

Сергей Николаевич Сергеев

АВТОМАТИЗИРОВАННАЯ СИСТЕМА
КЛАССИФИКАЦИИ СОДЕРЖАНИЯ
ЮРИДИЧЕСКО-ПРАВОВОЙ ДОКУМЕНТАЦИИ

представлена на соискание академической степени
доктора

Грузинский Технический Университет

Тбилиси, 0175, Грузия

Ноябрь, 2008 год

Авторские права © 2008 "Сергей Николаевич Сергеев"

საქართველოს ტექნიკური უნივერსიტეტი

”ინფორმატიკისა და მართვის სისტემების ფაკულტეტი”

ჩვენ, ქვემოთ ხელისმომწერნი ვადასტურებთ, რომ გავეცანით ”სერგო ნიკოლოზის-ძე სერგვეს” მიერ შესრულებულ სადისერტაციო ნაშრომს დასახელებით: ”იურიდიულ-სამართლებრივი დოკუმენტაციის შინაარსის კლასიფიკაციის ავტომატიზებული სისტემა” და ვაძლევთ რეკომენდაციას საქართველოს ტექნიკური უნივერსიტეტის ”ინფორმატიკისა და მართვის სისტემების ფაკულტეტის” სადისერტაციო საბჭოში მის განხილვას დოქტორის აკადემიური ხარისხის მოსაპოვებლად.

24.11.2008

ხელმძღვანელი: პროფ.ტ.მ.დ. ლევან ჩხაიძე

რეცენზენტი: პროფ.ტ.მ.დ. ზურაბ გასიტაშვილი

რეცენზენტი: პროფ.,ფ.მ.დ. რუსუდან ასათიანი

საქართველოს ტექნიკური უნივერსიტეტი

2008

ავტორი: სერგო ნიკოლოზის-ძე სერგეევი
დასახელება: ”იურიდიულ-სამართლებრივი დოკუმენტაციის
შინაარსის კლასიფიკაციის ავტომატიზებული
სისტემა”
ფაკულტეტი: ”ინფორმატიკისა და მართვის სისტემების ფაკულტეტი”
აკადემიური ხარისხი: დოქტორი
სხდომა ჩატარდა:

ინდივიდუალური პიროვნებების ან ინსტიტუტების მიერ
ზემომოყვანილი დასახელების დისერტაციის გაცნობის მიზნით მოთხოვნის
შემთხვევაში მისი არაკომერციული მიზნებით კოპირებისა და გავრცელების
უფლება მინიჭებული აქვს საქართველოს ტექნიკურ უნივერსიტეტს.

ავტორის ხელმოწერა

ავტორი ინარჩუნებს დანარჩენ საგამომცემლო უფლებებს და არც
მთლიანი ნაშრომის და არც მისი ცალკეული კომპონენტების გადაბეჭდვა ან
სხვა რაიმე მეთოდით რეპროდუქცია დაუშვებელია ავტორის წერილობითი
ნებართვის გარეშე.

ავტორი ირწმუნება, რომ ნაშრომში გამოყენებული საავტორო
უფლებებით დაცული მასალებზე მიღებულია შესაბამისი ნებართვა (გარდა
იმ მცირე ზომის ციტატებისა, რომლებიც მოითხოვენ მხოლოდ სპეციფიურ
მიმართებას ლიტერატურის ციტირებაში, როგორც ეს მიღებულია
სამეცნიერო ნაშრომების შესრულებისას) და ყველა მათგანზე იღებს
პასუხისმგებლობას.

დისერტაციას დიდი სიყვარულით ვუძღვნი ჩემს
ოჯახს და მეგობრებს.

РЕЗЮМЕ

В диссертационной работе «Автоматизированная система классификации юридически-правовой документации» рассматривается одна из важных задач интеллектуальной обработки данных – построение формальных моделей смыслового значения текстов на естественном языке с целью их автоматической классификации и решения поисковых задач. Проблема решается на примере автоматизации судебного делопроизводства по гражданским делам для текстов формулировок содержания гражданских исков, записанных на ограниченном естественном грузинском языке.

Прикладная ценность диссертационного исследования заключается в необходимости преодоления объективных и субъективных трудностей, вызванных особенностями традиционного уровня формализации создаваемых в процессе судопроизводства документов, в частности, наличием в этих документах ряда реквизитов, классифицирующих их формальное содержание, но в традиционной технологии ведения судебного делопроизводства формируемых «вручную» сотрудниками судебных органов. Столь значительная роль человеческого фактора в процессе формализации ответственного документа, определяющего его поисковый образ, неизбежно приводит к достаточно многочисленным ошибкам в процессе ручного кодирования, что неизбежно сказывается на работе судебных органов, препятствуя их объективности и независимости, и в результате приводит к негативному социальному эффекту от недостаточного уровня их работы по сравнению с запросами общества.

В диссертационной работе рассматривается конкретная задача автоматизации судебного делопроизводства – сопоставление встречающихся в одном и том же документе формулировок содержания гражданского иска на естественном грузинском языке и их кодового образа, выбранного сотрудником суда из утвержденных классификаторов содержания судебных исков по гражданским делам.

Как научная проблема, задача проверки взаимного соответствия их содержания формулируется в диссертации следующим образом. Даны два текста на естественном грузинском языке, ограниченном предметной областью судебного делопроизводства по гражданским делам. Первый текст является формулировкой содержания конкретного гражданского иска. Вторым текстом является текст одной из статей принятого в судопроизводстве Грузии кодификатора форм статистической отчетности

Задача заключается в построении формализованной модели указанных двух текстов, позволяющей автоматически определить формальные образы их содержания и с помощью вычисляемой процедуры специального вида получить строгое определение их релевантности. При положительном решении задачи заданный в классификаторе код статьи кодификатора объявляется полностью формализованным образом содержания данного гражданского иска и в дальнейшем используется автоматизированной системой для решения классификационных, статистических и поисковых задач.

Предлагаемый в диссертационной работе подход к решению задачи состоит из ряда последовательных этапов, на которых производится иерархический анализ текста на ограниченном естественном грузинском языке. Основным ограничением языка предметной области является синтагматическая структура текста формулировки содержания гражданского иска, т.е. отсутствие в нем личных глагольных форм (в нем могут быть только масдары, существительные, прилагательные и наречия, а также некоторые вспомогательные слова). В диссертационной работе выделяются следующие алгоритмические этапы решения поставленной задачи:

- морфологический анализ текста на базе его лексического состава, определяемого словарями ограниченного языка предметной области;
- синтаксический анализ текста с учетом ограничений языка предметной области и эвристических правил литературного грузинского языка относительно взаимного расположения членов предложения;
- семантический анализ текста, выявление в нем основных семантических сущностей (актера, агента, адвербов, атрибутов) и его приведение к унифицированным причинно-квалитативным связям между выявленными сущностями, абстагированными от грамматических понятий;

– анализ возможной релевантности формальных образов двух гражданских исков путем сравнения полученных списковых структур, с использованием эвристических правил игнорирования несущественных адвербиативных и атрибутивных свойств главных сущностей (актера, агента), а также эвристических правил установления семантической эквивалентности близких по содержанию сущностей

Каждый из рассмотренных этапов осуществляется с помощью разработанного в диссертационной работе программного обеспечения на языке «Delphi», последовательно преобразующего текст на естественном языке в списковые структуры специального вида.

Важной стороной синтаксического и семантического анализа является использование современного подхода «разметки текстов» для разрешения их синтаксической и семантической неоднозначности. Для рассматриваемой предметной области построение системы разрешения неоднозначности существенным образом опирается на грамматические особенности грузинского языка и на ограничения, присущие языку конкретной предметной области формулировок содержания гражданского иска. Реализация системы разрешения неоднозначности основана на принципе генерации и отбора синтаксических, лексических и семантических гипотез путем интерактивно-диалогового взаимодействия с квалифицированным экспертом, а также путем использования экспертных знаний о текстовых корпусах судебной документации.

В результате полного решения задачи достигается принципиальное сокращение роли субъективного фактора в судебном делопроизводстве и соответственно повышение объективности судебного процесса, что в значительной степени повысит социальное и психологическое доверие общества к судебной системе за счет внедрения совершенных и наукоемких средств автоматизации.

ABSTRACT

The dissertational theses titled «Automated system for juridical documentation contents classification» deals with one of most important problems of intellectual data processing — devising formalized models of text meanings in a natural language with purpose of further automated classification and solving textual search problems. The problem is resolved using an example of automating civil lawsuit record keeping for texts formulating the contents of civil lawsuits written in limited natural Georgian language.

The practical value of the research presented lies in the need for overcoming objective and subjective difficulties caused by the specifics of traditional level for formalizing the documents created in process of civil lawsuit record keeping, including in particular specific requisites classifying the formal contents of lawsuit, which are added «manually» by legal workers in traditional lawsuit record keeping. This significant role of human factor in process of formalizing an important document which determines document's image for search processes, leads to an essential number of errors in manual encoding, which is bound to affect the effectiveness and independency of court bodies, and as a result leads to negative social effects from insufficient courts functionality compared to society's needs.

In the dissertational theses a specific task of automating lawsuit record-keeping is reviewed — comparing the phrasings in natural Georgian language found in a specific document to their coded image selected by a court worker, found within approved civil lawsuits contents' classifiers.

As a research problem, the task of checking mutual correspondence of their content is formulated in theses as follows. Given two texts in natural Georgian language, limited by subject of civil lawsuit record keeping. The first text is a formulated abstract of content of a specific civil lawsuit. Second text is one of the articles of an article from statistical accounts form codifier accepted in Georgian law practice.

The problem is in devising a formalized model of two texts specified, which would allow to automate the formal images of these texts' contents and obtain their strict relevancy definition using specifically formatted computational procedure. If the answer is positive, the codifier article code in the classifier is considered the fully

formalized image of given civil lawsuit content, and is further used by automated system to perform classification, statistical and search tasks.

The approach to solving the problem suggested in the theses consists of several consequential stages of hierarchical analysis of text in limited natural Georgian language. The main limitation of subject field in the language is the syntagmic structure of civil lawsuit formulae, i.e. absence of personal verbal forms (it may contain only *masdari*, nouns, adjectives and adverbs, and also some auxiliary words). The following algorithmical stages to solution of task formulated are presented in the theses:

- text morphological analysis based on its lexical contents determined by vocabularies of subject field limited language;
- text syntax analysis taking into account the limitations of language subject field and heuristic rules of literary Georgian language related to relative arrangement of sentence parts;
- text semantical analysis, locating principal semantic substances (actor, agent, adverbs, attributes) and bringing it to the form of unified reasonal-qualitative connections between substances found, having grammar concepts put aside;
- analysis of possible relevancy of formalized images of two civil lawsuits via comparing the resulting listings, using the heuristic rules of ignoring non-existent adverbial and attributive properties of principal substances (actor, agent) and also heuristic rules of devising semantic equivalence of substances close in content.

Each of the stages described is performed using software in Delphi developed in process of research for the dissertational theses, which gradually transforms a text in natural language into specifically formatted listings.

Using the modern approach of text formatting is an important part of syntax and semantic analysis to resolve texts' syntax and semantic ambiguity. For the subject field in question devising a system for resolving ambiguity substantially relies on Georgian language grammatic peculiarities and limitations specific for civil lawsuit formulae language. The implementation of system for resolving ambiguities is based on generation and selection of syntax, lexical and symantical assumptions via dialogue interaction with a qualified expert in the field, and also via using expert knowledge on lawsuit documentation text massives.

As a result of fully solving the problem a principal reduction of subjective factor influence in judicial record keeping can be achieved, and correspondingly —

increased efficiency and objectiveness of court procedure, which would significantly increase social and psychological trust towards the judiciary system due to implementing modern automated techniques.

Содержание

0. Введение.	17
0.1. Научная новизна поставленной в диссертации проблемы.	17
0.2. Актуальность и практическое значение проблемы.	18
0.3. Научные особенности решения поставленной в диссертации проблемы.	19
0.4. Краткая постановка проблемы, решению которой посвящена диссертационная работа.	20
Глава 1. Постановка задачи создания формализованных моделей семантики юридически-правовой документации на естественном грузинском языке (на примере юридических формул гражданского иска) и использование аппарата грамматической разметки текстов для ее решения.	22
1.1. Актуальность поставленной в диссертации проблемы.	22
1.2. Научная новизна поставленной в диссертации проблемы.	25
1.3. Концептуальные основы формальной модели синтаксиса текста гражданского иска на ограниченном естественном грузинском языке.	27
1.4. Концептуальные основы формальной модели семантики текста гражданского иска на ограниченном естественном грузинском языке.	30
1.5. Постановка задачи автоматизированной классификации, релевантного поиска и релевантного сравнения текстов в принятых формальной модели семантики текста гражданского иска на ограниченном естественном грузинском языке.	35
Глава 2. Обзор методов морфологической, синтаксической и семантической разметки текстовых корпусов на естественных языках.	38
2.1. Обзор текстовых корпусов, имеющих морфологическую, синтаксическую и семантическую разметку (на примере национального корпуса русского языка)	38
2.2. Обзор проблемы разрешения синтаксической неоднозначности при морфологической, синтаксической и семантической разметке текстов на естественных языках и методы ее интерактивного разрешения.	59
Глава 3. Лингвистические основы формального описания юридически-правовой документации на ограниченном естественном грузинском языке ..	63
3.1. Формальное описание флексивного словоизменения существительных грузинского языка.	63
3.2. Формальное описание синтаксической структуры синтагмы грузинского языка.	74
Глава 4. Экспериментальное исследование системы описания семантики юридических формул гражданского иска	87
4.1. База экспериментального исследования семантики юридических формул гражданского иска.	87

4.2. Определение лексического состава текстов юридических формул гражданского иска	89
4.3. Полный морфологический анализ текстов юридических формул гражданского иска.	90
4.4. Уточненная морфологическая классификация основ, выделенных из текстов юридических формул гражданского иска.	93
4.5. Выделение образующих элементов семантико-синтаксической структуры юридических формул гражданского иска – акторов.	96
4.6. Выделение в тексте юридических формул гражданского иска возможных маркеров разделения на семантически условно-независимые подформулы. ..	97
4.7. Разделение текста юридических формул гражданского иска на две части с помощью выделенных маркеров.	99
4.8. Маркирование прямого дополнения в тексте юридических формул гражданского иска, на которое непосредственно распространяется действие актора.	100
4.9. Маркирование косвенных дополнений в тексте юридических формул гражданского иска, уточняющих семантику прямого дополнения, на которое косвенно распространяется действие актора.	102
4.10. Маркирование атрибутов в тексте юридических формул гражданского иска, описывающих атрибутивные свойства прямого дополнения, на которые также косвенно распространяется действие актора.	103
4.11. Маркирование адвербов или адвербиативных конструкций (обстоятельств или их списков) в тексте юридических формул гражданского иска, описывающих условия или обстоятельства применения действия, представленного актором, к прямому дополнению, на которые также косвенно распространяется действие актора.	105
4.12. Маркирование атрибутов внутри адвербиативных конструкций (обстоятельств или их списков) в тексте юридических формул гражданского иска, описывающих атрибутивные свойства, относящиеся непосредственно к адвербу.	107
4.13. Построение грамматически-независимого представления текстов юридических формул гражданского иска.	109
4.14. Маркирование ключевых семантических элементов формализованного представления содержания текстов юридических формул гражданского иска (актора и агента).	116
4.15. Маркирование адвербиативных семантических элементов формализованного представления содержания текстов юридических формул гражданского иска.	123
4.16. Маркирование составных семантических элементов – вложенных агентов формализованного представления содержания текстов юридических формул гражданского иска.	125
4.17. Маркирование семантических элементов – атрибутов в тексте формализованного представления содержания юридических формул гражданского иска.	132

4.18. Пример решения задачи определения принадлежности текста судебного решения к статье классификатора судебной отчетности, путем установления релевантности двух текстов и сравнения маркированных представлений их семантики.....	140
4.19. Примеры установления релеванности текстов судебного решения и текстов классификатора судебной отчетности путем сравнения маркированных представлений их семантики (тексты всех примеров заимствованы из 4.1).....	145
Глава 5. Автоматизированная система обработки документов гражданского судопроизводства Грузии «Фемида».....	156
5.1. Общее описание автоматизированной системы «Фемида»	156
5.2. Описание функционирования автоматизированной системы «Фемида»	161
5.3. Описание функционирования автоматизированной системы «Фемида» в режиме проверки релеванности текста формулировки судебного решения и текста статьи классификатора судебной отчетности	174
Выводы.....	181
Использованная литература.....	185

Список таблиц

Табл. 2.1.1.1	38
Табл. 2.1.1.2	38
Табл. 2.1.1.3	39
Табл. 2.1.1.4	40
Табл. 2.1.1.5	40
Табл. 2.1.1.6	40
Табл. 4.18.6.1	143

Список изображений

Рис. 4.1.2.1.....	88
Рис. 5.2.1.1.....	162
Рис. 5.2.1.2.....	163
Рис. 5.2.1.3.....	164
Рис. 5.2.1.4.....	165
Рис. 5.2.2.1.....	167
Рис. 5.2.3.1.....	168
Рис. 5.2.4.1.....	169
Рис. 5.2.4.2.....	170
Рис. 5.2.5.1.....	171
Рис. 5.2.5.2.....	172
Рис. 5.2.5.3.....	173
Рис. 5.3.1.1.....	175
Рис. 5.3.1.2.....	176
Рис. 5.3.2.1.....	177
Рис. 5.3.3.1.....	178
Рис. 5.3.4.1.....	179
Рис. 5.3.5.1.....	180

Использованные аббревиатуры

NOM	именительный падеж существительного или прилагательного
GEN	родительный падеж существительного или прилагательного
IN	существительное в постпозиционной форме «внутри»;
FROM	существительное в постпозиционной форме «из»
WITH	существительное в постпозиционной форме «вместе»
GENPL	родительный падеж существительного во множественном числе
AJECT	обстоятельная форма прилагательного
ADVERB	наречие или существительное в инструментальном или адвербиативном падеже
POSTPOS	отдельно записываемый послелог
UNION	союз
PRONOUN	местоимение
NOUN	основы существительных или прилагательных субстантивного характера, не порождающие личных глагольных форм
MASD	основы отглагольных существительных, порождающие личные глагольные формы

0. Введение.

0.1. Научная новизна поставленной в диссертации проблемы.

Научная новизна диссертационного исследования заключается в использовании формальных моделей смыслового значения текстов на естественном языке, в частности, естественном грузинском языке, для автоматизации судебного делопроизводства по гражданским делам, являющихся формулировкой содержания конкретного гражданского иска, записанного лицом, входящим в состав судебного органа и имеющим на это действие юридическое право.

Принятый в работе оригинальный научный подход заключается в построении специальной модели семантики текстов судебной документации, записанных на ограниченном естественном грузинском языке. Основная функция модели текста – возможность автоматически построить формальный образ его содержания и получить строго формальное определение задач автоматизированной классификации, позволяющих свести релевантность конкретных текстов к вычисляемым процедурам специального вида.

Важной стороной принятого научного подхода является использование современных принципов, т.н. «разметки текстов» для разрешения их синтаксической и семантической неоднозначности. Для рассматриваемой предметной области построение системы разрешения неоднозначности существенным образом опирается на грамматические особенности грузинского языка и ограничения, присущие языку конкретной предметной области формулировок содержания гражданского иска. Реализация системы разрешения неоднозначности основана на принципе генерации и отбора синтаксических, лексических и семантических гипотез путем интерактивно-диалогового взаимодействия с квалифицированным экспертом, а также путем использования экспертных знаний о текстовых корпусах судебной документации.

0.2. Актуальность и практическое значение проблемы.

Истоки настоящей работы находятся в одном из важных приложений автоматизированных систем управления – обработки документов, содержащих информационно значимые тексты на версии естественного языка, принятой в проблемной сфере автоматизации. Данная работа посвящена одной из подобных задач, возникающих при создании систем автоматизированного ведения судебного делопроизводства по гражданским и уголовным делам. Внедрение средств автоматизации в эти подразделения государственных органов Грузии является одним из самых многообещающих путей повышения объективности и независимости судебных органов, повышения уровня их работы до уровня мировых стандартов. С другой стороны, попытки решения этой задачи сталкиваются с рядом объективных и субъективных трудностей, вызванных особенностями традиционного уровня формализации создаваемых в процессе судопроизводства многочисленных документов, вполне удовлетворительных с точки зрения немашинного делопроизводства, но становящимся заметным препятствием в процессе внедрения автоматизации.

Среди источников искажения выходных документов автоматизированных систем указанной области наиболее критичным является несоответствие между текстовым образом формулировки содержания иска с ее формализованным кодом для принятого в Грузии классификатора судебной отчетности. Эта задача на практике играет очень заметную роль как в оценке работы суда через регламентированные формы статистической отчетности, так и для решения различных поисковых задач, связанных с такой социально ответственной деятельностью, как контроль над решением судебных заседаний, оценка различных сторон функционирования судебной системы, поиск конкретных дел и т.д.

Внедрение совершенных и наукоемких средств автоматизации, необходимо для принципиального сокращения роли субъективного фактора и соответственно повышения объективности судебного процесса, что в значительной степени увеличит социальную и психологическую роль в формировании доверия общества к судебной системе.

Предлагаемый в диссертационной работе научный подход радикального решения этой задачи заключается в создании автоматизированных систем,

обладающих расширенными «интеллектуальными» возможностями, с достаточной для практики эффективностью и обеспечивающих обнаружение и исправление возможных противоречий в судебной документации.

0.3. Научные особенности решения поставленной в диссертации проблемы.

Поставленная в диссертационной работе научная задача имеет ряд прототипов, известных из теории поисковых систем, ставящих своей целью поиск семантически адекватных документов не только по лексическому совпадению найденных слов текста с заданным поисковым образом. Разработка подобных систем ведется с 50-ых годов, развившись, в частности, в научное направление т.н. «контент-анализ».

Основные трудности решения задачи сосредоточены в грамматическом и в последующем семантическом анализе текста на естественном языке. Успешность решения этой проблемы определяется сложностью описания грамматики естественного языка, а также возникновением порождаемых при этом неоднозначностей. Не случайно наибольшие успехи в этой области достигнуты для естественного английского языка с его минимальными морфологическими средствами и однозначными правилами синтаксиса, в частности, жестко фиксированной последовательностью слов в предложении. Для таких флективных языков со сравнительно свободной последовательностью слов в предложении, как немецкий или русский, несмотря на успехи в построении модели морфологии, даже задача синтаксического анализа далека от своего полного решения, не говоря уже о трудностях в моделировании семантики текста для лексически достаточно развитых предметных областей.

