знака (в том числе и пропуска).

В данном случае исходный алфавит образуют слова текста, т.е. каждое слово воспринимается как имя соответствующего понятия и в дальнейшем им оперируют, как элементом. Это предположение базируется на том, что автор исходного текста-инструкции полагает: реципиенту понятен смысл каждого использованного термина, а также и то, как изменится состояние участвующих в действии объектов после окончания определенного действия. Например, пусть в тексте есть предложение: «вернуть кран K_j на два деления влево». Это означает, что понятно о каком кране идет речь и известно (по построению) в каком конкретно состоянии он находится, например, в состоянии $SK7_j$. По окончании данного действия i этот кран будет находиться в другом состоянии, например, $SK7_j(j=i+1)$.

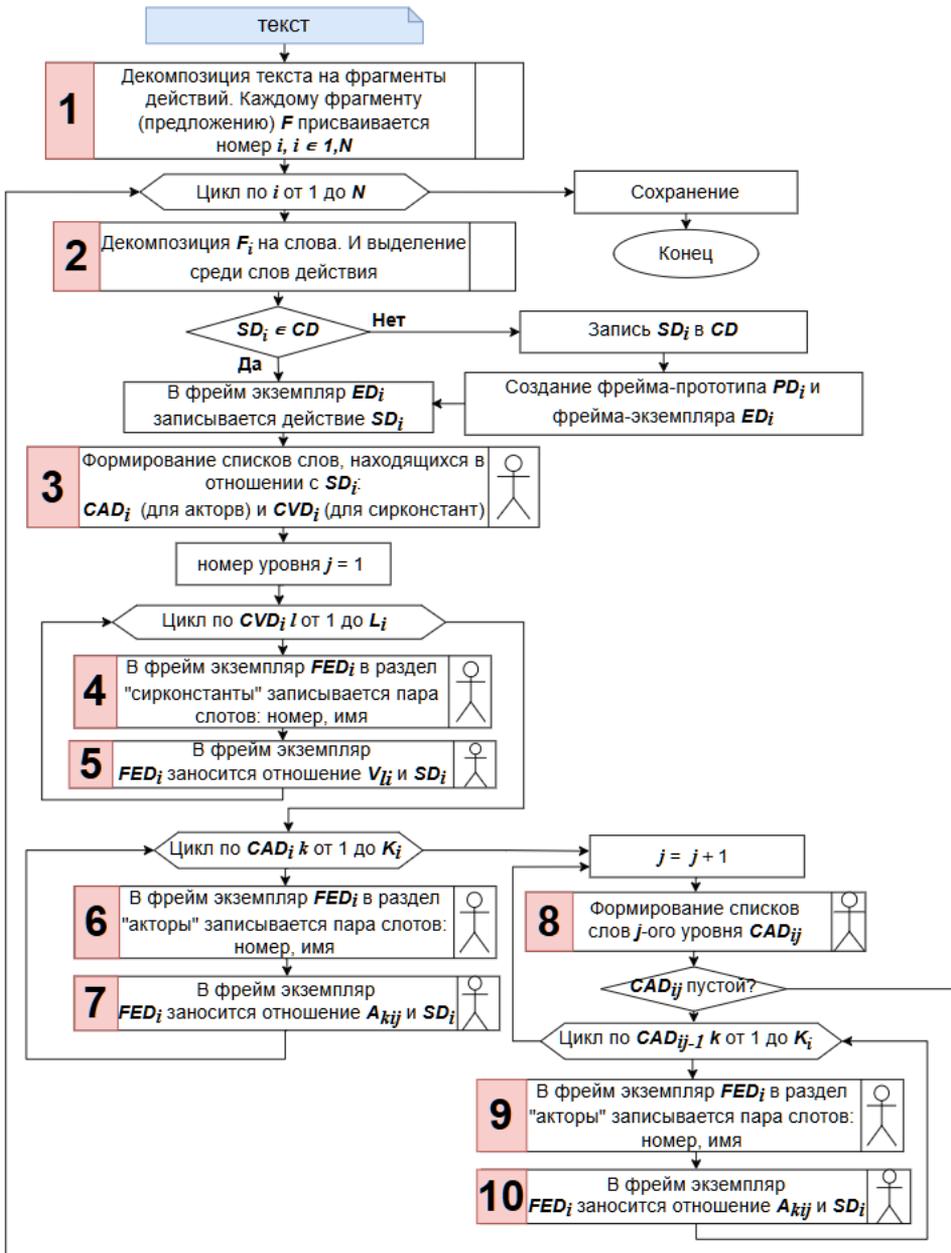


Рис. 2 Блок-схема алгоритма



Обобщенный алгоритм решения поставленной задачи представлен на рис. 2.

F – фрагмент текста (предложение);

i – номер предложения;

S – слово в предложении F_i ;

D_i – действие в предложении i ;

SD – слово (оборот), описывающее действие;

CD – список слов, описывающих действия;

PD_i – фрейм-прототип действия в предложении i ;

ED_i – фрейм-экземпляр действия в предложении i ;

j – номер уровня окружения слов данного предложения;

$CVD_i = \{V_{li}\}$, $l \in [1, L_i]$, список слов, описывающих обстоятельства реализации данного действия;

$CAD_i = \{A_{ki}\}$, $k \in [1, K_i]$, список слов, обозначающих предметы (акторов), непосредственно участвующие в данном действии;

FED_i – фрейм-экземпляр для действия D_i .

OVD_i – список отношений, в которых сирконстанты из $\{V_{li}\}$, находятся с данным действием;

OV_{li} – отношение из списка OVD_i под номером l , в котором этот сирконстант находится с данным действием;

OA_{ki} – имя актора из списка CAD_i под номером k , в котором этот актор находится с данным действием.

Изложим ряд положений, раскрывающих особенности излагаемого подхода решения поставленной задачи.

1. Любой конструктивный процесс осуществляется по шагам [17]. В данном случае реализуются шаги разных уровней. Основным шагом, назовем его внешним, является построение формализма, отображающего содержание отдельного предложения исходного текста. В ходе его осуществления реализуется множество так называемых внутренних шагов. На каждом внутреннем шаге решается «судьба» каждого из составляющих предложения слов. При этом фиксируется отдельное слово и выявляются отношения, в которых соответствующая сущность находится с другими объектами, представленными в данном предложении. Все такого рода сведения определенным образом фиксируются.
2. Предложение текста отображается в виде отдельной подсети (ориентированного графа), характерная структура которой представлена на рис. 3.

Ее образуют следующие компоненты: «ядро» соответствует слову, описывающему в предложении действие, слова, описывающие имена «входных» объектов (непосредственных участников этого действия), а также их состояние на начало действия, слова описывающие имена «выходных» объектов – всех задействованных (и вновь сформированных) в действии объектов, а также соответствующие сирконстанты, значения которых отвечает состоянию всех компонентов на момент окончания данного действия. Таким образом, граф, отображающий структуру предложения представляет собой «гроздь» весьма своеобразного вида: одна группа дуг входит, а другая

выходит из ядра, причем входные дуги в свою очередь составляют два подмножества: одни выходят из акторов, другие из сирконстант, но все они входят в «ядро», отображающее в данном предложении действие.