Естественный грузинский язык представляет в этом отношении структуру более сложную, чем индо-европейские языки. Это объясняется такими важнейшими чертами иберийско-кавказских языков, как их эргативность, что в сочетании с вытекающей отсюда полиперсоналией грузинского глагола приводит к невозможности использования для них даже ограниченного аппарата морфо-синтаксического анализа, развитого для индо-европейских языков. На сегодняшний день прецедента решения этой задачи для естественного грузинского языка не существует.

Успешность решения задач, относящихся к классу «контент-анализа», требует для грузинского языка развития достаточно оригинальных научных подходов, сочетающихся со значительными ограничениями на язык, действующими в избранной предметной области.

В данном диссертационном исследовании наиболее важным из них является подтверждаемая в юридической практике составления формул гражданского иска допустимая грамматическая форма принятых текстов. Она может являться только сингамами грузинского языка, не имеющими сказуемого, т.е. не содержащими личных глагольных форм, а только-существительные, прилагательные и наречия, объединенные в специальную структуру, установленную сложившимися в этой области традициями описания содержания документов. Это позволяет с высокой степенью адекватности и без нарушения требований минимальной загруженности оператора диалогами грамматического содержания достигнуть успешного решения поставленной в диссертации проблемы.

0.4. Краткая постановка проблемы, решению которой посвящена диссертационная работа.

Проблема в общей постановке может быть сформулирована следующим образом.

Даны два текста на естественном грузинском языке, ограниченном предметной областью судебного делопроизводства по гражданским делам. Первый текст является формулировкой содержания конкретного гражданского иска. Вторым текстом является текст одной из статей принятого в судопроизводстве Грузии кодификатора форм статистической отчетности

Задача заключается в построении формализованной модели указанных двух текстов, позволяющей автоматически определить формальные образы их содержания и получить строго формальное определение их релевантности, после чего установление степени релевантности конкретных текстов становится вычисляемой процедурой специального вида. При положительном решении задачи заданный в классификаторе код статьи кодификатора объявляется полностью формализованным образом содержания данного гражданского иска и

используется в дальнейшем во всех статистических и поисковых задачах автоматизированной системы.

Формализованная модель, в свою очередь, разбивается на следующие компоненты:

- модель синтаксиса текста гражданского иска на ограниченном естественном грузинском языке;

- модель семантики текста гражданского иска на ограниченном естественном грузинском языке;

- модель установления релевантности формальных образов двух формул гражданского иска, преобразованных соответственно в две схемы причинно-квалитативных связей между выявленными в формулах сущностями.

Каждая из указанных моделей, в свою очередь, разбивается на частные модели, реализующие отдельные этапы морфологического, синтаксического и семантического анализа текстов предметной области, а также установление релевантности между их формальными образами.

Глава 1. Постановка задачи создания формализованных моделей семантики юридически-правовой документации на естественном грузинском языке (на примере юридических формул гражданского иска) и использование аппарата грамматической разметки текстов для ее решения.

1.1. Актуальность поставленной в диссертации проблемы.

Одной из самых примечательных черт современного состояния теории и практики создания автоматизированных систем управления является существенное развитие их «интеллектуализации» - перенесение центра тяжести решаемых проблем с традиционных задач статистики и выборки данных по различным логическим условиям на задачи углубленного анализа данных с целью интеллектуальной поддержки принятия решения непосредственных пользователей систем. Это позволяет сделать существенный шаг в повышении эффективности человеко-машинного взаимодействия, по существу стимулируя новые постановки вопроса прагматической ценности самой автоматизации управления во многих прикладных областях.

Успешное развитие этих направлений связано, с одной стороны, с накоплением и переосмыслением опыта эксплуатации уже созданных АСУ, а с другой стороны, - со все более заметными успехами в создании теоретической базы информатики, в том числе, с плодотворными теоретическими исследованиями в области прикладной лингвистики, как научной базы для переориентации средств общения на языки, возможно более близкие к естественным языкам. Подобно тому значительному шагу в области построения автоматизированных систем, каким было создание несколько десятилетий тому назад концепции реляционных баз данных и средств общения с ними, являвшихся развитием достаточно близкого к человеческому мышлению языка исчисления предикатов, ведущиеся в настоящее время интенсивные исследования в направлении создания средств общения на естественных языках

призваны обеспечить следующий качественный скачок в обеспечении «дружественного» функционирования человека как составного элемента человеко-машинных систем, которому автоматизированная система должна создать эффективные средства преодоления сложности стоящих перед нею новых практических задач.

Истоки настоящей работы находятся в имеющей значительную актуальность области практических задач, являющейся одной из важных приложений автоматизированных систем управления – обработки документов, содержащих информационно значимые тексты на версии естественного языка, принятой в проблемной сфере автоматизации. В данной работе рассматривается одна из подобных задач, возникающая при создании систем автоматизированного ведения судебного делопроизводства по гражданским и уголовным делам. Внедрение средств автоматизации в эти подразделения государственных органов Грузии является одним из самых многообещающих путей для повышения объективности и независимости судебных органов, повышения уровня их работы до уровня мировых стандартов.

Решение указанной задачи сталкивается с рядом объективных и субъективных трудностей, вызванных особенностями традиционного уровня формализации создаваемых в процессе судопроизводства многочисленных документов, вполне удовлетворительных с точки зрения немашинного делопроизводства, но становящимися заметным препятствием в процессе внедрения автоматизации.

Среди источников нарушения формальной корректности документов, которая существенно сказывается на обеспечиваемых автоматизированной системой выходных документах, одним из наиболее критичных является несоответствие между текстовым образом формулировки содержания иска с ее формализованным кодом для принятого в Грузии классификатора судебной отчетности. Эта задача, представляющаяся первоначально достаточно частной, в действительности играет ключевую роль как в оценке работы суда через регламентированные формы статистической отчетности, так и в различных поисковых задачах, прежде всего, связанных с такой ответственной проблемой, как контроль над решением судебных заседаний, оценка различных сторон функционирования судебной системы, поиск конкретных дел и т.д.

Очевидной причиной ошибок этого характера (и как следствие, снижения юридического качества судебных документов) является то, что указанные реквизиты традиционно заполняются оператором – ответственным сотрудником суда, отвечающим за судебное делопроизводство.

Использование человеческой интуиции в создании ответственных документов, как хорошо известно, предполагает длительный процесс отбора, подготовки, тестирования и, в необходимых случаях, замены кадров. С этой точки зрения, внедрение средств автоматизации, с одной стороны, необходимо для принципиального сокращения роли субъективного фактора и соответственно повышения объективности судебного процесса, что имеет очень большую социальную и психологическую роль для формирования доверия общества к судебной системе. С другой стороны, попытка повышения уровня формализации документов до требований, вытекающих из необходимости автоматизированного решения нетривиальных задач принятия решения, неизбежно наталкивается на трудности формализации содержания текстов, созданных экспертами, излагающими их на естественном языке.

Хотя квалификация сотрудников судебной системы Грузии в последние годы существенно возросла, участие человека в процессе создания документа, могущего иметь формализованную интерпретацию, как это хорошо известно из фундаментальных принципов функционирования автоматизированных систем, неизбежно включает в этот процесс человеческий фактор и, как следствие, является потенциальным источником нарушения целостности: возникновение в базах данных реквизитов с противоречащими друг другу значениями, что равнозначно риску искажения смысла всех содержащих эти реквизиты документов. Как показывает практика, масштабы подобного «шума» в автоматизированных системах не являются исчезающим фактором, а достаточно часто влекут за собой искажение адекватности всей цепочки выходных документов с соответствующими негативными социальными и юридическими последствиями.

Безусловно, проводимая в последние годы специальная подготовка персонала, участвующего в формировании машинного образа судебной документации (в том числе с участием международных организаций), позволила заметно снизить возможность возникновения недостатков подобного характера, однако, как показывает практика, полностью исключить их оказалось почти

невозможным, поскольку это является неотъемлемым следствием участия человека в процессе обработки информации. В частности, все еще заметная часть документов судебного делопроизводства содержит неадекватные коды классификатора судебной отчетности, что вызывает цепочку несоответствий выходных данных при решении самых различных с общественно-политической точки зрения задач, связанных с оценкой работы судебных органов.

Одним из современных путей радикального решения этой задачи является создание автоматизированных систем, обладающих «интеллектуальными» возможностями, с достаточной для практики эффективностью обеспечивающих обнаружение и исправление возможных противоречий подобного типа.

1.2. Научная новизна поставленной в диссертации проблемы.

Научной новизной изложенной выше практически актуальной задачи, фактически недоступной для недостаточно наукоемких методов исследования и проектирования автоматизированных систем, является успешно развиваемая в последние годы теория представления смысла текстов на естественном языке, в частности, ее развитие для естественного грузинского языка.

Попытки формализации семантики текстов на естественных языках ведутся уже несколько десятилетий. Они мотивированы необходимостью решения ряда важнейших для практики задач, связанных с проблемами поиска релевантных текстов в больших массивах библиотечного типа, управления роботами, принятия решения в сложных системах, в которых критическим фактором являются затраты на подготовку специалистов в сочетании со сложностью языка общения с операторами, а также значительное количество других задач, возникающих при автоматической (или автоматизированной) обработке текстов, не ограниченных условиями жесткого кодирования их формы.

Накопленный до настоящего времени опыт показывает, что решение подобных задач в самой общей формулировке остается делом отдаленного будущего, однако достигнуто решение отдельных частных задач, успешно использующих ограничения, присущие конкретным предметным областям. Принятый в работе оригинальный научный подход заключается в построении

специальной модели семантики текстов судебной документации, записанных на ограниченном естественном грузинском языке. В концептуальной постановке эта задача может быть сформулирована следующим образом.

Дан текст на естественном грузинском языке, ограниченном предметной областью судебного делопроизводства по гражданским и уголовным делам, который является формулировкой содержания конкретного гражданского иска, записанного имеющим на это юридическое действие право лицом, входящим в состав судебного органа, например, секретарем судебного заседания. Задача в общей формулировке заключается в построении модели указанного текста, позволяющей автоматически построить формальный образ его содержания и получить строго формальное определение различных задач определения релевантности, что дает возможность свести релевантность конкретных текстов к вычисляемым процедурам специального вида.

Трудность построения подобных моделей заключается в том, что на них оказывает решающее воздействие важнейшая черта, присущая всем естественным языкам, – феномен их синтаксической и семантической неоднозначности (см. обзор литературы в гл.2). Следствием этого, к сожалению, неизбежного феномена при изучении естественных языков является невозможность полной формализации синтаксиса и семантики текстов, по крайней мере в настоящее время, пока наука все еще весьма далека до исчерпывающего формального представления экстралингвистических знаний, которые окружают предметную область формализуемых текстов.

Одним из наиболее эффективных методов, способных обеспечить инженерно эффективное решение этой задачи, является организация интерактивно-диалогового разрешения неоднозначности. Она основана на генерации синтаксических, лексических и семантических гипотез, имеющих вид поддеревьев (обычно бинарных), множество которых оказывается достаточно ограниченным для того, чтобы предъявить их квалифицированному эксперту-пользователю автоматизированной системы, получить его ответы и внести полученные в базу данных (в общем случае – в базу знаний) представлений текстов, добившись тем самым устранения всех неоднозначностей. Базы данных, над которыми произведена подобная трансформация, получили название «текстов, оформленных в виде деревьев» («Treebanking»); в

русскоязычной научной литературе преимущественно используется термин «синтаксически (семантически) размеченные деревья».[84]

Конкретная постановка и способ решения задачи синтаксической разметки текстов отражает реализацию научной новизны поставленной в диссертации проблемы. Ее построение существенным образом опирается на грамматические особенности грузинского языка и ограничения, присущие языку конкретной предметной области формулировок содержания гражданского иска.

1.3. Концептуальные основы формальной модели синтаксиса текста гражданского иска на ограниченном естественном грузинском языке.

1.3.1. Существенным ограничением предметной области, вытекающих из юридической специфики корректных формул гражданского иска, является допустимая грамматическая форма принятых текстов. Они могут являться только *синтагмами* грузинского языка, не имеющими сказуемого, т.е. не содержащими личных глагольных форм, а только существительные, прилагательные и наречия, объединенные в специальную структуру, более подробно рассмотренную в последующих главах. Возможно также перечисление синтагм, разделенных союзами или знаками препинания (обычно запятыми). [17]

Это ограничение играет фундаментальную роль в принятой в настоящей работе концепции построения формальной модели содержания текста на ограниченном естественном грузинском языке. Благодаря этому ограничению, отпадает необходимость учета морфологически наиболее сложного элемента грузинского языка – его глагола, как известно, характеризующегося полиперсонализмом, что существенно облегчает все задачи морфологического и синтаксического разбора. Кроме того, наличие базы данных масдаров как отглагольных имен, семантически тесно связанных с вербальной структурой предложения, но не требующих учета сложной ранговой структуры личных глагольных форм, позволяет существенно упростить и унифицировать важнейшую часть алгоритмов распознавания – блок морфологического анализа.

1.3.2. Основы всех словоформ, входящих в синтагму, принадлежат заранее определенному словарю и связаны общепринятыми для грузинского языка грамматическими правилами, регулирующими морфологические признаки именных словоформ и подчинение их взаимных связей установленным правилам синтаксиса грузинского языка. [2]

1.3.3. Для представления синтаксической структуры текста формулы гражданского иска на ограниченном естественном грузинском языке и построения формального образа его смысла предполагается использование специализированной базы знаний, ориентированной на решении поставленной задачи. Эта база данных одновременно является инструментальным средством для проведения интерактивной синтаксической и семантической разметки.

Базой фактов системы является перечень основ ограниченного естественного языка предметной области, лексический состав которого жестко ограничен предметной областью юридической и связанной с процессуальными действиями областью (в общем случае приблизительно 500 основ), в сочетании с перечислением морфологических и синтаксических элементов грузинского языка, используемых в рассматриваемых текстах.

Базой правил системы является перечень морфологических и синтаксических правил для построения корректной синтагмы, удовлетворяющей условиям (1.3.1)-(1.3.2), а также перечень семантических правил для унификации описания смысла текста с целью подготовки гипотез для их синтаксической разметки. Наконец, база правил системы содержит перечень правил и критериев для непосредственного сравнения формул унифицированного вида с целью решения задачи релевантности.

1.3.4. Дальнейшее уточнение формальной постановки содержательной задачи происходит путем морфологического и синтаксического анализа заданного текста, в результате которого строится формальный машинный образ их смысла, позволяющий, в частности, строго классифицировать их, а также поставить и вычислить степень релевантности, что в свою очередь является прямым ответом исходной задачи. [1]

Ввиду сложности модели, происходящей от грамматической сложности грузинского языка, ее формальное описание производится по иерархической схеме, на каждом уровне которой определяется частная подмодель, получающая данные от подмодели предыдущего уровня и сообщающая результаты своей

работы подмодели последующего уровня. Процесс работы подмодели заключается в построении специальных списковых структур, исходящих от последовательности слов заданной синтагмы и модифицируемых на подмодели каждого уровня путем связывания с полученными маркерами грамматической и семантической информации, вплоть до получения структур конечного вида, для которых решение задачи классификации и проверки релевантности является уже полностью алгоритмизируемой процедурой.

В случае, если текст не является корректным с точки зрения рассмотренных выше ограничений (например, в нем найдены не представленные в словаре основы или недопустимые правила словообразования и словосочетания), анализ текста завершается, выдается диагностическое сообщение.

1.3.4.1. Первой, расположенной на низшем иерархическом уровне подмоделью, является морфологический анализатор корректных словоформ грузинского языка заданной предметной области, который обеспечивает выделение из заданного текста содержащих его основ с одновременным отделением флективных окончаний, что позволяет получить полную морфологическую оценку распознанных словоформ. В сочетании с найденными основами результат именуется *списком сущностей* системы.

1.3.4.2. Второй, следующей по иерархии подмоделью, является первый шаг синтаксического анализа заданного текста, заключающийся в выявлении среди списка сущностей *актора* синтагмы – образующего элемента ее синтаксической структуры, которому синтаксически подчинены все остальные слова текста, но оно само не подчинено ни одному из них. Актором синтагмы, как вытекает из принятых ограничений, может являться только отглагольное имя в номинативной форме, преимущественно находящееся в последней позиции текста. Если синтагма состоит из нескольких подсинтагм, соединенных союзами, поиск акторов ведется по отдельности в каждом из них.

1.3.4.3. Третьей подмоделью является второй шаг синтаксического анализа заданного текста, заключающийся в выявлении среди списка сущностей *предметов (агентов) акции*, на которые непосредственно распространяется действие, определяемое выявленным ранее актором синтагмы. Используемые правила, определяемые традиционной схемой синтаксического подчинения в грузинском языке, опираются на то, что агентами акции могут быть, большей

частью, существительные в генетиве, расположенные в синтагме непосредственно перед актором.

1.3.4.4. Четвертой подмоделью является третий шаг синтаксического анализа заданного текста, заключающийся в выявлении среди списка сущностей необязательных *адвербиативных оценок* - прилагательных или наречий, определяющих условия или обстоятельства применения действия, представленного актором к предметам акции. Традиционная синтаксическая структура грузинского языка предполагает использование для подобных связей инструментального падежа от инструментальных наречий или специальных обстоятельных форм генетива, с необязательными ограничениями в последовательности слов.

1.3.4.5. Пятой подмоделью является третий шаг синтаксического анализа заданного текста, заключающийся в выявлении среди списка сущностей необязательных *атрибутивных свойств*, представляющих тонкую синтаксическую структуру атрибутивных конструкций, обычно имеющих синтаксическую квалификацию обстоятельства. Традиционная синтаксическая структура грузинского языка предполагает использование для подобных подчинений инструментального или генитивного падежа имен прилагательных в сочетании с позицией атрибутивных слов, непосредственно предшествующих сущности, на которую распространяется действие атрибута.

1.4. Концептуальные основы формальной модели семантики текста гражданского иска на ограниченном естественном грузинском языке.

Принятый в настоящей работе подход к решению задачи построения формальных образов смысла текста основывается на преобразовании синтаксической структуры синтагмы в структуру *причинно-кавалитативных связей* между сущностями и их синтаксическими свойствами, выявленными в формулах гражданского иска, которое осуществляется по отдельной иерархической схеме, включающей в себя несколько подмоделей.

Формальная модель семантики иллюстрируется синтаксическим и семантическим анализом типичных юридических формул гражданского иска,

выбранных с помощью стохастической процедуры из имеющейся базы данных текстов, содержащихся в решениях судебных инстанций Грузии по гражданским делам.

1.4.1. Для решения поставленной выше задачи получения формализованного описания семантики текстов юридических формул гражданского иска, в первую очередь необходимо такое преобразование морфологических и синтаксических структур грамматических конструкций, которое в наибольшей возможной степени было бы свободно от грамматических свойств естественного грузинского языка, на котором они были первоначально записаны. Это означает, что в явной форме должны быть представлены глубинные структуры языка, определяющие семантику текста, что требует устранения тех черт естественного языка, которые определяются его эволюцией и традициями, а также специфику сублингвистической среды, в которой существует предметная область, порождающая юридические формулы гражданского иска как результат специфических социально-коммуникативных процессов.

Данная задача сводится к следующим подзадачам.

1.4.1.1. Приведение всех лексических единиц, встречающихся в тексте, к нормализованной форме с опущенными морфологическими признаками. Эта задача для существительных и прилагательных означает преобразование словоформы в именительный падеж единственного числа, рассматривается подробно в гл.3.

1.4.1.2. Перестановка (транспозиция) всех встречающихся в тексте синтаксических единиц в последовательность развития их причинных связей, которая в большинстве случаев не совпадает с последовательностью словоформ в естественном языке. В результате формируется последовательность основ, отражающая развитие семантического подчинения в тексте юридических формул гражданского иска, которая уже мало зависит от особенностей конкретного языка-носителя текста (в данном случае грузинского).

Для рассматриваемой предметной области, с учетом принятых выше модельных ограничений, эта последовательность формализуется следующим образом (в порядке семантического подчинения сверху-вниз):

Основные синтаксические элементы:

– актор (единственный в принятом представлении синтаксической структуры);

– прямое дополнение (единственный в принятом представлении синтаксической структуры);

Дополнительные синтаксические элементы, не обязательные в синтаксической структуре:

– косвенное дополнение;

– атрибуты, подчиняющиеся прямому дополнению;

– адвербы или адвербиативные конструкции, подчиняющиеся прямому дополнению;

– атрибуты внутри адвербиативных конструкций, подчиняющиеся непосредственно адвербу.

1.4.2. Центральной задачей получения формализованного описания семантики текстов юридических формул гражданского иска является маркирование ключевых семантических элементов формализованного представления содержания текстов юридических формул гражданского иска – актора и агента. Содержательно оно заключается в установлении главной причинной связи синтагмы, направленной от актора к предмету акции – агенту, на который непосредственно распространяется действие актора, что обеспечивает его представление как причинного следствия актора.

Хотя семантическое понятие актора является, с точки зрения принятой терминологии, близким аналогом одноименного синтаксического понятия, но оно принадлежит не внешней структуре текста, а к его внутренней, глубинной структуре, поскольку в данном случае имеет место не синтаксическое подчинение всех остальных слов текста (через иерархическую структуру), а маркирование ведущего семантического элемента ситуации, которую описывает данный текст и которая может быть однозначно переведена на другие языки.

Семантическое понятие агента является аналогом синтаксического понятия прямого дополнения, однако отличается от него игнорированием грамматического управления (обычно – родительного падежа) и направлением главной причинной связи текст от актора к представленному агентом объекту, на который непосредственно распространяется действие актора.

1.4.3. Следующей задачей получения формализованного описания семантики текстов юридических формул гражданского иска является

маркирование составных семантических элементов – вложенных агентов в формализованном представлении содержания текстов юридических формул гражданского иска, что отражает такую распространенную синтаксическую конструкцию естественного грузинского языка, как вложенные (распространенные) дополнения.

С их помощью семантика агента может быть существенно уточнена словоформами, являющимися существительными в родительном падеже, синтаксически подчиненными агенту (прямому дополнению), благодаря этому строится сложный агент, имеющий иерархическую структуру развивающую семантику основного агента, непосредственно подчиненного актору.