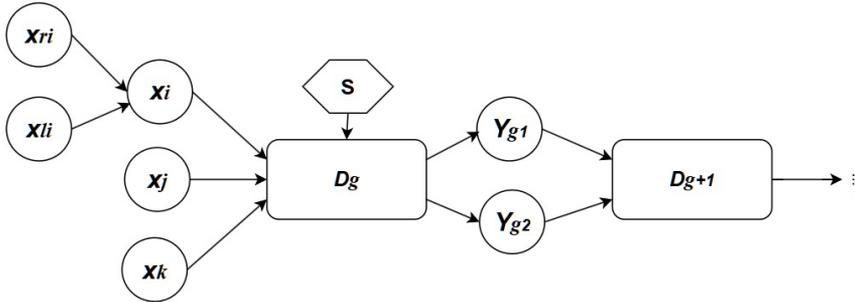


Рис. 3 Отображение структуры отдельного предложения

3. Процедура формализации очередного предложения начинается с выявления сказуемого (предикатного слова или оборота), описывающего собственно действие, о котором идет речь в данном предложении, что является ключевым моментом представляемого подхода. Предикатное слово и является тем первым словом, которое отображается «центральной» вершиной ориентированного графа, представляющего впоследствии структуру рассматриваемого предложения (см. Рис. 3). Отталкиваясь от него осуществляются все последующие (внутренние) шаги, приводящие к формированию остальных компонентов формируемого графа, ибо помимо собственно действия в предложении обычно описаны и отношения данного действия к ряду объектов (акторов действия), а также сирконстанты, характеризующие обстоятельства описываемого действия. Продолжая рассматривать описанный выше пример, отмечаем: исходное слово (обозначение действия) «повернуть», повернуть что? – «кран К7», куда? – «влево», на сколько? – «на два деления».
4. В результате формализации исходный текст представляется в виде семантической сети, элементами которой выступают соответствующие фреймы. Каждый из них несет в себе сведения о представленных в тексте сущностях, а также связях с другими сущностями, отражая динамику изменения их состояния.
5. Формирование графа, отображающего данное предложение, осуществляется по схеме «в ширь». В данном случае это означает, что для каждого рассматриваемого слова выявляются слова, находящиеся с ним только в прямом (непосредственном) отношении. Как уже отмечено, вначале выявляется предикатное слово предложения и формируется множество слов, которые находятся в непосредственном отношении со сказуемым. Это так называемые семантические актанты, представляющие собой акторов, объекты или инструменты действия, а также обстоятельства их участия [18]. Особенность этих слов в том, что они

обозначают только те предметы, которые находятся с данным предикатом в непосредственных отношениях. Будем говорить, что эти слова, т.е. подлежащее и дополнения образуют ярус первого уровня (заметим, ярус первого уровня единственный). Каждому из этих слов могут быть поставлены в соответствие зависимые слова, которые образуют второй ярус. Компоненты последнего также могут содержать зависимые слова и т.д. (ярусов не первого уровня может быть несколько). Важно подчеркнуть, что в каждый данный ярус включаются только слова, обозначающие сущности, находящиеся с образующим данный ярус словом в непосредственном отношении (первый ярус образует предикат предложения, другие ярусы образуют акторы действий). (Рис. 4)

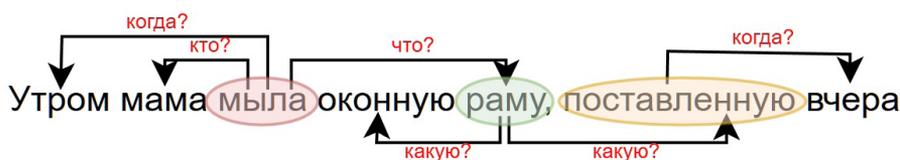


Рис. 4 Пример графа, построенного по предложению

- В текстах встречаются сложные предложения – предложения, составленные из простых предложений (простое предложение содержат одно сказуемое). Если предложение сложносочинённое, т.е. в нем несколько простых предложений, соединенных сочинительными союзами, то все простые предложения рассматриваются и формализуются последовательно, как отдельные предложения исходного текста. Если же предложение сложноподчинённое, т.е. содержит простые предложения, соединенные подчинительными союзами, то они формализуются по описанной выше схеме формирования соответствующего яруса.
- Все результаты осуществления шагов, реализованных в ходе рассмотрения данного предложения, фиксируются в соответствующих фреймах. Формируется два множества фреймов. Первый образуют фреймы, где зафиксирована структура описания рассматриваемого действия (указаны участники данного действия и их отношения к описываемому действию). Во фреймах второго множества отражается история изменения состояния упомянутых в тексте объектов. Для этого в них фиксируется состояние объекта на конец описываемого действия (предыдущие состояния были зафиксированы на предшествующих шагах или заданы в качестве исходных). Правила отображения результатов каждого шага формализации обуславливают принятые схемы формирования соответствующих фреймов (рис. 5).
- Важно рассматривать участвующие в данном действии объекты в состоянии, в котором они находятся именно на начало описываемого в предложении действия. Это условие обеспечивается тем, что поскольку в текстах-инструкциях важна последовательность описываемых действий, то на начало очередного действия, в котором участвует данный объект, он находится в актуальном состоянии. В этой связи состоянию каждого актора на момент окончания данного действия

присваивается номер рассматриваемого предложения. Следовательно, номера предложений, номера действия и состояния участвующих в действиях объектов находятся в однозначно идентифицируемой связи. Это обстоятельство позволяет автоматизировать процесс синтеза текста в целом из построенных формализмов его отдельных предложений [19].