В связи с этим, важна семантическая классификация понятия агентов, имеющих вложенную структуру, поскольку она определяет подчиненность объектов, к которым применяются действия, представленные актором.

1.4.3.1. Подчиненный агент принадлежности или уточнения основного агента, образуемый родительным падежом одного существительного

1.4.3.2. Подчиненный агент принадлежности или уточнения основного агента, образуемый цепочкой родительных падежей нескольких существительных.

1.4.3.3. Подчиненный агент, оппозиционный основному агенту, образуемый родительным падежом существительного в сочетании со специальным союзом исключения.

Синтаксическая неоднозначность при решении указанной задачи возникает практически во всех случаях, когда словоформа, представляющая агента, нарушает эвристические правила, вытекающие из классификации 1.4.3.1-1.4.3.3.

1.4.4. Следующей задачей получения формализованного описания семантики текстов юридических формул гражданского иска является маркирование их адвербиативных семантических элементов, которые непосредственно раскрывают содержание как описанного актором действия, так и особенности его применения к агенту.

Аналогично 1.4.3, важное значение имеет семантическая классификация адвербиативных семантических элементов, которая определяет условия или обстоятельства применения действия, представленного актором. Для рассмотренных текстов мы принимаем следующую систему классификации

адвербов, которая существенным образом связана с особенностями предметной области юридических формул гражданского иска

1.4.4.1. Адverb юридического статуса действия, определяемого актором.

1.4.4.2. Адverb имущественного статуса действия актора по отношению к агенту.

1.4.4.3. Адverb пространственного направления действия актора, его логической принадлежности или связи между актором и другими семантическими элементами текста.

1.4.4.4. Адverb места выполнения действия актора.

1.4.4.5. Адverb логической связи, каузирующей действие актора.

Синтаксическая неоднозначность при решении указанной задачи возникает практически во всех случаях, когда словоформа, представляющая адverb, не включена в классификацию адвербов.

1.4.5. Завершающей задачей получения формализованного описания семантики текстов юридических формул гражданского иска является маркирование их атрибутивных семантических элементов, которые окончательно раскрывают содержание как описанного актором действия, так и особенности его применения к агенту. В рассматриваемой предметной области они являются семантической формализацией таких распространенных синтаксических конструкций, как определения (атрибуты), которые могут быть простыми и вложенными, благодаря чему уточняется иерархическая структура семантики основного агента, непосредственно подчиненного актору.

В связи с этим важна семантическая классификация атрибутивных конструкций, отражающая их синтаксическую роль, поскольку она определяет подчиненность объектов, к которым применяются действия, уточняемые атрибутами. Аналогично рассмотренным выше, эта классификация также существенным образом определяется особенностями предметной области юридических формул гражданского иска

1.4.5.1. Атрибуты, непосредственно подчиненные актору или агенту.

1.4.5.2. Иерархически лежащие ниже атрибуты, подчиненные уже подчиненным агентам принадлежности или уточнения основного агента.

Синтаксическая неоднозначность при решении указанной задачи возникает практически во всех случаях, когда словоформа, представляющая атрибут, нарушает вытекающие из 1.4.5.1-1.4.5.2 эвристические правила.

1.4.6. Результатом построения формализованного описания семантики текстов юридических формул гражданского иска, которое сводится к маркированию их различных семантических элементов, является списковая структура специального вида, отражающая унифицированное направление действия слева направо причинных и качественных связей, выявленных в тексте семантических сущностей.

В результате, при определенном упрощении сложной структуры причинно-качественных связей, характерной для естественного языка с его традиционным богатством отражения синтаксических феноменов, все причинные и качественные зависимости, отображенные в формальной модели формулы гражданского иска, с приемлемой, как показывают результаты проведенных экспериментальных исследований, степенью адекватности отображения содержания, могут быть формально описаны с помощью нескольких префиксных операторов элементами списковой структуры, содержащей основы естественного грузинского языка (в пределах ограничений предметной области).

1.5. Постановка задачи автоматизированной классификации, релевантного поиска и релевантного сравнения текстов в принятых формальной модели семантики текста гражданского иска на ограниченном естественном грузинском языке.

Решение задачи построения формальной модели текста на естественном грузинском языке, ограниченного предметной областью судебного делопроизводства по гражданским (фактически – и по уголовным) делам, являющейся формулировкой содержания конкретного гражданского иска, открывает возможность получить строго формальное определение различных задач, связанных с определением релевантности и свести релевантность конкретных текстов к вычисляемым процедурам специального вида. В первую очередь, становится доступной решение задачи автоматизированной классификации содержания формулировки решения суда по гражданскому иску в соответствии с официально принятым в государстве классификатором судебных решений. В формальной постановке она является естественной

надстройкой над задачей построения формализованного описания семантики текстов юридических формул гражданского иска и также опирается на идею маркирования составных семантических элементов – вложенных агентов формализованного представления содержания текстов. Обобщение состоит в том, что рассматривается параллельное построение моделей не одного, а двух текстов с последующим формализованным сравнением списковых структур. В точной постановке обобщенная задача может быть записана следующим образом.

Даны два текста на естественном грузинском языке, ограниченные предметной областью судебного делопроизводства по гражданским и уголовным делам. Первый текст является формулировкой содержания конкретного гражданского иска, записанного имеющим на это юридическое действие право лицом, входящим в состав судебного органа, например, секретарем судебного заседания. Вторым текстом является также записанный на ограниченном естественном грузинском языке текст одной из статей принятого в судопроизводстве Грузии кодификатора форм статистической отчетности, лексическая, грамматическая и семантическая структура которого в принципе не отличается, за исключением некоторых деталей, от структуры первого текста.

Задача общей формулировки заключается в построении модели указанных двух текстов, позволяющей автоматически построить формальные образы их содержания и получить строго формальное определение их релевантности, после чего установление степени релевантности конкретных текстов становится вычисляемой процедурой специального вида. При положительном решении задачи заданный в классификаторе код статьи кодификатора объявляется полностью формализованным образом содержания данного гражданского иска и используется в дальнейшем во всех статистических и поисковых задачах автоматизированной системы.

Формальные образы двух формул гражданского иска, преобразованные соответственно в две схемы причинно-квалитативных связей между выявленными в формулах сущностями, объявляются в принятом в настоящей работе подходе РЕЛЕВАНТНЫМИ, если выполняются следующие условия:

1.5.1. Списковые структуры идентичны с точностью до возможности игнорирования атрибутивных свойств, которые выполняются с применением заранее объявленных экспертных правил.

1.5.2. Между некоторыми текстуально различающимися сущностями формальных образов текста (актерами, агентами, адвербами) могут быть установлены заранее определенные экспертным путем отношения эквивалентности.

Глава 2. Обзор методов морфологической, синтаксической и семантической разметки текстовых корпусов на естественных языках

2.1. Обзор текстовых корпусов, имеющих морфологическую, синтаксическую и семантическую разметку (на примере национального корпуса русского языка)

Постоянно развивающийся национальный корпус русского языка к сентябрю 2008г. насчитывает 52 392 текста общим объемом 149 357 020 словоупотреблений. [101]

2.1.1. Статистика Корпуса

I. Распределение текстов по подкорпусам

Подкорпус	Число текстов	Число словоупотреблений
Основной корпус	42 387	147 577 522
в том числе со снятой омонимией:	2 215	5 884 661
Диалектный корпус	122	144 099
Поэтический корпус	9 675	2 586 710
Обучающий корпус	230	649 684

Табл. 2.1.1.1

II. Распределение текстов основного корпуса по видам и другим метапризнакам

Вид текста	Число текстов	Число словоупотреблений	% словоупотреблений
Художественные письменные тексты	3 893	58 547 176	39,7
Нехудожественные письменные тексты	37 249	83 218 964	56,4
Устная речь	1 245	5 810 482	3,9

Табл. 2.1.1.2

Художественные тексты

Жанр	Число текстов	Число словоупотреблений	% словоупотреблений
автобиографическая проза	3	323 808	0,6
детектив, боевик	65	4 796 573	8,2
детская литература	380	2 479 165	4,2
документальная проза	109	3 177 007	5,4

драматургия	92	1 236 635	2,1
детская литература	380	2 479 165	4,2
историческая проза	82	2 968 863	5,1
любовная история	29	622 838	1,1
нежанровая проза	2 008	35 199 610	60,1
приключения	11	1 246 928	2,1
фантастика	174	3 902 519	6,7
юмор и сатира	788	1 566 038	2,7
прочее	152	1 027 541	1,8
Всего:	3 893	58 547 525	100,0

Табл. 2.1.1.3

Нехудожественные тексты

Сфера функционирования	Число текстов	Число словоупотреблений	% словоупотреблений
бытовая	590	1 986 286	2,4
официально-деловая	1 244	1 814 812	2,2
производственно-техническая	131	233 569	0,3
публицистика	28 659	56 549 987	68,0
реклама	1 221	513 955	0,6
учебно-научная	4 578	18 269 877	22,0
церковно-богословская	735	2 540 674	3,1
электронная коммуникация	91	1 310 804	1,6
Всего:	37 249	83 219 964	100,0
Тематика текста	Число текстов	Число словоупотреблений	% словоупотреблений
администрация и управление	344	233 211	0,3
армия и вооруженные конфликты	705	1 668 166	2,0
астрология, парапсихология, эзотерика	54	49 607	0,1
бизнес, коммерция, экономика, финансы	3 447	2 721 353	3,3
дом и домашнее хозяйство	824	671 190	0,8
досуг, зрелища и развлечения	807	516 326	0,6
здоровье и медицина	856	2 227 828	1,5
информационные технологии	16	14 583	0,0
искусство и культура	3 289	5 236 518	6,3
криминал	658	496 855	0,6
мемуары, записные книжки, дневники	419	20 604 802	24,8
наука и технологии	5 446	17 334 780	20,8
образование	163	172 806	0,2
политика и общественная жизнь	11 217	15 526 635	18,7
право	506	1 359 877	1,6
природа	327	550 070	0,7
производство	1 007	912 649	1,1
религия	1 036	3 502 679	4,2
сельские хозяйство	211	129 089	0,2

спорт	1 377	1 779 891	2,1
техника	869	820 206	1,0
транспорт	169	162 310	0,2
философия	90	1 411 709	1,7
частная жизнь	2 927	5 730 035	6,9
Всего:	37 249	83 219 964	100,0

Табл. 2.1.1.4

Устная речь

Тип	Число текстов	Число словоупотреблений	% словоупотреблений
публичная речь	617	3 738 790	64,3
непубличная речь	445	470 597	8,1
речь кино	183	1 601 095	27,6
Всего:	1 245	5 810 482	100,0

Табл. 2.1.1.5

II. Распределение словоупотреблений по частям речи (только для корпуса со снятой грамматической омонимией, данные на 28 ноября 2007 г. - объём корпуса со снятой омонимией 5,5 млн).

Часть речи	Число словоупотреблений	% словоупотреблений
существительное	1 554 272	28,50
прилагательное	465 743	8,54
числительное	82 809	1,52
в том числе записанные прописью	39 827	0,73
в том числе записанные цифрами	42 982	0,79
числительное-прилагательное	21 081	0,39
глагол	931 687	17,08
наречие	222 502	4,08
предикатив	38 260	0,70
вводное слово	24 954	0,46
местоимение-существительное	443 205	8,13
местоимение-прилагательное	255 772	4,69
местоименное наречие	120 568	2,21
местоимение-предикатив (<i>некого, нечего</i>)	602	0,01
предлог	568 295	10,42
союз	433 815	7,95
частица	258 085	4,73
междометие	7 192	0,13
инициал	9 726	0,18
прочие (иностранные слова, звукоподражания)	15 781	0,29
Всего:	5 454 349	100,00

Табл. 2.1.1.6

2.1.2. Структура морфологической информации

Морфологическая информация, приписываемая произвольному слову в тексте, состоит из четырех полей или групп помет:

Лексема, которой принадлежит словоформа (указывается словарная запись данной лексики и ее принадлежность к той или иной части речи).

Множество грамматических признаков данной лексики, или словоклассифицирующие характеристики (например, род для существительного, переходность для глагола).

Множество грамматических признаков данной словоформы, или словоизменительные характеристики (например, падеж для существительного, число для глагола).

Информация о нестандартности грамматической формы, орфографических искажениях и т. п.

Морфологический разбор (или множество морфологических разборов), приписанный каждой словоформе в составе поисковой выдачи. В корпусе со снятой грамматической омонимией указывается полный разбор, в корпусе без снятой грамматической омонимии – только лексема и часть речи.

В основу метаязыка грамматических помет, ввиду предполагаемой широкой международной аудитории пользователей Корпуса, положена система сокращенных помет («тегов») на основе латинского алфавита. В то же время предусмотрена возможность использования при поиске традиционных названий категорий на русском языке (в форме «грамматические признаки»).

Ниже приводим инвентарь всех используемых в корпусе грамматических помет. Для пояснения в скобках даются примеры.

Части речи

S – существительное (*яблоны, лошадь, корпус, вечность*)

A – прилагательное (*коричневый, таинственный, морской*)

NUM – числительное (*четыре, десять, много*)

A-NUM – числительное-прилагательное (*один, седьмой, восьмидесятый*)

V – глагол (*пользоваться, обрабатывать*)

ADV – наречие (*сгоряча, очень*)

PRAEDIC – предикатив (*жаль, хорошо, пора*)

PARENTH – вводное слово (*кстати, по-моему*)

S-PRO – местоимение-существительное (*она, что*)

A-PRO – местоимение-прилагательное (*который, твой*)

ADV-PRO – местоименное наречие (*где, вот*)

PRAEDIC-PRO – местоимение-предикатив (*некого, нечего*)

PR – предлог (*под, напротив*)

CONJ – союз (*и, чтобы*)

PART – частица (*бы, же, пусть*)

INTJ – междометие (*увы, бабушки*)

Значения грамматических категорий

Род:

m – мужской род (*работник, стол*)

f – женский род (*работница, табуретка*)

m-f – «общий род» (*задира, пьяница*)

n – средний род (*животное, озеро*)

Одушевленность:

anim – одушевленность (*человек, ангел, утопленник*)

inan – неодушевленность (*рука, облако, культура*)

Число:

sg – единственное число (*яблоко, гордость*)

pl – множественное число (*яблоки, ножницы, детишки*)

Падеж:

nom – именительный падеж (*голова, сын, степь, сани, который*)

gen – родительный падеж (*головы, сына, степи, саней, которого*)

dat – дательный падеж (*голове, сыну, степи, саням, которому*)

acc – винительный падеж (*голову, сына, степь, сани, который/которого*)

ins – творительный падеж (*головой, сыном, степью, санями, которым*)

loc – предложный падеж (*[о] голове, сыне, степи, санях, котором*)

gen2 – второй родительный падеж (*чашка чаю*)

acc2 – второй винительный падеж (*постричься в монахи; по два человека*)

loc2 – второй предложный падеж (*в лесу, на осí*)

voc – звательная форма (*Господи, Серёж, ребят*)

adnum – счётная форма (*два часа́, три шара́*)

Краткая/полная форма:

brev – краткая форма (*высок, нежна, прочны, рад*)

plen – полная форма (*высокий, нежная, прочные, морской*)

Степень сравнения:

comp – сравнительная степень (*глубже*)

comp2 – форма «*по*+сравнительная степень» (*поглубже*)

supr – превосходная степень (*глубочайший*)

Вид:

pf – совершенный вид (*пошёл, встречу*)

ipf – несовершенный вид (*ходил, встречаю*)

Переходность:

intr – непереходность (*ходить, вариться*)

tran – переходность (*вести, варить*)

Залог:

act – действительный залог (*разрушил, разрушивший*)

pass – страдательный залог (только у причастий: *разрушаемый, разрушенный*)

med – медиальный, или средний залог (глагольные формы на *-ся*: *разрушился* и т.п.)

Форма (репрезентация) глагола:

inf – инфинитив (*украшать*)

partcp – причастие (*украшенный*)

ger – деепричастие (*украшая*)

Наклонение:

indic – изъявительное наклонение (*украшаю, украшал, украшу*)

imper – повелительное наклонение (*украшай*)

imper2 – форма повелительного наклонения 1 л. мн. ч. на *-те* (*идемте*)

Время:

praet – прошедшее время (*украшали, украшавший, украсив*)

praes – настоящее время (*украшаем, украшающий, украшая*)

fut – будущее время (*украсим*)

Лицо:

1p – первое лицо (*украшаю*)

2p – второе лицо (*украшаешь*)

3p – третье лицо (*украшает*)

Прочие признаки:

persn – личное имя (*Иван, Дарья, Леопольд, Эстер, Гомер, Маугли*)

patrn – отчество (*Иванович, Павловна*)

famn – фамилия (*Николаев, Волконская, Гумбольдт*)

0 – несклоняемое (*шоссе, Седых*)

Часть указанных помет (а именно, второй винительный падеж, звательная форма, счётная форма, форма по+сравнительная степень, общий род, переходность, несклоняемость) присутствуют только в корпусе со снятой грамматической омонимией.

2.1.3. Множественные разборы

В отдельных случаях в морфологической разметке допускается указание у одной и той же словоформы нескольких разборов, а именно:

Для глагольных словоформ, входящих в видовые пары, указываются исходные формы обоих видов (например, форма *пришёл* считается принадлежащей и лексеме ПРИЙТИ, и лексеме ПРИХОДИТЬ).

Для глагольных форм в тех случаях, когда существуют лексемы с *-ся* и без *-ся*, указываются обе исходные формы (например, форма *разрушается* считается принадлежащей и лексеме РАЗРУШАТЬСЯ, и лексеме РАЗРУШАТЬ).

Для прилагательных, совпадающих с причастиями (*открытый*), в качестве исходной дается как лексема-прилагательное (ОТКРЫТЫЙ), так и глагол (ОТКРЫТЬ).

Ставится множественная помета в случаях, когда однозначный выбор лексемы или грамматического значения в данном контексте невозможен (*не видел родного отца* – **gen/acc**; *манекену* – **anim/inan**; *спазмами* – исходная форма СПАЗМ/СПАЗМА и т. п.)

В ряде случаев допускается множественная помета части речи для союзов/частиц типа *словно*, для *-o/-e*-форм типа *хорошо* (предикатив/наречие/прилагательное), для субстантивированных адъективов типа *всё, военный* (существительное/прилагательное), для форм *его, её, их* (притяжательное/личное местоимение); число случаев частеречной неоднозначности, по мере работы над корпусом будет уменьшаться.

2.1.4. Информация о нестандартности и особенностях записи

В Корпусе со снятой грамматической омонимией предусмотрен ряд помет, указывающих на нестандартность или особенности записи входящей в Корпус словоформы. Отсутствие таких особенностей обозначается пометой normal.

anom («Аномальная форма») – различного рода морфологические аномалии, возможные у устаревших или просторечных нелитературных форм

(*три дни* при нормативном *три дня*, *ляжь* при нормативном *ляг*)
distort («Искаженная форма») – орфографическое и/или фонетическое искажение слова, часто передающее различные особенности произношения

(*дэвушка*, *това'ищи*, *про-хо-ди*, *низнаю*).
cipr («Цифровая запись») – запись числительного, числительного-прилагательного или прилагательного (полностью или частично) при помощи цифр (73, LXXIII, 73-й, 22-летний). Для этих словоформ в поле «Лексема» также употребляется цифровая запись; число и падеж указываются только в тех случаях, когда выписано окончание (типа 14-му).

INIT («Инициал») – запись вида «заглавная буква с точкой» (M., P.). В поле «Лексема» инициал не раскрывается; грамматические признаки не указываются.
abbr («Сокращение») – сокращенная запись (тов., гг., ч.). В поле «Лексема» сокращение (кроме инициалов) раскрывается, указывается грамматическая форма, соответствующая контексту. Специально отметим, что акронимы вроде ООН, вуз и усеченные слова вроде зав, зам, записываемые без точки и не раскрываемые при чтении, не получают пометы abbr и трактуются как обычные слова (склоняемые или несклоняемые).

Кроме того, в корпусе с неснятой грамматической омонимией используется особая помета (bastard) для несловарной формы (не входящей в словарь автоматического анализатора, а порожденной по аналогии, например, форма вроде Махабхарата получает несколько гипотетических разборов, в том числе от псевдолексем махабхаронок, махабхарать и т. п.); по мере пополнения словаря анализатора число таких форм будет уменьшаться. С целью снижения «шума» при поиске по корпусу с неснятой грамматической омонимией иногда бывает целесообразно исключить поиск по подобным формам; для ряда задач, напротив, можно ограничить поиск именно ими.

2.1.5. О лексико-семантической информации в Корпусе

В настоящее время в Корпусе реализована система поиска по лексико-семантическим признакам, основанная на частичной семантической разметке текстов. [100]

При такой разметке большинству слов в тексте приписывается один или несколько семантических и словообразовательных признаков, например, 'лицо', 'вещество', 'пространство', 'скорость', 'движение', 'обладание', 'свойство человека', 'дими́нутив', 'отглагольное имя' и т.п. Используется фасетная классификация, при которой одно слово может попадать в несколько классов. На первом этапе поиск осуществляется по части имеющихся в словаре признаков.

Разметка текстов осуществляется автоматически с помощью программы Semmarkup в соответствии с Семантическим словарем Корпуса. Поскольку ручная обработка семантически размеченных текстов очень трудоемка, семантическая омонимия в Корпусе не снимается: многозначным словам приписывается несколько альтернативных наборов семантических признаков.

В основу семантической разметки положена система классификации русской лексики, принятая в базе данных "Лексикограф", которая разрабатывалась с 1992 г. Для нужд Корпуса был существенно увеличен словник, расширен состав и усовершенствована структура семантических классов, добавлены словообразовательные признаки.

Словник семантического словаря базируется на морфологическом словаре системы DIALING (общим объемом порядка 120 тыс. слов), представляющим собой расширение Грамматического словаря русского языка А. А. Зализняка. Текущая версия семантического словаря включает слова знаменательных частей речи: существительные, прилагательные, числительные, местоимения, глаголы и наречия.

При работе над семантической разметкой использовались сведения о значении слов и структуре семантических классов из следующих источников:

Словарь русского языка под ред. С. И. Ожегова

Словарь русского языка под ред. С. И. Ожегова и Н. Ю. Шведовой

Словарь русского языка в 4-х тт. под ред. А. П. Евгеньевой (МАС)

Словарь русского языка в 17-ти тт. (БАС)

Толковый словарь русского языка Д. Н. Ушакова

Русский семантический словарь под ред. Н. Ю. Шведовой, тт. 1-3

Толковый словарь русских глаголов под ред. Л. Г. Бабенко

Системный семантический словарь русского языка Л. М. Васильева

Новый объяснительный словарь синонимов русского языка под общим рук.

акад. Ю. Д. Апресяна (НОСС)

2.1.6. Структура лексико-семантической информации

Лексико-семантическая информация, приписываемая произвольному слову в тексте, состоит из трех групп помет:

разряд (например, имя собственное, возвратное местоимение);

собственно лексико-семантические характеристики (например, тематический класс лексемы, признаки каузативности, оценки);

деривационные (словообразовательные) характеристики (например, «диминутив», «отадъективное наречие»).

Лексико-семантическая информация имеет различную структуру для разных частей речи. Кроме того, каждый из разрядов существительных - имена предметные, непердметные и собственные - имеет свою структуру помет.

Собственно лексико-семантические пометы сгруппированы по следующим полям:

таксономия (тематический класс лексемы) – для имен существительных, прилагательных, глаголов и наречий;

мереология (указание на отношения «часть – целое», «элемент – множество») – для предметных и непердметных имен;

топология (топологический статус обозначаемого объекта) – для предметных имен;

каузация – для глаголов;

служебный статус – для глаголов;

оценка – для предметных и непердметных имен, прилагательных и наречий.

Словообразовательные характеристики включают несколько типов:

морфо-семантические словообразовательные признаки (например, «диминутив», «каритив», «семельфактив»);

разряд производящего слова (например, отглагольное существительное или отадъективное наречие);

лексико-семантический (таксономический) тип производящего слова (например, наречие, образованное от прилагательного размера);

морфологический тип словообразования (субстантивация, сложное слово).

В основу метаязыка лексико-семантических помет, ввиду предполагаемой широкой международной аудитории пользователей Корпуса, положена система сокращенных помет («тегов») на основе англоязычной нотации. В то же время предусмотрена возможность использования при поиске традиционных названий категорий на русском языке (в форме «семантические признаки»). Ниже приводится инвентарь всех доступных в настоящее время для поиска в Корпусе семантических помет. Для пояснения в скобках даются примеры.

Имена существительные (S)

Разряды

r:concr – предметные имена (*девочка, стол, молоко*)

r:abstr – не предметные имена (*вожделение, яркость, время*)

r:prorp – имена собственные (*Иван, Эйнштейн, Петроград*)

Предметные имена

Лексико-семантические пометы

Таксономия:

t:hum – лица (*человек, учитель*)

t:hum:etn – этнонимы (*эфиоп, итальянка*)

t:hum:kin – имена родства (*брат, бабушка*)

t:hum:supernat – сверхъестественные существа (*русалка, инопланетянин*)

t:animal – животные (*корова, жираф, сойка, ящерица, муравей*)

t:plant – растения (*береза, роза, трава*)

t:stuff – вещества и материалы (*вода, песок, тесто, жемчуг, шелк*)

t:space – пространство и место (*космос, город, тайга, овраг, вход*)

t:constr – здания и сооружения (*дом, шалаш, мост*)

t:tool – инструменты и приспособления (*молоток, палка, пуговица, машина*)

t:tool:instr – инструменты (*молоток, штопор, игла, карандаш*)

t:tool:device – механизмы и приборы (*телефон, сеялка, градусник*)

t:tool:transp – транспортные средства (*автобус, поезд, сани*)

t:tool:weapon – оружие (*сабля, пистолет, гаубица*)

t:tool:mus – музыкальные инструменты (*рояль, скрипка, колокол*)

t:tool:furn – мебель (*стол, диван, шкаф*)

t:tool:dish – посуда (*чашка, кастрюля, фляжка*)

t:tool:cloth – одежда и обувь (*платье, шляпа, ботинки*)

t:food – еда и напитки (*пирог, каша, молоко*)

t:text – тексты (*рассказ, книга, афиша*)

Мереология:

pt:part – части (*верхушка, кончик, половина*)

pt:partb & pc:hum – части тела и органы человека (*голова, сердце, ноготь*)

pt:partb & pc:animal – части тела и органы животных (*хвост, жало*)

pt:part & pc:plant – части растений (*лист, ветка, корень*)

pt:part & pc:constr – части зданий и сооружений (*комната, дверь, арка*)

pt:part & pc:tool – части приспособлений (*деталь, лопасть, крышка*)

pt:part & pc:tool:instr – части инструментов (*топорище, лезвие*)

pt:part & pc:tool:device – части механизмов и приборов (*дисплей, корпус, кнопка*)

pt:part & pc:tool:transp – части транспортных средств (*руль, колесо, капот*)

pt:part & pc:tool:weapon – части оружия (*дуло, курок, эфес*)

pt:part & pc:tool:mus – части музыкальных инструментов (*струна, гриф*)

pt:part & pc:tool:furn – части предметов мебели (*сиденье, подлокотник*)

pt:part & pc:tool:dish – части предметов посуды (*носик, горлышко*)

pt:part & pc:tool:cloth – части одежды и обуви (*рукав, каблук*)

pt:qtm – кванты и порции вещества (*капля, комок, порция*)

pt:set | pt:aggr – множества и совокупности объектов (*набор, букет, мебель, человечество*)

hi:class – имена классов (*животное, ягода, инструмент*)

Топология:

top:contain – вместилища (*кошелек, комната, озеро, ниша*)

top:horiz – горизонтальные поверхности (*пол, площадка*)

Оценка:

ev – оценка (неопределенная по признаку «положительная/отрицательная»)
(*озорник, махина*)

ev:posit – положительная (*умница, светило*)

ev:neg – отрицательная (*негодяй, вертихвостка*)

Словообразовательные пометы

d:dim – диминутивы (*зайчик, коробочка*)

d:aug – аугментативы (*детина, домище*)

d:sing – сингулятивы (*пылинка, изюминка*)

d:nag – nomina agentis (*писатель, создатель, докладчик*)

d:fem – nomina feminina (*немка, генеральша, доярка*)

Непредметные имена

Лексико-семантические пометы

Таксономия (тематический класс):

t:move – движение (*беготня, вынос, качка*)

t:move:body – изменение положения тела, части тела (*поклон*)

t:put – помещение объекта (*размещение, расстановка, погрузка, намотка*)

t:impact – физическое воздействие (*удар, втирание, обмолот*)

t:impact:creat – создание физического объекта (*лепка, отливка, плетение, сооружение, строительство*)

t:impact:destr – уничтожение (*слом, сожжение*)

t:changest – изменение состояния или признака (*укрепление, затвердение, осушение, конденсация, осложнение*)

t:be – бытийная сфера

t:be:exist – существование (*жизнь, наличие, бытие*)

t:be:appear – начало существования (*возникновение, рождение, формирование, учреждение, творение*)

t:be:disapp – прекращение существования (*смерть, казнь, ликвидация*)

t:loc – местонахождение (*местоположение*)

t:loc:body – положение тела в пространстве (*лежание*)

t:contact – контакт и опора (*прикосновение, объятие*)

t:poss – посессивная сфера (*обладание, приобретение, покупка, потеря, лишение*)

t:ment – ментальная сфера (*знание, абстракция, воображение, воспоминание, догадка*)

t:perc – восприятие (*осязание, слух, видимость, взгляд, зрелище*)

t:psych – психическая сфера (*апатия, безумие, вдохновение, спокойствие*)

t:psych:emot – эмоция (*восторг, раскаяние, печаль*)

t:psych:volit – воля (*намерение, решение*)

t:speech – речь (*дискуссия, молва, ахинея, реплика, подковырка*)

t:physiol – физиологическая сфера (*жажда, кровоизлияние, судорога, утомление, икота*)

t:weather – природное явление (*зарница, вьюга, зной*)

t:sound – звук (*шум, перезвон, хлопок, аплодисменты, диссонанс*)
t:color – цвет (*окраска, колорит, желтизна, прозелень*)
t:light – свет (*луч, полумрак, светлынь, иллюминация*)
t:taste – вкус (*вкуснота, горчинка, кислятина*)
t:smell – запах (*аромат, перегар*)
t:temper – температура (*прохлада, стужа, нагрев*)
t:time – время (*весна, годовщина, минута, современность*)
t:time:period – период (*межсезонье, путина, сенокос, стаж*)
t:time:moment – момент (*миг, мгновение*)
t:time:week – день недели (*понедельник*)
t:time:month – месяц (*январь*)
t:time:age – возраст (*детство, молодость, двадцатилетие*)
t:humq – свойство человека (*порядочность, безволие, остроумие*)
t:behav – поведение и поступки человека (*разгильдяйство, подхалимаж, неповиновение, ребячество, предательство*)
t:inter – взаимодействие и взаимоотношение (*взаимопомощь, вражда, схватка, драка*)
t:action – мероприятие (*аукцион, вернисаж, вечеринка, выборы, именины, заседание, культпоход*)
t:disease – болезнь (*ангина, диабет*)
t:game – игра (*жмурки, покер, домино, волейбол*)
t:sport – спорт (*спартакиада, акробатика, баскетбол*)
t:param – параметр (*высота, грузоподъемность*)
t:unit – единица измерения (*балл, килограмм, метр, минута*)

Мереология:

pt:part – часть (*начало, финал*)
pt:qtm – квант (*оборот, прыжок, кивок*)
pt:set – множество (*система, выборка, алгоритм*)

Оценка:

ev – оценка (*озорник, махина*)
ev:posit – положительная (*благоухание, загляденье, изюминка*)
ev:neg – отрицательная (*безвкусица, ахиня*)

Словообразовательные пометы

der:v – отглагольные имена (*выбор, демонстрация*)

der:a – отадъективные имена (*краснота, жадность*)

Имена собственные

Лексико-семантические пометы

Таксономия:

t:hum | t:hum:supernat – лица (*Людмила, Черномор*)

t:persn – имена (*Александр*)

t:patrn – отчества (*Сергеевич*)

t:famn – фамилии (*Пушкин*)

t:topon – топонимы (*Европа, Волга, Эльбрус, Москва, Преображенка*)

Словообразовательные пометы

d:dim – диминутивы (*Саша, Женечка, Николаич*)

Имена прилагательные (A)

Разряды

r:qual – качественные (*хороший, большой*)

r:rel – относительные (*деревянный, лунный*)

r:poss – притяжательные (*божий, отцов, мужнин*)

r:invar – неизменяемые (*беж, джерси*)

Лексико-семантические пометы

t:size – размер (*высокий, короткий*)

t:size:max – большой (*высокий, длинный*)

t:size:min – малый (*низкий, короткий*)

t:size:abs – абсолютный (*двухэтажный*)

t:dist – расстояние (*далекий, соседний*)

t:dist:max – большое (*дальний, отдаленный*)

t:dist:min – малое (*близкий, недалекий*)

t:quant – количество (*большой, достаточный, трехкратный*)

t:quant:max – большое (*обильный, многочисленный*)

t:quant:min – малое (*ничтожный, малочисленный*)

t:quant:abs – абсолютное (*двухтысячный, восьмимиллионный*)

t:place – место (*левый, придорожный, теменной*)

t:dir – направление (*обратный, подветренный*)

t:time – время (*прошлый, ночной*)

t:time:dur – длительность (*долгий, краткий*)

t:time:dur:max – большая (*долгий, продолжительный*)
t:time:dur:min – малая (*краткий, кратковременный*)
t:time:dur:abs – абсолютная (*восьмичасовой*)
t:time:age – возраст (*зрелый*)
t:time:age:max – большой (*старый, древний*)
t:time:age:min – малый (*молодой, малолетний*)
t:time:age:abs – абсолютный (*трехлетний*)
t:speed – скорость (*проворный*)
t:speed:max – большая (*скорый, быстрый*)
t:speed:min – малая (*медленный, тягучий*)
t:physq – физические свойства (*мягкий, вязкий*)
t:physq:form – форма (*кривой, круглый*)
t:physq:color – цвет (*красный, бесцветный*)
t:physq:taste – вкус (*кислый, приторный*)
t:physq:smell – запах (*ароматный, тухлый*)
t:physq:temper – температура (*горячий, ледяной*)
t:physq:weight – вес (*тяжелый, легкий*)
t:humq – качества человека (*умный, верный, ловкий*)

Оценка:

ev – оценка (*толковый, мешковатый*)
ev:posit – положительная (*везучий, ладный*)
ev:neg – отрицательная (*продажный, сварливый*)

Словообразовательные пометы

d:dim – диминутивы (*тихонький, крохотный*)
d:aug – аугментативы (*здоровенный, злющий*)
d:atten – аттенуативы (*угловатый, жуликоватый*)
d:habit – хабитивы (*глазастый, пузатый*)
d:carit – каритивы (*безглазый, бездыханный*)
d:potent | d:impot – потенциальные (*плавучий, недееспособный*)
d:potent – POSSIBILITY – (*плавучий, плодородный, занимательный*)
d:impot – IMPOSSIBILITY – (*несоизмеримый, недееспособный*)
der:s – отыменные прилагательные (*домашний, железный*)
der:v – отглагольные прилагательные (*ковкий, навязчивый, кочевой*)
der:adv – отадвербиальные прилагательные (*поздний, здешний*)

Имена числительные (NUM, A-NUM)

Разряды:

г:card – количественные (*два, пять, десять*)

г:card:pauc – числительные малого количества (*два, три, четыре, оба, пол, полтора*)

г:ord – порядковые (*первый, второй, десятый*)

Местоимения, в том числе:

S-PRO – местоимения-существительные (*он, кто*)

A-PRO – местоимения-прилагательные (*его, какой*)

ADV-PRO – местоимения-наречия (*где, как*)

Разряды:

г:pers – личные (*я, он*)

г:ref – возвратные (*себя*)

г:poss – притяжательные (*мой, его, свой*)

г:rel – вопросительные/относительные (*кто, который, когда*)

г:dem – указательные (*этот, такой*)

г:indet – неопределенные (*некоторый, некогда*)

г:neg – отрицательные (*никакой, ничей*)

г:spes – кванторные (определятельные) (*всякий, каждый, любой*)

Глаголы (V)

Лексико-семантические пометы:

t:move – движение (*бежать, дергаться, бросить, нести*)

t:move:body – изменение положения тела, части тела (*согнуть, нагнуться, примоститься*)

t:put – помещение объекта (*положить, вложить, спрятать*)

t:impact – физическое воздействие (*бить, колоть, вытирать*)

t:impact:creat – создание физического объекта (*выковать, смастерить, сшить*)

t:impact:destr – уничтожение (*взорвать, сжечь, зарезать*)

t:changest – изменение состояния или признака (*взрослеть, богатеть, расширить, испачкать*)

t:be – бытийная сфера (*жить, возникнуть, убить*)

t:be:exist – существование (*жить, происходить*)

t:be:appear – начало существования (*возникнуть, родиться, сформировать, создать*)

t:be:disapp – прекращение существования (*умереть, убить, улетучиться, ликвидировать, искоренить*)
t:loc – местонахождение (*лежать, стоять, положить*)
t:loc:body – положение тела в пространстве (*сидеть*)
t:contact – контакт и опора (*касаться, обнимать, облокотиться*)
t:poss – посессивная сфера (*иметь дать, подарить, приобрести, лишиться*)
t:ment – ментальная сфера (*знать, верить, догадаться, помнить, считать*)
t:perc – восприятие (*смотреть, слышать, нюхать, чуют*)
t:psych – психическая сфера (*гипнотизировать, сочувствовать, настроиться, терпеть*)
t:psych:emot – эмоция (*радоваться, обидеть*)
t:psych:volit – воля (*решить*)
t:speech – речь (*говорить, советовать, спорить, каламбурить*)
t:behav – поведение человека (*куролесить, привередничать*)
t:physiol – физиологическая сфера (*кашлять, икать*)
t:weather – природное явление (*бушевать, вьюжить*)
t:sound – звук (*гудеть, шелестеть*)
t:light – свет (*гаснуть, лучиться*)
t:smell – запах (*пахнуть, благоухать*)

Каузация:

ca:caus – каузативные глаголы (*показать, вертеть*)
ca:noncaus – некаузативные глаголы (*видеть, вертеться*)

Служебные глаголы:

aux:phase – фазовые (*начать, продолжать, прекратить*)
aux:caus – служебные каузативные (*вызвать, привести (к)*)

Словообразовательные пометы

d:pref – приставочные глаголы (*забегать, оглядеть*)
d:semelf – семельфактивы (*кивнуть, чихнуть, боднуть, качнуться*)
d:impf – вторичные имперфективы (-ива-, -ва-, -а-) (*выпивать, вбивать, прогонять*)

Наречия (ADV)

Лексико-семантические пометы:

t:place – место (*здесь, посередине*)
t:dir – направление (*туда, вверх*)

t:dist – расстояние (*далеко, близко*)
t:dist:max – большое (*далеко, вдали, вдалеке*)
t:dist:min – малое (*близко, вблизи*)
t:time – время (*тогда, поздно*)
t:time:dur – длительность (*вечно, недолго*)
t:time:dur:max – большая (*вечно, подолгу, всегда*)
t:time:dur:min – малая (*временно, недолго*)
t:speed – скорость (*быстро, медленно*)
t:speed:max – большая (*быстро, мигом*)
t:speed:min – малая (*медленно, неторопливо*)
t:quant – количество (*столько, достаточно*)
t:quant:max – большое (*много, навалом*)
t:quant:min – малое (*мало, чуть-чуть*)

Оценка:

ev – оценка (*беспечно, бойко*)
ev:posit – положительная (*бойко, безупречно*)
ev:neg – отрицательная (*бездарно, неловко*)

Словообразовательные пометы:

d:dim – диминутивы (*немножко, быстренько*)
d:atten – аттенуативы (*рановато, суховато*)
der:s – отыменные наречия (*вверху, дома*)
der:v – отглагольные наречия (*отродясь, стоймя*)
der:a – отадективные наречия (*быстро, обычно*)

Таксономия производящего слова-прилагательного

der:a & dt:size – размер (*высоко, коротко*)
der:a & dt:size:max – большой (*высоко, бесконечно*)
der:a & dt:size:min – малый (*коротко, низко*)
der:a & dt:physq – физические свойства (*твердо, плотно*)
der:a & dt:physq:form – форма (*плоско, прямо*)
der:a & dt:physq:color – цвет (*красно, добела*)
der:a & dt:physq:taste – вкус (*горько, вкусно*)
der:a & dt:physq:smell – запах (*смердно, зловонно*)
der:a & dt:physq:temper – температура (*тепло, прохладно*)

der:a & dt:physq:weight – вес (*тяжело, легко*)

der:a & dt:humq – качества человека (*внимательно, грубо*)

2.1.7. Параметры текстов.

В национальном Корпусе используется сравнительно простая система метаразметки, предназначенная не для специалистов по корпусной лингвистике, работающих с универсальной международной классификацией, а для рядового пользователя (в том числе и для лингвиста, не знакомого с терминологией корпусных исследований). [104]

Интерфейс для упрощенного метатекстового поиска устроен так, что параметры текста объединяются в несколько блоков:

I. «Паспорт текста»

Автор текста: имя, пол, дата рождения (или примерный возраст)

Название текста.

Время создания текста (может указываться точно или приблизительно; при поиске может использоваться формат «<не> раньше такой-то даты» или «<не> позже такой-то даты»).

Объем текста (в количестве слов; при поиске может использоваться формат «<не> более такого-то объема» или «<не> менее такого-то объема»). Дается пояснение: для художественных произведений принято, что обычная длина рассказа — менее 5 тыс. слов; обычная длина повести — от 5 до 15 тыс. слов; обычная длина романа — более 15 тыс. слов.

Блок II состоит из трех поисковых массивов: «нехудожественная проза», «художественная проза», «драматургия». Первые два массива имеют несколько разные структуры параметров, поэтому оформляются по отдельности.

1. Художественные тексты

Жанр текста (включается также помета «нежанровая проза»): историко-приключенческая, криминальная, любовная литература, сатира и юмор, фантастика, и т.п.

Тип текста (при обозначении типа широко используется самоидентификация текста; список типов, в принципе, открытый и дается в окне поиска в алфавитном порядке): анекдот, боевик, детектив, повесть, притча, рассказ, роман, сказка, триллер, эпопея, эссе и т.п.

Хронотоп текста (приблизительное указание на место и время описываемых в тексте событий; включается также помета «хронотоп не определен»); в частности, различается доисторический период, античность, Средние века, Новое время, Россия: 19 век, Россия: 20 век (до 1914), Россия/СССР: война 1914–1918, революция, гражданская война, 20-е гг., 30-е гг., война 1941–1945, послевоенный период (до 1952), 50-е гг., 60–80 гг., перестройка, Россия: постсоветский период. Для художественных текстов указание на хронотоп дается вместо указания на тематику (как более информативное).

2. Нехудожественные тексты

Сфера функционирования текста (параметр призван отражать в первую очередь языковые особенности): бытовая, официально-деловая, производственно-техническая, публицистическая, учебно-научная, церковно-богословская.

Тип текста (при обозначении типа широко используется самоидентификация текста; список типов в принципе открытый и дается в окне поиска в алфавитном порядке; включается также помета «тип не определен»): автобиография, акт, дневник, договор, документ, закон, заметка, заявление, инструкция, информационное сообщение, кодекс, комментарий, листовка, обзор, объявление, отзыв, отчет, очерк, письмо, постановление, проповедь, путеводитель, резюме, реклама, рекомендация, рецензия, рецепт, сочинение, справочник, статья, учебник, характеристика, хроника, эссе, и т.п.

Тематика текста (в принципе, у одного текста тем может быть несколько; список открытый): бизнес, коммерция, экономика, финансы; война и вооруженные конфликты; дом и домашнее хозяйство; здоровье и медицина; зрелища и развлечения; искусство; криминал; наука (по разделам и отраслям); политика и общественная жизнь; право; производство; сельское хозяйство; спорт; природа; частная жизнь и т.п.

При разработке параметров метатекстовой разметки составителями Национального корпуса был учтен мировой опыт, прежде всего опыт составителей Британского национального корпуса. В англоязычной литературе существует целый ряд предложений по классификации текстов для создания представительных корпусов, но было принято решение опираться в основном на рекомендации Дж. Синклера (так называемый стандарт EAGLES, принятый во

многих современных системах автоматической обработки текстов). Эти рекомендации были адаптированы к русскому материалу С. А. Шаровым и составили первый вариант метаразметки (условно — «международный» вариант, или вариант Синклера-Шарова). В настоящее время ведется работа по внесению этой информации в настоящий Корпус. Их использование в корпусе облегчит сопоставимость результатов метаразметки в русском и других Национальных корпусах и будет наиболее удобно для специалистов по корпусной лингвистике разных стран (не обязательно славистов и русистов).