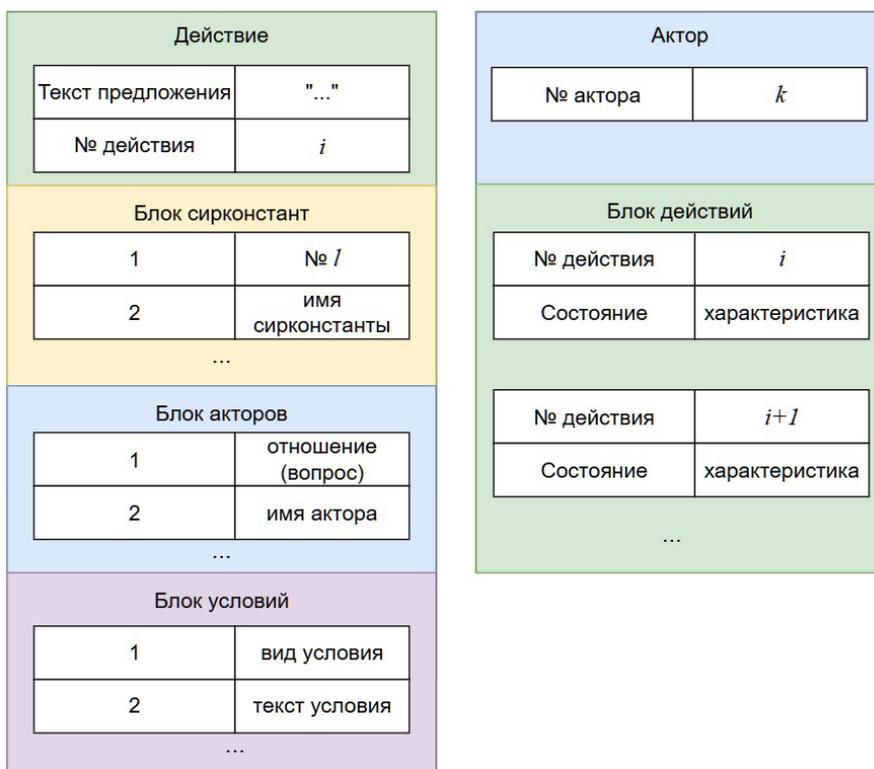


Рис. 5 Структура фреймов

5. ЗАКЛЮЧЕНИЕ

Предложенный метод формализации текстов, представленных в виде инструкций, основывается на том, что они состоят из определенной последовательности побудительных предложений. Это обеспечивает возможность утверждать, что в основе структуры каждого предложения исходного текста-инструкции лежит его предикат, а остальные слова предложения лишь раскрывают, уточняют описываемое действие.

Изложенный метод представляет собой конкретную форму реализации конструктивного процесса, что обеспечивает возможность решения поставленной задачи:



построить пошаговый, допускающий четкую реализацию автоматизированного процесса, в ходе которого формируется формальная модель заданного в вербальной форме исходного текста-инструкции. Эта модель представляет собой семантическую сеть, связывающую между собой сформированную совокупность заполненных фреймов.

Наличие формальных моделей таких текстов обуславливает возможности решения широкого круга задач анализа и синтеза пакетов документации конкретной предметной области.