2.2. Обзор проблемы разрешения синтаксической неоднозначности при морфологической, синтаксической и семантической разметке текстов на естественных языках и методы ее интерактивного разрешения.

Не будет преувеличением сказать, что проблема разрешения языковой неоднозначности по сложности занимает первое место среди всех задач автоматической обработки текстов. Сколь бы тщательно ни были разработаны грамматики такой системы и ее словари, анализирующие модули системы на любом этапе непрерывно сталкиваются с необходимостью выбора вариантов анализа текстового материала – будь-то на уровне морфологии, синтаксиса или семантики. При этом по большому счету неважно, имеем мы дело с реальной омонимией обрабатываемого фрагмента текста или же с неоднозначностью мнимой, возникающей из-за несовершенства действующих в системе правил – тем более, что между этими разновидностями неоднозначности нет резкой границы (то, что представляется неоднозначным в коротком сегменте текста, может получить разрешение в более широком контексте; напротив, вполне недвусмысленное высказывание может оказаться омонимичным при его расширении, ср. однозначное

Я ждал его

и неоднозначное

Я ждал его вчера (обстоятельство «вчера» относится или к процессу ожидания, или к дополнению, выраженному местоимением «его»). [99]

Неудовлетворительная ситуация с разрешением неоднозначности в системах автоматической обработки текста носит универсальный характер и, вообще говоря, не зависит от лингвистической модели, лежащей в основе текстового анализатора. Удивляться здесь не приходится: то, что с легкостью делает человек в процессе понимания текста, опираясь при выборе интерпретации неоднозначных элементов на здравый смысл, знания о мире и широкий контекст коммуникации, пока недоступно никаким компьютерным системам. Информация о мироустройстве, по-видимому, вообще не поддается сколько-нибудь масштабной формализации. Да и степень формализации чисто языковой семантики, достигнутая в компьютерной лингвистике, пока далеко не достаточна для того, чтобы эксплицировать те нетривиальные сведения о смысле высказывания, которыми необходимо располагать для разрешения неоднозначности текста. Сказанное, однако, отнюдь не означает, что попытки решить проблему неоднозначности при автоматическом переводе вообще лишены перспективы. Частичное решение этой проблемы вполне возможно, и всякая система анализа естественного языка располагает арсеналом более или менее действенных средств, направленных на сокращение неоднозначности в ходе обработки текста – от простого игнорирования редких лексических единиц или синтаксических конструкций до использования масштабных статистических процедур, определяющих частотность встречаемости отдельных языковых элементов.

На данный момент наиболее эффективным является метод интерактивно-диалогового синтаксического разрешения неоднозначности. Ведется диалог, позволяющий пользователю выбирать между синтаксическими гипотезами, имеющими вид бинарных поддеревьев. Такой метод особенно эффективен, если разрешение лексической и синтаксической неоднозначности производится для предложения одновременно. [93]

Чтобы нагляднее охарактеризовать суть комбинированного алгоритма синтаксического анализа, включающего компонент корпусной статистики, коротко напомним основные стадии главного, правилowego алгоритма, используемого в системе ЭТАП-3. В процессе СинТА текста ЭТАП-3 использует результаты морфологического и предсинтаксического анализа. Единицей СинТА всегда является предложение. Морфологический анализ в ЭТАПе-3

осуществляется без учета какого-либо контекста. Это означает, например, что слово дуло в предложении

Из щели дуло

будет представлено в морфологической структуре как пара объектов

{(ДУТЬ, прош, несов, сред, ед);

(ДУЛО, ед, род)}. [86]

На стадии предсинтаксического анализа происходит частичное разрешение лексической и грамматической омонимии по линейному контексту. В частности, в данном предложении блок предсинтаксического анализа оставит для дуло лишь глагольную интерпретацию, а именная интерпретация будет исключена. Результатом работы блока собственно синтаксического анализа является построение дерева зависимостей для исходного предложения, в котором каждый узел соответствует одному, и только одному слову предложения, а все ветви помечены именами синтаксических отношений. На начальной стадии этого блока синтаксические правила, или синтагмы, порождают множество минимальных поддеревьев (два узла, связанные синтаксическим отношением) – своего рода кирпичиков, из которых будет построено все дерево зависимостей. Затем осуществляется выбор вершины будущего дерева зависимостей и его непосредственное построение. На этой стадии происходит снятие оставшейся после предсинтаксического анализа лексической и грамматической неоднозначности, а также разрешение синтаксической неоднозначности (иными словами, удаление части минимальных поддеревьев). Здесь используется система различных фильтров, центральную роль в которой играет механизм приоритетов. В этот момент в действие вступает новый статистический блок, играющий роль дополнительного фильтра. Основные идеи, легшие в основу его создания, были предложены в работе Чардин (2001). Статистический блок взвешивает минимальные поддеревья, а также цепочки минимальных поддеревьев длиной в три слова в пространстве поиска алгоритма синтаксического анализа на основании частоты встречаемости фрагментов такого вида в деревьях зависимостей корпуса. В итоге в систему возвращаются новые значения приоритетов связей, которые вычисляются с учетом вновь приписанных весов. Цепочки длиной в три слова взвешиваются в случаях, когда между двумя словами сформированы две или более альтернативные синтаксические связи и у них нет альтернативы на один уровень выше по

дереву. Если такие альтернативы есть, то взвешиваются минимальные поддеревья. Такой подход связан с тем, что в противном случае пришлось бы учитывать условную вероятность полученных приоритетов, что привело бы к чрезмерному усложнению работы. Цепочки минимальных поддеревьев можно рассматривать как n -граммы 2-го и 3-го порядка или, соответственно, древесные биграммы и триграммы. Однако требование приписывания весов конкретным связям не позволяет в полной мере реализовать преимущества n -граммной модели. Значения статистических приоритетов связей заведомо меньше единицы (поскольку за единицу принимается вся совокупность встречаемостей конкретных конфигураций в корпусе), в то время как значение положительного приоритета связи, присваиваемого регулярными правилами, целочисленное и всегда не меньше единицы. Благодаря этому обстоятельству, при конфликте между эвристической и статистической стратегиями, в механизме учета приоритетов предпочтение отдается результатам работы эвристических правил, созданных экспертами-лингвистами. Следует отметить, что при наличии соответствующих корпусных данных предлагаемый комбинированный алгоритм может использоваться при СинтА предложений не только на русском, но и на английском языке как, впрочем, и на любых других языках, анализ которых осуществляется на данных принципах. [84]

Глава 3. Лингвистические основы формального описания юридически-правовой документации на ограниченном естественном грузинском языке

3.1. Формальное описание флексивного словоизменения существительных грузинского языка.

Построение полностью формализованного описания раздела юридически-правовой документации на ограниченном естественном грузинском языке исходит из того, что согласно принятым в гл.1 предположениям, в текстах формулировок гражданского иска наиболее часто встречающимися частями речи являются имена существительные и прилагательные. Исходя из этого, модель морфологического анализа должна обеспечивать точное решение задачи морфологического разбора заданной именной словоформы, в том числе выделение основы и определение падежа и/или послелога (постпозиции).

В работе использован современный подход к описанию полной парадигмы грузинских существительных, основанный на применении метасинтаксической нотации Бэкуса-Наура (по существу являющейся видоизменением контекстно-свободных грамматик Хомского). В частности, нами используется известная форма грамматик описания парадигм грузинских существительных, изложенная в работах Э.Годуадзе. [42]

Одним из спорных вопросов при описании грамматик естественных языков является выбор класса формального языка в соответствии с известной классификацией Хомского. С точки зрения точной теории, для этого достаточен уровень грамматики, эквивалентной некоторому конечному автомату. Однако на практике представление формальной модели естественного языка в виде контекстно-свободной грамматики, хотя и является, строго говоря, избыточным, обладает значительными преимуществами с точки зрения наглядности изложения формального аппарата, удобства проверки ее адекватности и внесения модификации, неизбежно возникающих при расширении лексики, охватываемой моделью, тем более, если принять во внимание традиции изложения формальных языков, накопленные в литературе по языкам программирования. В нашей работе принято изложение грамматики

имен существительных грузинского языка алгебраическими средствами на уровне контекстно-свободной модели.

3.1.1. Алгебраическое описание флексивного формоизменения имен грузинского языка состоит из следующих основных разделов:

- классификация именных форм по происходящим в них морфонологическим изменениям основы и распределению падежных окончаний;

- определение простых (нерасширенных) морфологических признаков для различных видов грамматического числа;

- определение используемых для данной основы послелогов (постпозиций), исходя из их морфологических особенностей;

- определение расширенных (имеющих эмфатические частицы) словоформ в их связи с нерасширенными формами;

- классификация именных форм с частицами;

- определение флексивных форм с помощью частицы, являющейся краткой формой вспомогательного глагола «Быть» ("ღ") (в юридическо-правовой документации встречается только в порядке исключения);

- определение флексивных форм с помощью частицы косвенной речи "ო" (в юридическо-правовой документации встречается только в порядке исключения);

Изложение модели, согласно традиции построения нотации Бэкуса-Наура, состоит из метасинтаксических продукционных правил получения значения нетерминалов (в смысле Хомского), в правых частях которых содержится определяемый нетерминал, а в левой - список альтернатив для их сведения к конкатенации с другими нетерминалами и/или литералами грамматики, текстуально определяющими основы и совокупности падежных окончаний, послелогов и частиц (в терминологии Хомского – терминальные символы грамматики). [79]

3.1.2. Основы имен существительных и прилагательных грузинского языка в предлагаемой модели разбиты на несколько групп, в соответствии с их морфонологическими свойствами. Для каждой группы в продукционных правилах приводится только по одной представительной основе.

Предполагается, что в базе данных основ каждая из них классифицирована в соответствии с ее морфонологическими свойствами и снабжена признаком, определяющим ее конкретную парадигму.

- общие (неличные) консонантные основы (ConsStem);
- консонантные основы личных имен (ConsStemPerson);
- общие (неличные) вокальные односложные основы (VocStemShrA) или многосложные основы, испытывающие акопию последней гласной (VocStemLngA):
 - общие (неличные) вокальные односложные неакопируемые основы (VocStemShrB) или многосложные неакопируемые основы (VocStemLngB);
 - вокальные основы, испытывающие симультанное изменение основы (сочетание синкопии и акопии) (VocStemSimultLng);
 - вокальные основы, завершающиеся на неакопируемую в единственном числе гласную "ო" (VocStemI);
 - вокальные основы личных имен (VocStemPerson);

Ввиду необходимости учета возможных фонетических изменений основы в различных частях ее парадимы, в продукционных правилах она приводится в трех формах: номинативной единственного числа, генитивной единственного числа и множественного числа, образованного суффиксом «*ებ*». С учетом изложенного, продукционные правила для определения нетерминалов, представляющих основы, имеют следующий вид:

- <ConsStem>::=*კაც/კედელ/მინდორ/ხუცეს*
- <ConsStemGen>::=*კაც/კედელ/მინდვრ/ხუც*
- <ConsStemPlur>::=<ConsStemGen>
- <ConsStemPerson>::=*დავით*
- <ConsStemPersonGen>::=<ConsStemPerson>
- <VocStemShrA>::=*ძმა/ხე*
- <VocStemShrCenA>::=*ძმ/ხ*
- <VocStemShrPlurA>::=*ძმ/ხე*
- <VocStemShrB>::=*ბუ/მზა*
- <VocStemShrGenB>::=<VocStemShrB>
- <VocStemShrPlurB>::=*ბუ/მზ*

- <VocStemLngA>::=*ღარაბა|სიცოცხლე|მხარე*
- <VocStemLngGenA>::=*ღარაბ|სიცოცხლ|მხრ*

<VocStemLngPlurA>::=ღარაბ/სიცოცხლე/მხარე

<VocStemLngB>::=სატრფო/გოგონა

<VocStemLngGenB>::=<VocStemLngB>

<VocStemLngPlurB>::=სატრფო/გოგონ

<VocStemI>::=პონი

<VocStemIGen>::=<VocStemI>

<VocStemIPlur>::=პონ

<VocStemSimultLng>::=ქახანა

<VocStemSimultLngGem>::=ქახან

<VocSimultLngPlus>::=<VocSimultLngGen>

<VocStemPerson>::=გაჟა

<VocStemPersonGen>::=<VocStemPerson>

3.1.3. Следующий важнейший нетерминал (PlurStem) описывает основу, которая в широком смысле представляет все словоформы множественного числа, порожденные суффиксом «*გბ*». В частности, в него входят все основы, которые могут быть получены конкатенацией суффикса «*გბ*» как с консонантами, так и с вокальными исходными основами, представленными соответствующими нетерминалами:

<PlurStem>::=<ConsStemPlur>*გბ*|

<VocStemPlur>*გბ*|<VocStemIPlur>*გბ*|

<VocStemSimultLngPlur>*გბ*

3.1.4. В число тех терминалов, которые описывают формы склонения по падежам, входят Nominative, Ergative, Dative, Genitive, Instrum, Adverbial, Vocative. Форма каждого падежа получается путем конкатенации соответствующих нетерминалов основы (они могут быть в исходном или модифицированном варианте) с терминалом, представляющим падежное окончание.

Множественное число описывает составной нетерминал (PlurStem), который, с точки зрения продукционных правил образования падежных форм, не отличается от нетерминала, представляющего консонантную основу (ConsStem), поскольку они неотличимы с точки зрения образования парадигмы.

Формы эргативного и адвербиального падежа делятся на два класса: завершающиеся гласной или согласной. Ниже приводятся иллюстративные примеры нетерминалов, соответствующие эргативному падежу:

$$\begin{aligned} \langle \text{Ergative-C} \rangle &::= \langle \text{VocStem} \rangle \mathcal{A} \\ &\quad \langle \text{VocStemI} \rangle \mathcal{A} \\ &\quad \langle \text{VocStemSimultLng} \rangle \mathcal{A} \\ &\quad \langle \text{VocStemPerson} \rangle \mathcal{A} \\ \langle \text{Ergative-V} \rangle &::= \langle \text{ConsStem} \rangle \mathcal{A} \mathcal{S} \\ &\quad \langle \text{PlurStem} \rangle \mathcal{A} \mathcal{S} \\ &\quad \langle \text{ConsStemPerson} \rangle \mathcal{A} \mathcal{S} \\ \langle \text{Ergative} \rangle &::= \langle \text{Ergative-C} \rangle | \langle \text{Ergative-V} \rangle \end{aligned}$$

3.1.5. Отдельного описания требуют нетерминалы, представляющие формы множественного числа, образованные архаическими суффиксами «бьб-оьб». К их числу относятся нетерминалы ArchPlurNomin, ArchPlurGen, ArchPlurVocat. Например, нетерминал для представления родительного падежа:

$$\begin{aligned} \langle \text{ArchPlurGen} \rangle &::= \langle \text{ConsStem} \rangle \mathcal{O} | \langle \text{VocStem} \rangle \mathcal{O} | \\ &\quad \langle \text{VocStemSimultLng} \rangle \mathcal{O} \end{aligned}$$

3.1.6. Приведенные выше правила продукции объединяются нетерминалом SimplForm, который представляет все именные флективные словоформы, которые не содержат послелогов и частиц:

$$\begin{aligned} \langle \text{SimplForm} \rangle &::= \langle \text{Nominative} \rangle | \langle \text{Ergative} \rangle | \\ &\quad \langle \text{Dative} \rangle | \langle \text{Genetive} \rangle | \langle \text{Instrum} \rangle | \langle \text{Adverbial} \rangle | \\ &\quad \langle \text{Vocative} \rangle | \langle \text{ArchPlurNomin} \rangle | \langle \text{ArchPlurGen} \rangle | \\ &\quad \langle \text{ArchPlurVocat} \rangle \end{aligned}$$

3.1.7. Для описания словоформ, содержащих послелоги (постпозиции), целесообразно ввести понятие стемоида – конкатенации основы, падежного окончания (иногда – пустого) и послелога. Нетерминалы, представляющие стемоиды, могут быть классифицированы следующим образом:

– нетерминалы, представляющие стемоиды, образованные послелогом, требующими именительного падежа (StemoidNomin-C). Стемоид этого типа, т.е.

та часть словоформы, которая непосредственно примыкает к падежному окончанию, завершается на согласную;

– нетерминалы, представляющие стемоиды, образованные послелогоми, требующими дательного, родительного и адвербиативного падежа (StemoidDative, StemoidGen, StemoidAdverb). В зависимости от того, какой литерой завершается стемоид (согласной или гласной), каждый из этих нетерминалов, в свою очередь, разделяется на два определяющих его нетерминала: StemoidDative-V, StemoidDative-C; StemoidGen-V, StemoidGen-C; StemoidAdverb-V, StemoidAdverb-C;

– нетерминалы, представляющие стемоиды, образованные послелогоми, требующими множественного числа, образованного суффиксом «*ꞑꞑ*» (StemoidPlur);

– нетерминалы, представляющие стемоиды, образованные послелогоми, требующими множественного числа, образованного архаическим суффиксом «*ꞑꞑ-ꞑꞑꞑ*» (StemoidGenArch), например, в полном представлении

$$\langle \text{StemoidGenArch} \rangle ::= \langle \text{ConsStem} \rangle | \langle \text{VocStem} \rangle | \\ \langle \text{VocStemSimultLng} \rangle$$

Как видно из приведенных продукционных правил, стемоиды, образованные послелогоми, требующими множественного числа, образованного архаическим суффиксом «*ꞑꞑ-ꞑꞑꞑ*», объединяют как консонантные, так и вокальные основы, принадлежащие общей (неличной) части основ, а также основам симультанного типа склонения, но не включают основ личных имен, а также основ, заканчивающихся на гласную "o".

3.1.8. Учитываемые в модели 13 послелогои распределены на 11 классов согласно тому, какие они требуют падежи и морфонологические изменения основы, а также какое падежное окончание необходимо перед конкретным послелогом. В обозначениях соответствующих нетерминалов, в зависимости от последней литеры послелогоа, используется маркер «V» (послелог завершается на гласную) или «C» (послелог завершается на согласную). Эти классы могут быть перечислены следующим образом:

– послелогои именительного падежа (нетерминал PostPositNominBound-C);

- три класса послелогов дательного падежа (нетерминалы PostpositDativBound-C, PostpositDativFree-C, PostpositDativFree-V);
- три класса послелогов родительного падежа (нетерминалы PostpositGenBounC-C; PostpositGenBoundV-C; PostpositGenBound-V);
- два класса послелогов инструментального падежа (нетерминалы (PostpositInstrBoundC-C, PostpositGenBoundV-C);
- два класса послелогов адвербиативного падежа (нетерминалы (PostpositAdverbBound-C, PastpositAdverbBound-V).

3.1.9. Для описания содержащих послелого словоформ вводятся следующие пять нетерминалов: PostpositFormNomin, PostpositFormDative, PostpositFormGen, PostpositFormInstr, PostpositFormAdverb, каждый из которых образуется через нетерминалы соответствующего стемоида, как с помощью гласной «а» расширенной формы падежного окончания, так и без нее. Например, для стемоидов дательного падежа записываются следующие продукционные правила:

$$\langle \text{PostpositFormDative-C} \rangle ::= \langle \text{StemoidDative-C} \rangle$$

$$\langle \text{PostpositDativeBound-C} \rangle |$$

$$\langle \text{StemoidDative-V} \rangle \mathbf{b} \langle \text{PostpositDativBound-C} \rangle |$$

$$\langle \text{StemoidDative} \rangle \mathbf{b} \langle \text{PostpositDativFree-C} \rangle$$

$$\langle \text{PostpositFormDative-C} \rangle ::= \langle \text{StemoidDative} \rangle$$

$$\langle \text{PostpositDativeFree-V} \rangle$$

$$\langle \text{PostpositFormDative} \rangle ::= \langle \text{PostpositFormDative-C} \rangle |$$

$$\langle \text{PostpositFormDative-V} \rangle$$

3.1.10. Все правила для нетерминалов, представляющие содержащие послелого словоформы, объединяются составным нетерминалом SimpePostpositForm. Он полностью описывает словообразование той части парадигмы грузинских существительных и прилагательных, которая не содержит никаких частиц.

<SimplePositForm>::=<PostpositFormNomin>|
 <PostpositFormDative>|
 <PostpositFormGen>|
 <PostpositFormInstr>|
 <PostpositFormAdverb>

3.1.11. Для формального описания дальнейшей части парадигмы именных словоформ в первую очередь следует ввести нетерминалы, представляющие словоформы, имеющие эмфатическую частицу (гласную) "ღ".

<Emphatic>::=ღ

Правила продукции форм, содержащие эмфатическую частицу, удобно определить через нетерминал, представляющий преэмфатические формы PreEmphaticForm, которые содержат в себе простые формы, завершающиеся на согласную, и простые формы, содержащие послелогои.

<PreEmphaticForm>::=<Ergative-C>|
 <Dative>|<Genetive>|
 <Instrum-C>|<Adverbial>|
 <ArhPlurGen>|
 <PostpositFormNomin-C>|
 <PostpositFormDative-C>|
 <PostpositFormGen-C>
 <PostpositFormInstr-C>|
 <PostpositFormAdverb-C>

3.1.12. Словоформы, содержащие послелогои, представляются с помощью нетерминала EmphaticForm, который определяется при помощи конкатенации нетерминалов PreEmphaticForm и Emphatic:

<EmphaticForm>::=<PreEmphaticForm><Emphatic>

3.1.13. Возможные в именах грузинского языка частицы делятся на три класса:

<Particle-ve-V>::=ʒʒ/ʁʁ

<Particle-ts-C>::=ʒ

<Particle-ts-V>::=ʒʁ

Для представления флективных форм, к которым конкатенируются частицы этих классов, вводятся нетерминалы PreparticleForm-C, PreparticleForm-DatGen, PreparticleForm-V. В частности, первый из них представляет ту часть словоформ, которая полностью совпадает с преэмфатическими словоформами; второй нетерминал выделяет для каждого грамматического числа нераспространенные формы дательного и родительного падежа; третий нетерминал представляет как простые, так и обладающие послелогом нерасширенные формы, завершающиеся на гласную. В частном случае:

<PreParticleForm-V>::=<Nominative>|

<Ergative-V>|

<Instrum-V>|

<ArchPlurNomin>|

<PostPositFormDative-V>|

<PostPositFormGen-V>|

<PostPositFormAdverb-V>

Для представления флективных словоформ, содержащих частицы ʒʒ/ʁʁ, ʒ, ʒʁ, вводятся нетерминалы ParticleForm-C, ParticleForm-V. Первый из них описывает именные словоформы, содержащие частицу "ʒ", а второй – содержащие частицы "ʒʒ/ʁʁ" и "ʒʁ". В частном случае:

<ParticleForm-C>::=<PreParticleForm-V><Particle-ts-C>|

<EmphaticForm><Particle-ts-C>

3.1.14. Для описания словоформ, содержащих частицу "a", являющуюся краткой формой вспомогательного глагола «быть», вводится нетерминал ToBe.