Литература

1. *Anderson P.W.* More Is Different Science 1972, no. 177(4047), pp. 393–396.
2. Серебрякова С.В., Милостивая.И. Семантическая эмерджентность как переводческая проблема // Вестник ВолГУ. Серия 2, Языкознание. 2017. Т. 16. № 3. С. 48–57.
3. *Walter Benedikt, Hammes Jakob, Piechotta Marco, Rudolph Stephan.* A Formalization Method to Process Structured Natural Language to Logic Expressions to Detect Redundant Specification and Test Statements // IEEE 25th International Requirements Engineering Conference (RE) – Lisbon. 2017. pp. 263–272.
4. *Воронов М.В., Зотов Д.Р., Кремков М.В.* Автоматизация процессов формализации технологий // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: естественные и технические науки. 2022, № 4. С. 5–10.
5. Большой энциклопедический словарь: В 2 т. / Гл. ред. А.М. Прохоров. Т. 1. М., 1991.
6. *Сандакова М.В.* Прилагательное в позиции атрибута и предиката // Вестник Вятского государственного университета. Филологические науки, 2007, № 16. С. 94–100.
7. *Новиков А.И.* Текст и его смысловые доминанты. М.: Ин-т языкознания РАН, 2007. 224 с.
8. *Пащенко Ю.А.* Предикативность и предикат в лингвистике и логике // Вестник ТГПИ Гуманитарные науки С. 70–72. [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/v/predikativnost-i-predikat-v-lingvistike-i-logike> (дата обращения: 23.08.2024).
9. *Арутюнова Н.Д.* Сокровенная связка (к проблеме предикативного отношения) / Изв. АН СССР, 1980. Т. 39. № 4. С. 347–358.
10. *Немченко В.Н.* Введение в языкознание: учебник для вузов / В.Н. Немченко. – 2-е изд., перераб. и доп. – М.: Издательство Юрайт, 2015. – 679 с.
11. *Zakharov Victor, Krassovitskiy Alexander, Meirambekkyzy Zhazirakhanym, Ualiyeva Irina, Khoroshilov Aleksander, Khoroshilov Alexey.* Automatic Creation Technologies of Declarative Tools for Clustering Media Documents // International Conference on Engineering Technologies and Computer Science (EnT). 2019. pp. 39–42.
12. *Теньер Л.* Основы структурного синтаксиса. М.: Прогресс, 1988. 656 с.
13. *Макаренко Н.Л.* Вербоцентрическая теория синтаксической структуры высказывания в курсе РКИ и современного русского языка / Материалы VII Республиканского научно-практического семинара Лингводидактика – Минск, МГПУ, 2020. С. 66–75.
14. *Дмитриевский А.А.* Практические заметки о русском синтаксисе // Филологические записки. IV. 1877. С. 3–38.
15. *Sihole Vusi, Marshal Linda.* Building Concept Hierarchies for the Internet of Things Patterns Using Domain-specific Dependency Knowledge // Open Innovations (OI). 2019. pp. 1–7.
16. Математическая энциклопедия: гл. редактор И.М. Виноградов – Советская энциклопедия, т. 2, 1979. 1104 с.
17. *Воронов М.В.* Конструктивно-имитационное моделирование слабоструктурированных систем. // Известия МАН ВШ. 2007, № 4(42), С. 156–165.



18. Батура Т.В. Семантический анализ и способы представления смысла текста в компьютерной лингвистике // Программные системы и продукты, 2014. № 4, С. 45–57.
19. Воронов М.В., Пименов В.И. Формализация регулятивных текстов // Информатика и автоматизация, 2021, Том 20, № 3. С. 562–589.



Constructive Method of Formalization of Instruction Texts

Mikhail V. Voronov*

Moscow State University of Psychology and Education (MSUPE), Moscow, Russia

ORCID: <https://orcid.org/0000-0001-7839-6250>

e-mail: mivoronov@yandex.ru

Elena I. Pichugina**

Moscow State University of Psychology and Education (MSUPE), Moscow, Russia

e-mail: pichugina070698@gmail.com

The problem of formalization of text data, which lies at the intersection of linguistics, systems theory and information technology, is considered. The focus is on the process of translating the content presented in the instructions into formal structures. The method proposed is to extract content from the sentence of the instruction text, which allows for an effective conversion of verbal information into formalized data.

Keywords: regulatory text, instruction, constructive algorithm, formal structure, frame.

For citation:

Voronov M.V., Pichugina E.I. Constructive Method of Formalization of Instruction Texts. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2024. Vol. 14, no. 4, pp. 63–77. DOI: <https://doi.org/10.17759/mda.2024140404> (In Russ., abstr. in Engl.).