С его помощью может быть введен нетерминал PreToBeForm, представляющий предедеривационные словоформы. В свою очередь, они разделяются на нетерминалы, представляющие завершающиеся на гласные предедеривационные словоформы PreToBeForm-V и на нетерминалы, представляющие завершающиеся на согласные предедеривационные словоформы. В свою очередь, эти нетерминалы являются продукциями от нетерминалов, представляющие словоформы, предшествующие частицам, а также словоформы, которые содержат частицы.

3.1.15. Представление именных словоформ через нетерминал ToBeForm в случае, если предедеривационная форма завершается на согласную, получается при помощи конкатенации с ней дополнительной эмфатической частицы "з" и такой же самой частицы, представляющей краткую форму вспомогательного глагола «быть». В случае, если предедеривационная форма завершается на гласную, представление именных словоформ через нетерминал ToBeForm получается через непосредственную конкатенацию предедеривационной формы с частицей "з", представляющей краткую форму вспомогательного глагола «быть».

<FormToBe> ::= <PreToBeForm-C> <Emphatic> <ToBe> |

<PreToBeForm-V> <ToBe>

3.1.16. Для описания частицы косвенной речи (третьего лица) "ѳ" вводится нетерминал IndirectSpeech вместе с нетерминалом для представления именной словоформы, содержащей частицу косвенной речи FormIndirectSpeech. С помощью этих нетерминалов объединяются продукционные правила для определения преиндиректных словоформ и частиц косвенной речи (не только третьего, но и первого и второго лица).

Что касается нетерминала PreIndirectSpeechForm, его продукционное определение может быть образовано, в первом случае, как нетерминалами, описывающими формы с послелогоми, так и не содержащие их, во втором случае, - как расширенные, так и нерасширенные словоформы, а также содержащие частицы и краткую форму вспомогательного глагола. Это продукционное правило может быть записано в полной форме:

<PreIndirectSpeechForm>::=<SimpleForm>|

<SimplPostpesitForm>|

<EmphaticForm>|

<ParticleForm>|<FormToBe>

<IndirectSpeech>::=∅

<FormIndirectSpeech>::=<PreIndirectSpeechForm><IndirectSpeech>

3.1.17. Для самого общего представления всех элементов парадигмы имен существительных и прилагательных грузинского языка вводится завершающий нетерминал Form (в терминал Пролога – цель грамматического разбора имени), для которого записывается общее продукционное правило, содержащее в качестве альтернатив все приведенные выше продукционные правила для определения нетерминалов, представляющих самостоятельно существующие члены парадигмы. Кроме того, вводится отдельный нетерминал ModifierForm и определяющее его продукционное правило, представляющее формы прилагательного, используемого в качестве грамматического определения и находящегося в т.н. «линейном» падеже (не имеющем падежного окончания):

<ModifierForm>::=<ConsStem>| <VocStem>| <ConsStemPerson>|

<VocStemI>|<VocStemPerson>|<VocStemSimultLng>

<Form>::=<PreIndirectSpeechForm>|<FormIndirectSpeech>|<ModifierForm>

С помощью нетерминала <Form>, совместно со всеми содержащими его определения продукционными правилами (самые типичные из которых перечислены выше), возможно решение задач как полного синтеза словоформ парадигмы грузинского существительного или прилагательного, так и анализа любой заданной словоформы на предмет подтверждения того, что она является корректным словообразованием от некоторой именной основы грузинского языка, или не принадлежит к ним.

3.2. Формальное описание синтаксической структуры синтагмы грузинского языка.

Построение синтаксического раздела полностью формализованного описания юридически-правовой документации на ограниченном естественном грузинском языке опирается на модель морфологического анализа, которая обеспечивает точное решение задачи морфологического разбора заданной именной словоформы, в том числе выделение основы и определение падежа и/или послелого (постпозиции). После того, как всем содержащимся в тексте формулам гражданского иска словоформам дана полная морфологическая квалификация (для имен – выделение основы, падежа и грамматического числа), решение задачи синтаксического анализа текста становится непростой, но достаточно доступной задачей.

Согласно принятому выше (см. 3.1) подходу к описанию полной парадигмы грузинских существительных, для описания модели синтаксической структуры текста также используется метасинтаксическая нотация Бэкуса-Наура (являющаяся видоизменением контекстно-свободных грамматик Хомского). Аналогично 3.1, в качестве формального аппарата для изложения синтаксиса синтагмы грузинского языка используется представление формальной модели естественного языка в виде контекстно-свободной грамматики, хотя альтернативные модели фактически не рассматривались. Изложение использованных в работе синтаксических правил в основном следует работе М.Векуа и др. [22]

3.2.1. Модель синтаксической структуры текста формулы гражданского иска на ограниченном естественном грузинском языке предполагает следующие ограничения.

3.2.1.1. Юридически корректная формула гражданского иска может являться только *синтагмой* грузинского языка – синтаксически связанный текст, не имеющий сказуемого, т.е. не содержащий *личных* глагольных форм.

3.2.1.2. Членами синтагмы могут быть только имена существительные и прилагательные грузинского языка, а также наречия. В отдельных случаях допускается использование местоимений (только третьего лица).

3.2.1.3. Сложно-сочиненные синтагмы должны иметь возможность разделения их на простые синтагмы, каждая из которых рассматривается как

отдельный текст (допускаются семантические связи между текстами при помощи местоимений).

3.2.1.4. Сложно-подчиненные синтагмы не допускаются.

3.2.1.5. В каждой синтагме допускается только один **актор**, описывающий семантически специфицируемое в синтагме действие. Им может являться только отглагольное имя (масдар). Актору синтаксически подчинены все остальные слова текста, но он сам не подчинен ни одному из них.

3.2.1.6. Основы всех входящих в синтагму словоформ принадлежат заранее определенному словарю и связаны общепринятыми для грузинского языка грамматическими правилами, регулирующими морфологические признаки именных словоформ и подчинение их взаимных связей установленным правилам синтаксиса грузинского языка.

3.2.1.7. Все входящие в синтагму словоформы подчиняются нормативным для грузинского языка грамматическим правилам, регулирующим их взаимные связи и подчинения на основании морфологических признаков.

3.2.1.8. Согласно нормам современного литературного грузинского языка, словоформа-определение должна предшествовать определяемой словоформе.

3.2.1. Модель синтаксической структуры текста формулы гражданского иска на ограниченном естественном грузинском языке предполагает, что все входящие в текст словоформы удовлетворяют приведенной в 3.1 модели морфологической структуры частей речи грузинского языка и приведенным в 3.2.1 ограничениям. В результате морфологического разбора входящих в текст словоформ, выполненного согласно принятой модели, каждая из них преобразуется в специальную структуру, которая для масдаров, существительных и прилагательных имеет следующую спецификацию (примеры структур см. ниже в гл.4).

3.2.1.1. Исходная словоформа, совпадающая с приведенной в тексте.

3.2.1.2. Выделенная из словоформы основа (с учетом ее морфонологического изменения в некоторых падежах).

3.2.1.3. Морфологическая квалификация словоформы: падеж, грамматическое число, послелог (постпозиция), обстоятельная форма прилагательного (в линейном падеже). Отметим, что в рассмотренном в данной работе экспериментальном материале (см. ниже гл.4) встречаются только слитные послелоги «внутри» (*ო*), «из» (*ღაბ*), «вместе» (*თაბ*), а также отдельно

записываемые послелогои – в гл.4 представлены только *ḡāḡā* «из-за», *ḡāḡāḡā* «без, вне» и *ḡāḡāḡāḡā* «об, насчет».

Придание всем словоформам текста приведенной выше структуры, с точки зрения продукционной модели, равнозначно связыванию любой из них с одним из следующих нетерминалов, однозначно описывающих морфологическую квалификацию словоформы. При переходе к модели синтаксиса, каждая словоформа лишается своего текстуального представления (т.е. теряет статус терминала) и приобретает в продукционных правилах статус нетерминала. Ниже следует определение всех нетерминалов продукционных правил модели синтаксической структуры текста формулы гражданского иска, предполагаемых как имеющие свое значение после выполнения морфологического анализа текста.

MasdFormNom – словоформа есть именительный падеж масдара;

MasdFormGen – словоформа есть родительный падеж масдара;

NounFormNom – словоформа есть именительный падеж существительного;

NounFormGen – словоформа есть родительный падеж существительного;

AjectFormNom – словоформа есть именительный падеж прилагательного;

AjectFormGen – словоформа есть родительный падеж прилагательного;

NounFormPost_In – существительное в постпозиционной форме «внутри»;

NounFormPost_From – существительное в постпозиционной форме «из»;

NounFormPost_With – существительное в постпозиционной форме «вместе»;

AjectFormAdverb – обстоятельная форма прилагательного;

NounForm_Instr – существительное в инструментальном падеже (или наречие);

NounForm_Advr – наречие или существительное в адвербиативном падеже (или наречие);

Adverb – наречие, не производимое от существительного;

Postpos_Because – отдельно записываемый послелог (постпозиция) «из-за»;

Postpos_Without – отдельно записываемый послелог (постпозиция) «без, вне»;

Postpos_About – отдельно записываемый послелог (постпозиция) «об, насчет»;

Conjunct – союз (используется только «И»);

Pronoun – местоимение (3-го лица);

3.2.2. Алгебраическое описание синтаксических связей между входящими в синтагму словоформами (представление текста формулы гражданского иска на ограниченном естественном грузинском языке) состоит из следующих основных разделов:

– продукционные правила формального описания образующих элементов семантико-синтаксической структуры синтагмы – акторов;

– продукционные правила формального описания агента – прямого дополнения синтагмы, на которое непосредственно распространяется действие актора.

– продукционные правила формального описания косвенных дополнений в синтагме, уточняющие семантику прямого дополнения, на которые косвенно распространяется действие актора.

– продукционные правила формального описания сочетания актора с подчиненными ему косвенными дополнениями;

– продукционные правила формального описания атрибутов в синтагме, описывающие атрибутивные свойства агента, на которые также косвенно распространяется действие актора.

– продукционные правила формального описания сочетания актора и агента с подчиненными ему атрибутами;

– продукционные правила формального описания адвербов или адвербиативных конструкций (обстоятельств или их списков) в синтагме, описывающие условия или обстоятельства применения действия, представленного актором к агенту, или на косвенное распространение действие актора;

– продукционные правила формального описания сочетания актора с подчиненными ему адвербами, а также образования адвербиативных конструкций;

– продукционные правила формального описания атрибутов, подчиненных адвербиативным конструкциям (обстоятельствам или их спискам) в синтагме, описывающие атрибутивные свойства, относящиеся непосредственно к адвербу;

- продукционные правила формального описания сочетания адвербиативных конструкций с подчиненными им атрибутами;
- продукционные правила формального описания полного текста синтагмы в общей форме.

3.2.3. Продукционные правила формального описания образующих элементов семантико-синтаксической структуры синтагмы – *акторов*;

Согласно принятым выше ограничениям в текстах юридических формул гражданского иска, задача поиска ведущего элемента синтагмы для построения ее синтаксической структуры – актора может быть однозначно решена путем выделения отглагольных имен (масдаров), которые формально определены нетерминалами после выполнения модуля морфологического анализа. При этом учитывается, что масдары могут находиться либо в именительном, либо в родительном падеже, причем, во втором случае употребляется отдельно записываемый послелог შესახებ «об». Кроме того, актор должен являться последней словоформой синтагмы.

Для записи соответствующих продукционных правил вводится нетерминал <Actor>, представляющий актор:

<Actor>::=<MasdFormNom>|

<MasdFormGen><Postpos_About>

В качестве дополнительного контроля синтаксической корректности синтагмы, представляющей текст юридической формулы гражданского иска, можно использовать ограничение, что в каждой синтагме допускается один и только один актор. Нарушение этого ограничения означает невозможность автоматического выполнения синтаксического анализа текста синтагмы, без использования дополнительных правил экстрасинтаксического характера, например, разделение текста юридической формулы гражданского иска на две и более синтагмы, каждая из которых содержит только один актор.

Примеры решения задачи:

ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება

ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლების შესახებ

3.2.4. Продукционные правила формального описания агента – прямого дополнения синтагмы, на которое непосредственно распространяется действие актора.

Согласно принятым выше ограничениям в текстах юридических формул гражданского иска, задача поиска агента, на который непосредственно распространяется действие актора может быть однозначно решена путем выделения существительных, находящихся в родительном падеже и расположенных слева от актора, которые формально определены нетерминалами после выполнения модуля морфологического анализа.

Хотя в простейшем случае задача выглядит тривиальной

აღიშენების დაცისრება,

в более общем случае возможно появление ряда трудностей, существенно усложняющих решение задачи. Агент может вообще отсутствовать, например:

მემკვიდრედ ცნობა

В более сложных случаях возможно проявление неоднозначности указанного грамматического правила, например, когда в тексте может присутствовать более одного существительного, находящегося в родительном падеже:

ბიბის ნასციდობის ხელშეკრულების ბათილად ცნობა

В этом случае может быть использовано эвристическое правило, основывающееся на распространенной (но не обязательной) грамматической традиции грузинского языка, согласно которой прямое дополнение обычно располагается на последней позиции среди всех имен существительных в родительном падеже; в данном случае оно может быть успешно использовано:

ბიბის ნასციდობის ხელშეკრულების ბათილად ცნობა

В более сложных случаях необходимо решение задачи методами, лежащими вне продукционных правил, например, путем интерактивного маркирования текста (см. гл.4).

Для записи соответствующих продукционных правил вводится нетерминал <Agent>, представляющий словоформу текста, являющуюся агентом. Это правило должно учитывать следующие альтернативы:

– агент непосредственно предшествует актору;

– между агентом и актором расположены словоформы, не являющиеся существительными в родительном падеже; описание подобной структуры правой части продукционного правила в рамках контекстно-свободных

грамматик затруднительно, хотя в данном случае оно доступно путем искусственного введения специального нетерминала <Non_Genitive_Forms_List>, описывающего любую последовательность словоформ, не являющихся существительными в родительном падеже; продукционные правила для определения этого нетерминала для упрощения изложения опускаются, поскольку они должны содержать все определенные выше нетерминалы, за исключением <NounFormGen >:

<Agent>::=< NounFormGen ><Actor>|

<NounFormGen ><Non_Genitive_Forms_List><Actor>

3.2.5. Продукционные правила формального описания косвенных дополнений в синтагме, уточняющих семантику прямого дополнения (актора), а также формального описания сочетания актора с подчиненными ему косвенными дополнениями

Согласно принятым выше ограничениям в текстах юридических формул гражданского иска, задача поиска косвенных дополнений в синтагме, на которые косвенно распространяется действие актора, может быть однозначно решена в случае, когда в синтагме имеется последовательность существительных, находящихся в родительном падеже и расположенных непосредственно слева от актора, которые формально определены нетерминалами после выполнения модуля морфологического анализа. В этом случае косвенные дополнения прямо подчинены прямому дополнению и играют семантическую роль субстантивного уточнения свойств объекта, на которое направлено действие актора – прямого дополнения. Типичный пример:

ბობის ნასყიდობის ხელშეკრულების ბათილად ცნობა

Как показывает следующий пример:

ავტორის თანხმობის გარეშე ნაწარმების აკრძალვა

ბავშვის საცხოვრებელი ადგილის განსაზღვრა,

в синтагме между косвенными дополнениями, а также между ними и актором возможно помещение других нетерминалов, не являющихся существительными в родительном падеже, например, прилагательных или отделяемых постпозиций; учет этой синтаксической особенности синтагмы

возможен также путем введения определенного выше специального нетерминала <Non_Genitive_Forms_List>.

Для определения продукционных правил необходимо введение нетерминала <Indirect_Object_Construct>, представляющего синтаксическую конструкцию, содержащую список косвенные дополнений в сочетании с агентом; далее необходимо записать следующее продукционное правило:

$$\begin{aligned} \langle \text{Indirect_Object_Construct} \rangle ::= & \langle \text{NounFormGen} \rangle \langle \text{Agent} \rangle | \\ & \langle \text{NounFormGen} \rangle \langle \text{Non_Genitive_Forms_List} \rangle \langle \text{Agent} \rangle | \\ & \langle \text{NounFormGen} \rangle \langle \text{Indirect_Object_Construct} \rangle | \\ & \langle \text{NounFormGen} \rangle \langle \text{Non_Genitive_Forms_List} \rangle \langle \text{Indirect_Object_Construct} \rangle \end{aligned}$$

3.2.6. Продукционные правила формального описания атрибутов в синтагме, показывающих атрибутивные свойства актора или агента, а также продукционные правила формального описания сочетания актора и агента с подчиненными ему атрибутами.

Согласно принятым выше ограничениям в текстах юридических формул гражданского иска, задача поиска атрибутов в синтагме, уточняющих семантику прямого дополнения (агента), на которое также косвенно распространяется действие актора, может быть однозначно решена в случае, когда в синтагме имеются прилагательные в различных формах склонения, которые непосредственно предшествуют агенту. При этом прилагательные могут быть, как правило, в именительном падеже, например:

შესაბამისი კომპენსაციის გადახდის შესახებ

Число атрибутов в этом случае не ограничено, они могут быть связаны как запятыми, так и союзом «И»:

მიყენებული მორალური და მატერიალური ზიანის ანაზრაურების შესახებ

Кроме того, атрибутивные свойства могут быть определены не только для агента, но и непосредственно для актора; в этом случае возможно использование обстоятельной формы прилагательного:

ანდერძის ნაწილობრივი ბათილობა

При этом в силе остается правило, вытекающее из принятых выше ограничений: в синтагме атрибуты должны находиться слева (чаще всего – непосредственно слева) от словоформы, свойства которой они уточняют.

Для определения продукционных правил необходимо введение нетерминалов <Attrib_Agent_List> и <Attrib_Actor_List>, представляющих список атрибутов в сочетании с относящимся к ним агенту или актору, для которых записываются следующие продукционные правила:

<Attribute>:=<AjectFormNom>

<Attrib_Agent_List>::=<Attribute><Agent>|

<Attribute><Attrib_Agent_List>

<Attrib_Actor_List>::=<Attribute><Actor>|

<Attribute><Attrib_Actor_List>

3.2.7. Продукционные правила формального описания адverbов или адverbиативных конструкций (обстоятельств или их списков) и сочетания актора с подчиненными ему адverbами, а также образования адverbиативных конструкций;

Согласно принятым выше ограничениям в текстах юридических формул гражданского иска, задача поиска адverbов - в синтагме, уточняющей семантику прямого дополнения (агента) или самого актора, в частности, определение свойств действия, определяемого актором. В текстах, характерных для рассматриваемой предметной области, адverbами могут являться следующие части речи:

– существительные или местоимения, имеющие слитные послелоги, например:

ადმინისტრაციულ აქტში შეცდომის გასწორება,

ბინიდან გამოსახლება;

შვილის მშობლებთან დაბრუნება;

შეცვლისათვის გვარის მიკუთვნება;

– существительные, имеющие отделенные послелоги, например:

გარემოებათა გამო საქმის განახლება;

– существительные в обстоятельном падеже, неотличимые от наречий, например:

აქტის ბათილად ცნობა.

Число подряд расположенных атрибутов, вообще говоря, не ограничено, они могут быть связаны как запятыми, так и союзом «И»:

მემკვიდრედ და მესაკუთრედ ცნობის შესახებ.

При этом в силе остается правило, вытекающее из принятых выше ограничений: в синтагме адвербы должны находиться слева (чаще всего – непосредственно слева) от словоформы, обстоятельства действия которой они уточняют.

Для определения продукционных правил необходимо введение нетерминалов $\langle \text{AdverbForm_Agent_List} \rangle$ и $\langle \text{AdverbForm_Actor_List} \rangle$, представляющих список адвербов в сочетании с относящимися к ним агента или актора, для которых записываются следующие продукционные правила:

$$\langle \text{AdverbForm} \rangle ::= \langle \text{NounFormPost_In} \rangle | \langle \text{NounFormPost_From} \rangle | \langle \text{NounFormPost_With} \rangle |$$
$$\langle \text{AjectFormAdverb} \rangle | \langle \text{NounForm_Instr} \rangle | \langle \text{NounForm_Advr} \rangle |$$
$$\langle \text{Adverb} \rangle$$
$$\langle \text{AdverbForm_Agent_List} \rangle ::= \langle \text{AdverbForm} \rangle \langle \text{Agent} \rangle |$$
$$\langle \text{AdverbForm} \rangle \langle \text{AdverbForm_Agent_List} \rangle$$
$$\langle \text{AdverbForm_Actor_List} \rangle ::= \langle \text{AdverbForm} \rangle \langle \text{Actor} \rangle |$$
$$\langle \text{AdverbForm} \rangle \langle \text{AdverbForm_Actor_List} \rangle$$

3.2.8. Продукционные правила формального описания атрибутов, подчиненных адвербиативным конструкциям (обстоятельствам или их спискам) в синтагме, описывающих атрибутивные свойства, относящиеся непосредственно к адвербу. [2]

Согласно принятым выше ограничениям в текстах юридических формул гражданского иска, задача поиска атрибутов в синтагме, уточняющих семантику прямого дополнения (агента), может быть однозначно решена в случае, когда в

синтагме имеются прилагательные в различных формах склонения, которые непосредственно предшествуют адвербу. В отличие от рассмотренного в п.3.2.6 общего случая атрибутов, атрибуты внутри адвербиативных конструкций синтаксически подчинены только подчиняющему их адвербу, поскольку являются синтаксическим объектом более низкого уровня, чем атрибуты, подчиненные непосредственно прямому дополнению или актору. Иерархически они имеют семантическую роль предикативного уточнения условия, связанного с обстоятельствами применения действия, представленного адвербом. Число подряд расположенных подчиненных атрибутов также не ограничено.

Подчиненные атрибуты могут быть, прежде всего, в именительном падеже, например:

გამოყენებული ნაწარმოებისათვის წიგნის აკრძალვა

Возможна обстоятельная форма прилагательного, применяемая для образования дополнения в т.н. линейном падеже:

ადმინისტრაციულ აქტში შეცდომის გახწორება

Кроме того, возможно использование существительных в обстоятельном падеже, неотличимых от наречий:

ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება.

При этом в силе опять остается правило, вытекающее из принятых выше ограничений: в синтагме подчиненные атрибуты должны находиться слева (чаще всего – непосредственно слева) от словоформы, свойства которой они уточняют.

Для определения продукционных правил необходимо введение нетерминалов <ExactAttr_Agent_List> и <ExactAttr_Actor_List>, представляющих списки подчиненных атрибутов в сочетании с подчиняющими им агентом или актором, для которых записываются следующие продукционные правила:

< ExactAttribute>:=<AjectFormNom>|<AjectFormAdverb>|<NounForm_Advr>

< ExactAttr_Agent_List>::=< ExactAttribute><Agent>|

< ExactAttribute><ExactAttr_Agent_List>

<ExactAttr_Actor_List>::=< ExactAttribute><Actor>|

< ExactAttribute><ExactAttr_Actor_List>

3.2.9. Продукционные правила формального описания полного текста синтагмы в общей форме.

Для самого общего представления всей синтагмы, являющейся текстовым образом юридически корректной формулы гражданского иска на ограниченном естественном грузинском языке вводятся объединенные продукционные правила, включающие в себя все описанные выше нетерминалы. Для этого необходимо ввести завершающий нетерминал <Syntagma> (в терминал Пролога – цель грамматического разбора синтагмы), для которого записывается общее продукционное правило, содержащее в качестве альтернатив все приведенные выше продукционные правила для определения нетерминалов, представляющих самостоятельно существующие члены парадигмы. Приведенные выше продукционные правила для упрощения записи сгруппированы в продукционные группы, отражающие особенности их структуры. На завершающем этапе построения системы productions нетерминал <Syntagma> с игнорированием некоторых второстепенных частных случаев приводит к последовательности двух основных групп – конкатенации двух нетерминалов:

– <Agent_Group>, отображающих все возможные комбинации словоформ, имеющих синтаксическую интерпретацию как агента или его возможных расширений;

– <Actor_Group>, отображающих все возможные комбинации словоформ, имеющих синтаксическую интерпретацию как актора или его возможных расширений;

<Agent_Group>::=<Agent>|

<Indirect_Object_Construct>|

<Attrib_Agent_List>|

<AdverbForm_Agent_List>|

< ExactAttr_Agent_List>::=< ExactAttribute><Agent>

<Actor_Group>::=<Actor>|

<AdverbForm_Actor_List>|

<Attrib_Actor_List>|

<ExactAttr_Actor_List>|

<Syntagma>::=<Agent_Group><Actor_Group>

С помощью нетерминала <Syntagma>, совместно со всеми содержащими его определения продукционными правилами, возможно решение задач полного синтеза и анализа любой заданной синтагмы на предмет подтверждения того, что она является корректной синтаксической конструкцией ограниченного естественного грузинского языка, или не принадлежит к такому тексту.

Глава 4. Экспериментальное исследование системы описания семантики юридических формул гражданского иска

4.1. База экспериментального исследования семантики юридических формул гражданского иска.

Для проведения экспериментального исследования формального описания семантики юридических формул гражданского иска с помощью описанных в гл.2-3 моделей представления семантики, из имеющейся базы данных текстов, содержащихся в решениях судебных инстанций Грузии по гражданским делам, с помощью стохастической процедуры были выбраны тексты 16 решений, отражающие различные аспекты юридического содержания исков, типичных для практики гражданского права Грузии (согласно требованиям юридической этики, остальная информация об участниках соответствующих дел не приводится). База является совокупностью записей в распространенном формате Word for Windows, отделенных друг от друга стандартными разделителями строк; специальная система поиска записей не используется.

- (1) ადმინისტრაციულ აქტში შეცდომის გასწორებისა და მისგან გამოწვეული სამართლებრივი შედეგების შეცვლის შესახებ
- (2) ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების აკრძალვისა და შესაბამისი კომპენსაციის გადახდის შესახებ
- (3) ალიმენტის დაკისრების და შვილთან ურთიერთობის დღეების დადგენის შესახებ
- (4) ალიმენტის დანიშვნის და საცხოვრებელი ფართით უზრუნველყოფის შესახებ
- (5) ანდერძის ბათილად ცნობის, მემკვიდრედ და მესაკუთრედ ცნობის შესახებ
- (6) ანდერძის ნაწილობრივ ბათილობის და მემკვიდრედ ცნობა
- (7) არასაცხოვრებელი ფართის არაკანონიერად მითვისების შესახებ
- (8) არასრულწლოვანი შვილის მშობლებთან დაბრუნების, შეცვლისათვის დედის გვარის მიკუთვნების შესახებ
- (9) აქტის ბათილად ცნობა და მიყენებული მორალური და მატერიალური ზიანის ანაზრაურების შესახებ

- (10) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება
- (11) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლების შესახებ
- (12) ბავშვის საცხოვრებელი ადგილის განსაზღვრა, ალიმენტის დაკისრება
- (13) ბინიდან გამოსახლება და მატერიალური ზარალის ანაზღაურება
- (14) ბინის ნასყიდობის ხელშეკრულების ბათილად ცნობა
- (15) ბინის ყიდვა-გაყიდვის ხელშეკრულების ბათილად ცნობა
- (16) გადაწყვეტილების ბათილად ცნობა, ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების განახლება

4.1.2 Схема информационных потоков системы.

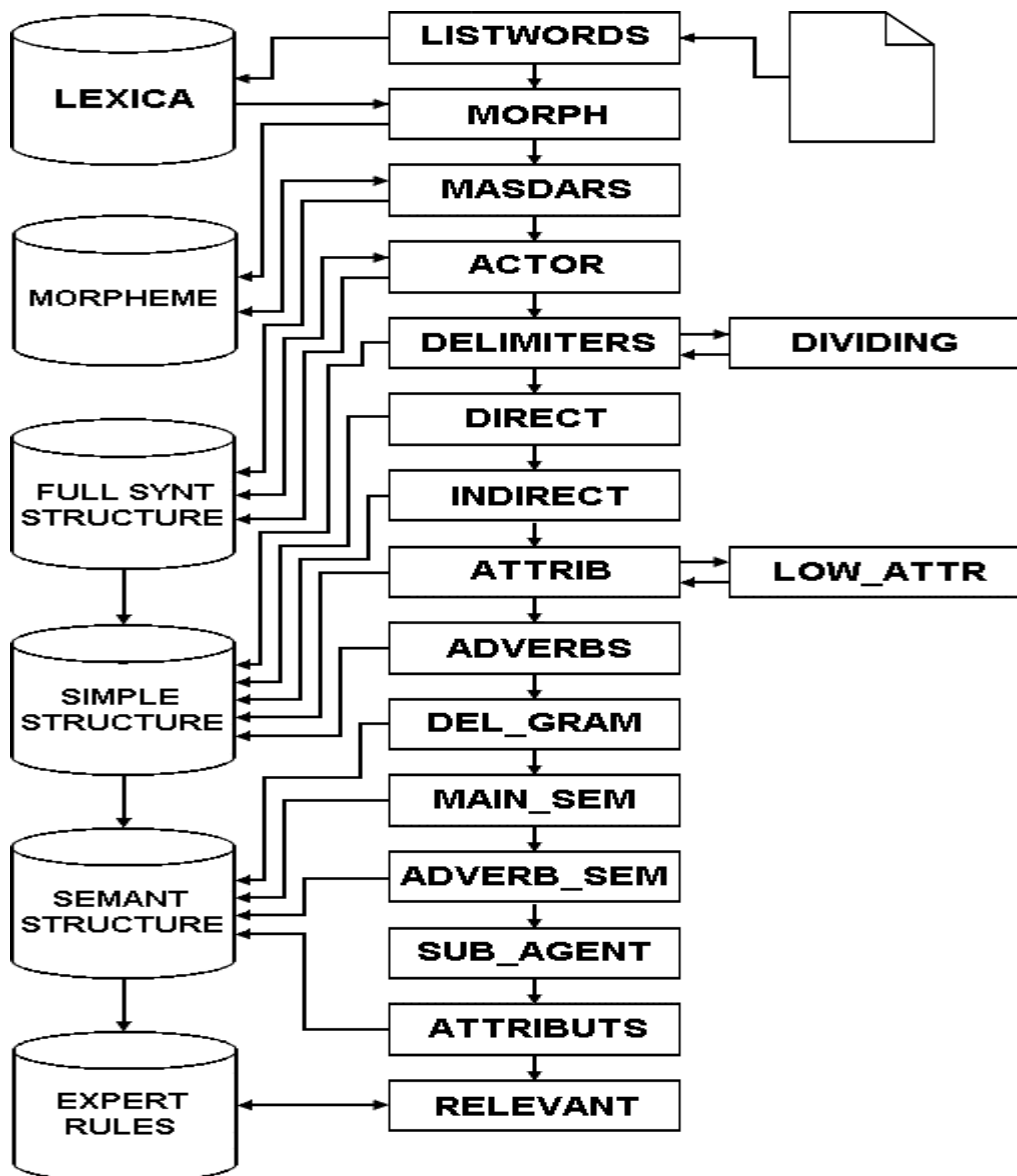


Рис 4.1.2.1

4.2. **Определение лексического состава текстов юридических формул гражданского иска**

Построение формального описания семантики является сложным многошаговым процессом обработки текстов на естественном языке (для наших приложений – на грузинском), исходной частью которого является автоматическое составление лексического состава текстов – списка всех словоформ грузинского языка, которые были использованы в базе экспериментального исследования.

Эта задача решается модулем LISTWORDS. Алгоритм поиска и выделения словоформ основан на сканировании текста базы путем использования стандартного алфавитного списка грузинских литер и списка возможных разделителей слов. Каждая уникальная словоформа входит в список только один раз, список упорядочивается в алфавитном порядке. Для приведенной в 4.1. базе экспериментального исследования результатом является следующий список словоформ:

ადგილის	განახლების	ნასყიდობის
ადმინისტრაციულ	განსაზღვრა	ნაწარმოების
ავტორის	გარემოებათა	ნაწილობრივ
აკრძალვისა	გარეშე	სამართლებრივი
ალიმენტის	გასწორებისა	საქმის
ანაზრაურების	გვარის	საცხოვრებელი
ანაზღაურება	და	უზრუნველყოფის
ანდერძის	დაბრუნების	ურთიერთობის
არაკანონიერად	დადგენის	ფართით
არასაცხოვრებელი	დაკისრება	ფართის
არასრულწლოვანი	დაკისრების	ყიდვა-გაყიდვის
აქტის	დანიშვნის	შედგების
აქტში	დედის	შესაბამისი
აღმოჩენილ	დღეების	შესახებ
ახლად	ზარალის	შეცდომის
ბავშვის	ზიანის	შეცვლის
ბათილად	თანხმობის	შეცვლისათვის
ბათილობის	კომპენსაციის	შვილთან
ბინიდან	მატერიალური	შვილის
ბინის	მემკვიდრედ	ცნობა
გადაწყვეტილების	მესაკუთრედ	ცნობის
გადახდის	მითვისების	წარმოების
გამო	მიკუთვნების	ხელშეკრულების
გამოსახლება	მისგან	
გამოყენებული	მიყენებული	
გამოწვეული	მორალური	
განახლება	მშობლებთან	

4.3. Полный морфологический анализ текстов юридических формул гражданского иска.

Каждая словоформа, входящая в список форм, которые были использованы в базе экспериментального исследования, обрабатывается алгоритмом полного морфологического анализа, результатом которого является представление каждой словоформы в виде двух компонентов: восстановленной по словоформе и ее возможному окончанию основы и ее морфологического атрибута.

Эта задача решается модулем MORPH. Алгоритм анализа основан на известной модели морфологии грузинских имен и удетр, в разработке которой принимал участие автор. Результаты работы алгоритма, помимо основы, дана морфологическая оценка, которая представлена следующими кодами: исходная словоформа, восстановленная основа и соответствующий ей маркер, представлены трехэлементными списками:

NOM – именительный падеж существительного или прилагательного;

GEN – родительный падеж существительного или прилагательного;

IN – существительное в постпозиционной форме «внутри»;

FROM – существительное в постпозиционной форме «из»;

WITH – существительное в постпозиционной форме «вместе»;

GENPL – родительный падеж существительного во множественном числе (архаическом);

AJECT – обстоятельная форма прилагательного;

ADVERB – наречие или существительное в инструментальном или адвербиативном падеже;

POSTPOS – отдельно записываемый послелог (постпозиция);

UNION – союз (используется только «И»);

PRONOUN – местоимение;

(ადგილის ადგილი **GEN**)

(ადმინისტრაციულ ადმინისტრაციული **AJECT**)

(ავტორის ავტორი **GEN**)

(აკრძალვისა აკრძალვა **GEN**)

(ალიმენტის ალიმენტი **GEN**)

(ანაზრაურების ანაზღაურება **GEN**)

(ანაზღაურება ანაზღაურება *NOM*)
 (ანდერძის ანდერძი *GEN*)
 (არაკანონიერად არაკანონიერად *ADVERB*)
 (არასაცხოვრებელი არასაცხოვრებელი *NOM*)
 (არასრულწლოვანი არასრულწლოვანი *NOM*)
 (აქტის აქტი *GEN*)
 (აქტში აქტი *IN*)
 (აღმოჩენილ აღმოჩენილი *AJECT*)
 (ახლად ახლად *ADVERB*)
 (ბავშვის ბავშვი *GEN*)
 (ბათილად ბათილად *ADVERB*)
 (ბათილობის ბათილობა *GEN*)
 (ბინიდან ბინა *FROM*)
 (ბინის ბინა *GEN*)
 (გადაწყვეტილების გადაწყვეტილება *GEN*)
 (გადახდის გადახდა *GEN*)
 (გამო გამო *POSTPOS*)
 (გამოსახლება გამოსახლება *NOM*)
 (გამოყენებული გამოყენებული *NOM*)
 (გამოწვეული გამოწვეული *NOM*)
 (განახლება განახლება *NOM*)
 (განახლების განახლება *GEN*)
 (განსაზღვრა განსაზღვრა *NOM*)
 (გარემოებათა გარემოება *GENPL*)
 (გარეშე გარეშე *POSTPOS*)
 (გასწორებისა გასწორება *GEN*)
 (გვარის გვარი *GEN*)
 (და და *UNION*)
 (დაბრუნების დაბრუნება *GEN*)
 (დადგენის დადგენა *GEN*)
 (დაკისრება დაკისრება *NOM*)
 (დაკისრების დაკისრება *GEN*)
 (დანიშვნის დანიშვნა *GEN*)
 (დედის დედა *GEN*)
 (დღეების დღე *GENPL*)
 (ზარალის ზარალი *GEN*)
 (ზიანის ზიანი *GEN*)
 (თანხმობის თანხმობა *GEN*)
 (კომპენსაციის კომპენსაცია *GEN*)
 (მატერიალური მატერიალური *GEN*)

(მემკვიდრედ მემკვიდრე *ADVERB*)
 (მესაკუთრედ მესაკუთრე *ADVERB*)
 (მითვისების მითვისება *GEN*)
 (მიკუთვნების მიკუთვნება *GEN*)
 (მისგან მისი *PRONOUN*)
 (მიყენებული მიყენებული *NOM*)
 (მორალური მორალური *NOM*)
 (მშობლებთან მშობლები *WITH*)
 (ნასყიდობის ნასყიდობა *GEN*)
 (ნაწარმოების ნაწარმოები *GEN*)
 (ნაწილობრივ ნაწილობრივ *AJECT*)
 (სამართლებრივი სამართლებრივი *NOM*)
 (საქმის საქმე *GEN*)
 (საცხოვრებელი საცხოვრებელი *NOM*)
 (უზრუნველყოფის უზრუნველყოფის *GEN*)
 (ურთიერთობის ურთიერთობა *GEN*)
 (ფართით ფართი *ADVERB*)
 (ფართის ფართი *GEN*)
 (ყიდვა-გაყიდვის ყიდვა-გაყიდვა *GEN*)
 (შედევების შედეგი *GENPL*)
 (შესაბამისი შესაბამისი *NOM*)
 (შესახებ შესახებ *ADVERB*)
 (შეცდომის შეცდომა *GEN*)
 (შეცვლის შეცვლა *GEN*)
 (შეცვლისათვის შეცვლა *FOR*)
 (შვილთან შვილი *WITH*)
 (შვილის შვილი *GEN*)
 (ცნობა ცნობა *NOM*)
 (ცნობის ცნობა *GEN*)
 (წარმოების წარმოება *GEN*)
 (ხელშეკრულების ხელშეკრულება *GEN*)

Существенной трудностью при решении задачи морфологического анализа словоформ является наличие неоднозначности в их грамматической интерпретации. Для рассматриваемой базы экспериментального исследования это проявляется только при анализе словоформы

(და და *UNION*),

которая может иметь интерпретацию не только как соединительного союза «И», но и как существительного «Сестра»:

(*და და NOM*).

В данном случае разрешение неоднозначности может быть достигнуто как путем статистического анализа, так и диалоговыми средствами. Наиболее удобно использование первого пути, поскольку семантическое понятие «Сестра» очень редко содержится в документах гражданского судопроизводства, тогда как соединение в один текст нескольких понятий с помощью соединительного союза является чрезвычайно распространенной конструкцией, встречающейся почти в 30-40 процентах располагаемых авторами реальных документах. Тем не менее использование диалогового режима является обязательной мерой, подтверждающей правильность решения, принятого на основании статистических критериев.[103]

4.4. Уточненная морфологическая классификация основ, выделенных из текстов юридических формул гражданского иска.

Проведенный в п.4.3 морфологический анализ словоформ недостаточен для решения поставленной в исследовании задачи, поскольку он исходит из общей классификации частей речи грузинского языка. С другой стороны, для автоматизированного построения модели семантики текстов юридических формул необходимо использование более тонких морфологических моделей, более полно отражающих семантические особенности встречающихся в текстах формул существительных.

Одной из важных для поставленных в работе целей является отделение имен субстантивного характера от отглагольных имен (масдаров), определяющих наименование действия (как известно, в грузинском языке отсутствует грамматическое понятие инфинитива). В рассмотренной в гл.2-3 модели грамматико-семантической структуры формул гражданского иска было высказано предположение, что исходным семантико-синтаксическим элементом этих формул всегда является отглагольное имя. Эта гипотеза подтверждается на всех примерах, принятых в качестве базы экспериментального исследования, а

также на существенно более широком фактическом материале, исследованном автором.

С формальной точки зрения, отделение масдаров от остальных имен возможно путем оценки парадигм, которые должны включать в себя не только словоформы, полученные склонением основы, но и связанные с ней стандартными деривационными средствами личные глагольные формы. Эта задача облегчается тем, что в масдаре, как правило, присутствует тематический суффикс, используемый в первой серии личного глагольного склонения, например:

წარმოება > წარმოებს.

Аналогично 4.3, ожидаемой трудностью при решении задачи уточненной морфологической классификации основ является наличие неоднозначности в их грамматической интерпретации (например, основы могут описывать как масдары, так и обычные субстантивные имена). Для рассматриваемой базы экспериментального исследования это, к счастью, не имеет места.

Эта задача решается модулем MASDARS. Алгоритм анализа основан на принятой выше модели морфологии грузинских имен и удетр, в разработке которой принимал участие автор. Результатом работы алгоритма является уточненная морфологическая (семантико-синтаксическая) оценка основы, которая представлена следующими кодами: восстановленная основа и соответствующий ей маркер в виде двухэлементных списков:

NOUN – основы существительных или прилагательных субстантивного характера, не порождающие личных глагольных форм;

MASD – основы отглагольных существительных, порождающие личные глагольные формы;

ADVERB – наречия, которые могут рассматриваться как формы обстоятельного или инструментального падежа от существительных или прилагательных;

для целей настоящего исследования эта грамматическая разница имеет чисто этимологический характер и не является существенной;

Остальные части речи не участвуют в рассматриваемой классификации.

(ადგილი *NOUN*)
 (ადმინისტრაციული *NOUN*)
 (აგტორი *NOUN*)
 (ალიმენტი *NOUN*)
 (ანდერძი *NOUN*)
 (არასაცხოვრებელი *NOUN*)
 (არასრულწლოვანი *NOUN*)
 (აქტი *NOUN*)
 (აღმოჩენილი *NOUN*)
 (ბავშვი *NOUN*)
 (ბინა *NOUN*)
 (გამოყენებული *NOUN*)
 (გამოწვეული *NOUN*)
 (გარემოება *NOUN*)
 (გვარი *NOUN*)
 (დედა *NOUN*)
 (დღე *NOUN*)
 (ზარალი *NOUN*)
 (ზიანი *NOUN*)
 (კომპენსაცია *NOUN*)
 (მატერიალური *NOUN*)
 (მიყენებული *NOUN*)
 (მორალური *NOUN*)
 (მშობლები *NOUN*)
 (ნასყიდობა *NOUN*)
 (ნაწარმოები *NOUN*)
 (ნაწილობრივი *NOUN*)
 (სამართლებრივი *NOUN*)
 (საქმე *NOUN*)
 (საცხოვრებელი *NOUN*)
 (ურთიერთობა *NOUN*)
 (ფართი *NOUN*)
 (შედეგი *NOUN*)
 (შესაბამისი *NOUN*)
 (შეცდომა *NOUN*)
 (შვილი *NOUN*)
 (ხელშეკრულება *NOUN*)
 (აკრძალვა *MASD*)
 (ანაზღაურება *MASD*)
 (ბათილობა *MASD*)
 (გადაწყვეტილება *MASD*)
 (გადახდა *MASD*)
 (გამოსახლება *MASD*)
 (განახლება *MASD*)
 (განსაზღვრა *MASD*)
 (გასწორება *MASD*)
 (დაბრუნება *MASD*)
 (დადგენა *MASD*)
 (დაკისრება *MASD*)
 (დანიშვნა *MASD*)
 (თანხმობა *MASD*)
 (მითვისება *MASD*)
 (მიკუთვნება *MASD*)
 (უზრუნველყოფა *MASD*)
 (ყიდვა-გაყიდვა *MASD*)
 (შეცვლა *MASD*)
 (ცნობა *MASD*)
 (წარმოება *MASD*)
 (არაკანონიერად *ADVERB*)
 (ახლად *ADVERB*)
 (ბათილად *ADVERB*)
 (მემკვიდრედ *ADVERB*)
 (მესაკუთრედ *ADVERB*)
 (ფართით *ADVERB*)
 (შესახებ *ADVERB*)
 (გამო *POSTPOS*)
 (გარეშე *POSTPOS*)
 (მისი *PRONOUN*)
 (და *UNION*)

4.5. Выделение образующих элементов семантико-синтаксической структуры юридических формул гражданского иска – акторов.