References

1. Anderson P.W. More Is Different. *Science*, 1972, No. 177 (4047), pp. 393–396.
2. Serebryakova S.V., Milostivaya A.I. Semanticheskaya emergentnost' kak perevodcheskaya problema. *Vestnik VolGU. Seriya 2, Yazykovedenie*. 2017. T. 16. No. 3. pp. 48–57.
3. Walter Benedikt, Hammes Jakob, Piechotta Marco, Rudolph Stephan. A Formalization Method to Process Structured Natural Language to Logic Expressions to Detect Redundant Specification and Test Statements. *IEEE 25th International Requirements Engineering Conference (RE) – Lisbon*. 2017. pp. 263–272.
4. Voronov M.V., Zotov D.R., Kremkov M.V. Avtomatizatsiya protsessov formalizatsii tekhnologiy. *Vestnik Sankt-Peterburgskogo Gosudarstvennogo Universiteta Tekhnologii i Dizaina. Seriya 1: Estestvennye i Tekhnicheskie Nauki*. 2022, No. 4. pp. 5–10.
5. Bol'shoy entsiklopedicheskiy slovar': V 2 t. Glavnyy redaktor A.M. Prokhorov. T. 1. Moscow, 1991.

***Mikhail V. Voronov**, Doctor of Technical Sciences, Professor, Moscow State University of Psychology and Education (MSUPE), Moscow, Russia, ORCID: <https://orcid.org/0000-0001-7839-6250>, mivoronov@yandex.ru

****Elena I. Pichugina**, Graduate Student, Moscow State University of Psychology and Education (MSUPE), Moscow, Russia, pichugina070698@gmail.com



6. Sandakova M.V. Prilagatel'noe v pozitsii atributa i predikata. Vestnik Vyatskogo Gosudarstvennogo Gumanitarnogo Universiteta Filologicheskie Nauki. 2007, No. 16, pp. 94–100.
7. Novikov A.I. Tekst i ego smyslovye dominanty. Moscow: In-t Yazykovedeniya RAN. 2007, 224 p.
8. Pashchenko Yu.A. Predikativnost' i predikat v lingvistike i logike. Vestnik TGPi Gumanitarnye Nauki, pp. 70–72.
9. Arutyunova N.D., Sokrovennaya svyazka: K probleme predikativnogo otnosheniya. Izvestiya AN SSSR, Seriya LiyA. 1980, T. 39, No 4, pp. 347–358.
10. Nemchenko V.N. Vvedenie v yazykovedenie: uchebnyk dlya vuzov / V.N. Nemchenko – 2-e izd., pererab., i dop – Moscow: Izdatel'stvo Yurayt, pp. 2015–679.
11. Zakharov Victor, Krassovitskiy Alexander, Meirambekkyzy Zhazirakhanym, Ualiyeva Irina, Khoroshilov Aleksander, Khoroshilov Alexey. Automatic Creation Technologies of Declarative Tools for Clustering Media Documents. International Conference on Engineering Technologies and Computer Science (EnT). 2019, pp. 39–42.
12. Tenyer L. Osnovy strukturnogo sintaksisa. Moscow: Progress. 1988, 656 p.
13. Makarenko N.L. Verbotsestricheskaya teoriya sintaksicheskoy struktury vyskazyvaniya v kurse RKI i sovremennogo russkogo yazyka. Materialy VII Respublikan'skogo nauchno-prakticheskogo seminaru Lingvodidaktika – Minsk, MGPU, 2020, pp. 66–75.
14. Dmitrievskiy A.A., Prakticheskie zametki o russkom sintaksise. Filologicheskie zapiski IV, 1877, pp. 3–38.
15. Sithole Vusi, Marshal Linda. Building Concept Hierarchies for the Internet of Things Patterns Using Domain-specific Dependency Knowledge. Open Innovations (OI). 2019, pp. 1–7.
16. Matematicheskaya entsiklopediya: glavnyy redaktor I.M. Vinogradov, t.v – Sovetskaya entsiklopediya, pp. 1979–1104.
17. Voronov M.V., Konstruktsionno-imitatsionnoe modelirovanie slabostrukturirovannykh sistem. Izvestiya MAN VSh № 4(42), 2007 g., pp.156–165.
18. Baturova T.V., Semanticheskii analiz i sposoby predstavleniya smysla teksta v komp'yuternoy lingvistike. Programmye sistemy i produkty. 2014, No4, pp. 45–57.
19. Voronov M.V., Pimenov V.I., Formalizatsiya regulativnykh tekstov. Informatika i Avtomatizatsiya. 2021, Tom 20, No3, pp. 562–589.

Получена 06.11.2024

Received 06.11.2024

Принята в печать 02.12.2024

Accepted 02.12.2024