Согласно принятой в гл.2-3 модели семантики юридических формул гражданского иска, опорным элементом для построения ее синтаксической структуры (что, в свою очередь, определяет семантическую структуру текста), являются выделенные в тексте отглагольные имена (масдары).

Эта задача решается модулем ACTORS. Алгоритм анализа основан на использовании построенного в 4.4 списка отглагольных имен (масдаров), каждый из которых предполагается возможным актором данного текста. Масдары могут находиться либо в именительном, либо в родительном падеже. Каждый масдар заменяется на список, состоящий из ключевого слова «**ACTOR**» и отнесенной к нему словоформы. В каждом тексте может быть выделен один или два масдара. Неоднозначность в грамматической интерпретации акторов не имеет места.

(1) ადმინისტრაციულ აქტში შეცდომის (**ACTOR** გასწორებისა) და მისგან გამოწვეული სამართლებრივი შედეგების (**ACTOR** შეცვლის შესახებ)

(2) ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების (**ACTOR** აკრძალვისა) და შესაბამისი კომპენსაციის (**ACTOR** გადახდის შესახებ)

(3) ალიმენტის (**ACTOR** დაკისრების) და შვილთან ურთიერთობის დღეების (**ACTOR** დადგენის შესახებ)

(4) ალიმენტის (**ACTOR** დანიშვნის) და საცხოვრებელი ფართით (**ACTOR** უზრუნველყოფის შესახებ)

(5) ანდერძის ბათილად (**ACTOR** ცნობის), მემკვიდრედ და მესაკუთრედ (**ACTOR** ცნობის შესახებ)

(6) ანდერძის ნაწილობრივ (**ACTOR** ბათილობის) და მემკვიდრედ (**ACTOR** ცნობა)

(7) არასაცხოვრებელი ფართის არაკანონიერად (**ACTOR** მითვისების შესახებ)

- (8) არასრულწლოვანი შვილის მშობლებთან (**ACTOR** დაბრუნების), შეცვლისათვის დედის გვარის (**ACTOR** მიკუთვნების შესახებ)
- (9) აქტის ბათილად (**ACTOR** ცნობა) და მიყენებული მორალური და მატერიალური ზიანის (**ACTOR** ანაზრაურების შესახებ)
- (10) ახლად აღმოჩენილ გარემოებათა გამო საქმის (**ACTOR** განახლება)
- (11) ახლად აღმოჩენილ გარემოებათა გამო საქმის (**ACTOR** განახლების შესახებ)
- (12) ბავშვის საცხოვრებელი ადგილის (**ACTOR** განსაზღვრა), აღიშენების (**ACTOR** დაკისრება)
- (13) ბინიდან (**ACTOR** გამოსახლება) და მატერიალური ზარალის (**ACTOR** ანაზღაურება)
- (14) ბინის ნასყიდობის ხელშეკრულების ბათილად (**ACTOR** ცნობა)
- (15) ბინის ყიდვა-გაყიდვის ხელშეკრულების ბათილად (**ACTOR** ცნობა)
- (16) გადაწყვეტილების ბათილად (**ACTOR** ცნობა), ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების (**ACTOR** განახლება)

4.6. Выделение в тексте юридических формул гражданского иска возможных маркеров разделения на семантически условно-независимые подформулы.

В качестве маркеров разделения используются:

- соединительный союз «И»;
- разделительная запятая.

Эта задача решается модулем DELIMITERS. Алгоритм анализа основан на сопоставлении последовательности масдаров и разделителей, которые должны находиться между масдарами; для рассматриваемых текстов встречается не более одного разделителя, обуславливающего деление текста на два семантически независимые подтекста; при этом второй масдар должен находиться в именительном падеже, тогда как первый обычно может находиться в родительном, но иногда и именительном падеже (примеры 12, 13, 16).

Неоднозначность в грамматической интерпретации акторов проявляется в других способах использования соединительного союза «И», не для связывания отдельных семантически условно-независимых подформул, а для объединения адвербов или атрибутов в адвербиативные и атрибутивные конструкции (примеры 5 и 9). Подобные источники грамматической неоднозначности с большим трудом поддаются формализации, вследствие чего для выделения этих конструкций используются только интерактивные средства. [102]

Ниже приводятся только те тексты из базы экспериментального исследования, в которых обнаруживаются разделители:

- (1) (ადმინისტრაციულ აქტში შეცდომის (**ACTOR** გასწორებისა)) (**PART** და) (მისგან გამოწვეული სამართლებრივი შედეგების (**ACTOR** შეცვლის შესახებ))
- (2) (ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების (**ACTOR** აკრძალვისა)) (**PART** და) (შესაბამისი კომპენსაციის (**ACTOR** გადახდის შესახებ))
- (3) (ალიმენტის (**ACTOR** დაკისრების)) (**PART** და) (შვილთან ურთიერთობის დღეების (**ACTOR** დადგენის შესახებ))
- (4) (ალიმენტის (**ACTOR** დანიშვნის)) (**PART** და) (საცხოვრებელი ფართით (**ACTOR** უზრუნველყოფის შესახებ))
- (5) (ანდერძის ბათილად (**ACTOR** ცნობის)) (**PART** ,) (მემკვიდრედ და მესაკუთრედ (**ACTOR** ცნობის შესახებ))
- (6) ანდერძის ნაწილობრივ (**ACTOR** ბათილობის) (**PART** და) მემკვიდრედ (**ACTOR** ცნობა)
- (8) (არასრულწლოვანი შვილის მშობლებთან (**ACTOR** დაბრუნების)) (**PART** ,) (შეცვლისათვის დედის გვარის (**ACTOR** მიკუთვნების შესახებ))
- (9) (აქტის ბათილად (**ACTOR** ცნობა)) (**PART** და) (მიყენებული მორალური და მატერიალური ზიანის (**ACTOR** ანაზრაურების შესახებ))
- (12) (ბავშვის საცხოვრებელი ადგილის (**ACTOR** განსაზღვრა)) (**PART** ,) (ალიმენტის (**ACTOR** დაკისრება))
- (13) (ბინიდან (**ACTOR** გამოსახლება)) (**PART** და) (მატერიალური ზარალის (**ACTOR** ანაზღაურება))

(16) (გადაწყვეტილების ბათილად (**ACTOR** ცნობა)) (**PART** ,)
(ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების (**ACTOR**
განახლება))

4.7. Разделение текста юридических формул гражданского иска на две части с помощью выделенных маркеров

Эта задача решается модулем DIVIDING. Алгоритм анализа непосредственно использует выделенные на предыдущем шаге маркеры разделения текста на семантически условно-независимые подформулы. Ниже приводятся только те тексты из базы экспериментального исследования, в которых были выделены разделители.

(1.1) ადმინისტრაციულ აქტში შეცდომის (**ACTOR** გასწორებისა)

(1.2) მისგან გამოწვეული სამართლებრივი შედეგების (**ACTOR**
შეცვლის შესახებ)

(2.1) ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების
(**ACTOR** აკრძალვისა)

(2.2) შესაბამისი კომპენსაციის (**ACTOR** გადახდის შესახებ)

(3.1) აღიქმების (**ACTOR** დაკისრების)

(3.2) შვილთან ურთიერთობის დღეების (**ACTOR** დადგენის შესახებ)

(4.1) აღიქმების (**ACTOR** დანიშვნის)

(4.2) საცხოვრებელი ფართით (**ACTOR** უზრუნველყოფის შესახებ)

(5.1) ანდერძის ბათილად (**ACTOR** ცნობის)

(5.2) მემკვიდრედ და მესაკუთრედ (**ACTOR** ცნობის შესახებ)

(6.1) ანდერძის ნაწილობრივ (**ACTOR** ბათილობის)

(6.2) მემკვიდრედ (**ACTOR** ცნობა)

(8.1) არასრულწლოვანი შვილის მშობლებთან (**ACTOR** დაბრუნების)

(8.2) შეცვლისათვის დედის გვარის (**ACTOR** მიკუთვნების შესახებ)

(9.1) აქტის ბათილად (**ACTOR** ცნობა)

(9.2) მიყენებული მორალური და მატერიალური ზიანის (*ACTOR* ანაზრაურების შესახებ)

(12.1) ბავშვის საცხოვრებელი ადგილის (*ACTOR* განსაზღვრა)

(12.2) აღიშენების (*ACTOR* დაკისრება)

(13.1) ბინიდან (*ACTOR* გამოსახლება)

(13.2) მატერიალური ზარალის (*ACTOR* ანაზღაურება)

(16.1) გადაწყვეტილების ბათილად (*ACTOR* ცნობა)

(16.2) ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების (*ACTOR* განახლება)

4.8. Маркирование прямого дополнения в тексте юридических формул гражданского иска, на которое непосредственно распространяется действие актора.

Эта задача решается модулем DIRECT. Алгоритм маркирования заключается в поиске имени, находящегося в родительном падеже и расположенного слева от актора, подчиняющего прямое дополнение. В некоторых случаях (см. пример 13.1) прямое дополнение может отсутствовать. В некоторых случаях, напротив, в тексте может существовать более одного существительного, находящегося в родительном падеже (см. пример 14). Разрешение возникающей в этом случае синтаксической неоднозначности может быть достигнуто как путем использования эвристических правил, так и диалоговыми средствами. В данном случае эвристические правила основываются на распространенной (но не обязательной) грамматической традиции грузинского языка, согласно которой прямое дополнение обычно располагается на последней позиции среди всех объектов, являющихся именами существительными в родительном падеже. Несмотря на сравнительную надежность этого эвристического правила, тем не менее использование диалогового режима является обязательной мерой, подтверждающей правильность решения, принятого на основании эвристических критериев.

(1.1) ადმინისტრაციულ აქტში (*DIRECT* შეცდომის) (*ACTOR* გასწორებისა)

- (1.2) მისგან გამოწვეული სამართლებრივი (**DIRECT** შედეგების) (**ACTOR** შეცვლის შესახებ)
- (2.1) ავტორის თანხმობის გარეშე გამოყენებული (**DIRECT** ნაწარმოების) (**ACTOR** აკრძალვისა)
- (2.2) შესაბამისი (**DIRECT** კომპენსაციის) (**ACTOR** გადახდის შესახებ)
- (3.1) (**DIRECT** ალიმენტის) (**ACTOR** დაკისრების)
- (3.2) შვილთან ურთიერთობის (**DIRECT** დღეების) (**ACTOR** დადგენის შესახებ)
- (4.1) (**DIRECT** ალიმენტის) (**ACTOR** დანიშვნის)
- (4.2) საცხოვრებელი (**DIRECT** ფართით) (**ACTOR** უზრუნველყოფის შესახებ)
- (5.1) (**DIRECT** ანდერძის) ბათილად (**ACTOR** ცნობის)
- (5.2) მემკვიდრედ და მესაკუთრედ (**ACTOR** ცნობის შესახებ)
- (6.1) (**DIRECT** ანდერძის) ნაწილობრივ (**ACTOR** ბათილობის)
- (6.2) მემკვიდრედ (**ACTOR** ცნობა)
- (7) არასაცხოვრებელი (**DIRECT** ფართის) არაკანონიერად (**ACTOR** მითვისების შესახებ)
- (8.1) არასრულწლოვანი (**DIRECT** შვილის) მშობლებთან (**ACTOR** დაბრუნების)
- (8.2) შეცვლისათვის დედის (**DIRECT** გვარის) (**ACTOR** მიკუთვნების შესახებ)
- (9.1) (**DIRECT** აქტის) ბათილად (**ACTOR** ცნობა)
- (9.2) მიყენებული მორალური და მატერიალური (**DIRECT** ზიანის) (**ACTOR** ანაზრაურების შესახებ)
- (10) ახლად აღმოჩენილ გარემოებათა გამო (**DIRECT** საქმის) (**ACTOR** განახლება)
- (11) ახლად აღმოჩენილ გარემოებათა გამო (**DIRECT** საქმის) (**ACTOR** განახლების შესახებ)
- (12.1) ბავშვის საცხოვრებელი (**DIRECT** ადგილის) (**ACTOR** განსაზღვრა)
- (12.2) (**DIRECT** ალიმენტის) (**ACTOR** დაკისრება)
- (13.1) ბინიდან (**ACTOR** გამოსახლება)

(13.2) მატერიალური (*DIRECT* ზარალის) (*ACTOR* ანაზღაურება)

(14) ბინის ნასყიდობის (*DIRECT* ხელშეკრულების) ბათილად
(*ACTOR* ცნობა)

(15) ბინის ყიდვა-გაყიდვის (*DIRECT* ხელშეკრულების) ბათილად
(*ACTOR* ცნობა)

(16.1) (*DIRECT* გადაწვევები) ბათილად (*ACTOR* ცნობა)

(16.2) ახლად აღმოჩენილ გარემოებათა გამო საქმის (*DIRECT*
წარმოების) (*ACTOR* განახლება)

4.9. Маркирование косвенных дополнений в тексте юридических формул гражданского иска, уточняющих семантику прямого дополнения, на которое косвенно распространяется действие актора.

Косвенные дополнения синтаксически подчинены прямому дополнению и являются существительными в родительном падеже, играющими семантическую роль субстантивного уточнения свойств объекта, на которое направлено действие актора – прямого дополнения. Последовательность косвенных дополнений жестко фиксирована и представлена схемой (начиная от прямого дополнения): «расположенный справа объект уточняет принадлежность или свойства объекта, непосредственно расположенного слева». Вследствие этого каждое косвенное дополнение, если оно найдено, дополнительно маркируется цифрой после ключевого слова «*OBJECT*».

Эта задача решается модулем INDIRECT. Алгоритм маркирования заключается в поиске (в направлении слева направо) имен, находящихся в родительном падеже и расположенных слева от актора, подчиняющего прямое дополнение, но не совпадающее с прямым дополнением. Типичным примером являются конструкции 2.1 и 14 со следующей очевидной подчиненностью объектов (без учета других элементов текста):

(*OBJECT2* ავტორის) (*OBJECT3* თანხმობის გარეშე) (*DIRECT*
ნაწარმოების).

(*OBJECT2* ბინის) (*OBJECT3* ნასყიდობის) (*DIRECT*
ხელშეკრულების)

Синтаксической неоднозначности при решении указанной задачи не возникает.

Ниже приводятся только те примеры из базы экспериментального исследования семантики юридических формул гражданского иска, в которых выделяются косвенные дополнения.

(2.1) (*OBJECT2* ავტორის) (*OBJECT3* თანხმობის გარეშე)
გამოყენებული (*DIRECT* ნაწარმოების) (*ACTOR* აკრძალვისა)

(3.2) შვილთან (*OBJECT2* ურთიერთობის) (*DIRECT* დღეების)
(*ACTOR* დადგენის შესახებ)

(8.2) შეცვლისათვის (*OBJECT2* დედის) (*DIRECT* გვარის) (*ACTOR*
მიკუთვნების შესახებ)

(12.1) (*OBJECT2* ბავშვის) საცხოვრებელი (*DIRECT* ადგილის)
(*ACTOR* განსაზღვრა)

(14) (*OBJECT2* ბინის) (*OBJECT3* ნასყიდობის) (*DIRECT*
ხელშეკრულების) ბათილად (*ACTOR* ცნობა)

(15) (*OBJECT2* ბინის) (*OBJECT3* ყიდვა-გაყიდვის) (*DIRECT*
ხელშეკრულების) ბათილად (*ACTOR* ცნობა)

(16.2) ახლად აღმოჩენილ გარემოებათა გამო (*OBJECT2* საქმის)
(*DIRECT* წარმოების) (*ACTOR* განახლება)

4.10. Маркирование атрибутов в тексте юридических формул гражданского иска, описывающих атрибутивные свойства прямого дополнения, на которые также косвенно распространяется действие актора.

Атрибуты синтаксически подчинены прямому дополнению и являются прилагательными в именительном падеже, играющими семантическую роль предикативного уточнения свойств объекта, на которое направлено действие актора – прямого дополнения. Исключением является пример 8, где атрибут «ნაწილობრივ» может быть представлен как обстоятельная форма прилагательного «ნაწილობრივი» (возможно также представление этой формы как адверба, но оно менее удобно для целей исследования). [30]

Эта задача решается модулем ATTRIB. Алгоритм решения основан на том, что расположение атрибутов-прилагательных фиксировано – они должны находиться слева (чаще всего – непосредственно слева) от прямого дополнения.

Синтаксическая неоднозначность при решении указанной задачи играет важную роль, так как прилагательные часто описывают не атрибутивные свойства прямого дополнения, а атрибутивные свойства адверба (см. ниже, например, в примере 1.1 прилагательная словоформа «ადმიბისტრაციულ»).

Разрешение возникающей в этом случае синтаксической неоднозначности может быть достигнуто как путем использования эвристических правил, так и диалоговыми средствами. В данном случае эвристические правила основываются на распространенной (но не обязательной) грамматической традиции грузинского языка, согласно которой атрибут обычно располагается непосредственно перед прямым дополнением. Это эвристическое правило, аналогично эвристике п.4.8, также является сравнительно надежным, однако и в этом случае использование диалогового режима является обязательной мерой, подтверждающей правильность решения, принятого на основании эвристических критериев.

Ниже приводятся только те примеры из базы экспериментального исследования семантики юридических формул гражданского иска, в которых выделяются атрибуты, подчиненные косвенному дополнению.

(1.2) მისგან გამოწვეული (*ATTR* სამართლებრივი) (*DIRECT* შედეგების) (*ACTOR* შეცვლის შესახებ)

(2.2) (*ATTR* შესაბამისი) (*DIRECT* კომპენსაციის) (*ACTOR* გადახდის შესახებ)

(4.2) (*ATTR* საცხოვრებელი) (*DIRECT* ფართით) (*ACTOR* უზრუნველყოფის შესახებ)

(6.1) (*DIRECT* ანდერძის) (*ATTR* ნაწილობრივ) (*ACTOR* ბათილობის)

(7) (*ATTR* არასაცხოვრებელი) (*DIRECT* ფართის) არაკანონიერად (*ACTOR* მითვისების შესახებ)

(8.1) (*ATTR* არასრულწლოვანი) (*DIRECT* შვილის) მშობლებთან (*ACTOR* დაბრუნების)

(9.2) (*ATTR* მიყენებული) (*ATTR* მორალური და მატერიალური) (*DIRECT* ზიანის) (*ACTOR* ანაზრაურების შესახებ)

(12.1) (*OBJECT2* ბავშვის) (*ATTR* საცხოვრებელი) (*DIRECT* ადგილის) (*ACTOR* განსაზღვრა)

(13.2) (*ATTR* მატერიალური) (*DIRECT* ზარალის) (*ACTOR* ანაზღაურება)

4.11. Маркирование адвербов или адвербиативных конструкций (обстоятельств или их списков) в тексте юридических формул гражданского иска, описывающих условия или обстоятельства применения действия, представленного актором, к прямому дополнению, на которые также косвенно распространяется действие актора.

Адвербы синтаксически подчинены прямому дополнению и играют семантическую роль определения широкого спектра свойств объекта, на которое направлено действие актора – прямого дополнения, а также, в некоторых случаях, определения свойств действия, определяемого самим актором.

Как показал статистический анализ текстов юридических формул гражданского иска, содержащихся в базе экспериментального исследования, адвербами могут являться следующие части речи:

– существительные или местоимения, имеющие слитные послелоги (например: აქტში, მისგან, მშობლებთან, შეცვლისათვის, ბინიდან);

– существительные, имеющие отделенные послелоги (например: გარემოებათა გამო);

– существительные в обстоятельном падеже, неотличимые от наречий (например: ბათილად, მემკვიდრედ, არაკანონიერად).

В некоторых случаях адверб может быть расширен связанными с ним атрибутами (не путать с атрибутами, подчиненными непосредственно прямому дополнению или актору – см. ниже).

Эта задача решается модулем ADVERBS. Алгоритм решения основан на том, что расположение адвербов фиксировано – они должны находиться либо слева (чаще всего – непосредственно слева) от прямого дополнения, либо между прямым дополнением и актором, преимущественно в тех случаях, когда адverb допускает интерпретацию наречия (см., например, пример 9.1).

Синтаксическая неоднозначность при решении указанной задачи играет наиболее важную роль, так как позиция адверба и его грамматические свойства допускают широкую вариацию, существенно препятствующую формализации синтаксической оценки. Разрешение возникающей в этом случае синтаксической неоднозначности может быть достигнуто как путем использования эвристических правил, так и диалоговыми средствами, причем сложность решения задачи заключается в возможном появлении многовариантности эвристических решений. Например, если объединить тексты примеров 9.1 и 10 в следующий юридически совершенно корректный текст:

(*ADVERBIAL* ახლად აღმოჩენილ გარემოებათა გამო) (*DIRECT* აქტის) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა),

то наличие в нем двух адвербов, в равной степени подчиненных актору, существенно затрудняет полностью формализованное выделение адверба и приводит к появлению двух гипотез об их наличии (в данном случае корректными оказываются обе).

Таким образом, в этом случае использование диалогового режима является еще более необходимой мерой, подтверждающей правильность решения или правильность выбора из нескольких вариантов решения, предложенных на основании эвристических критериев.

Ниже приводятся только те примеры из базы экспериментального исследования семантики юридических формул гражданского иска, в которых выделяются адвербы или адвербиативные конструкции.

(1.1) (*ADVERBIAL* აღმინისტრაციულ აქტში) (*DIRECT* შეცდომის) (*ACTOR* გასწორებისა)

(1.2) (*ADVERBIAL* მისგან გამოწვეული) (*ATTR* სამართლებრივი) (*DIRECT* შედეგების) (*ACTOR* შეცვლის შესახებ)

- (5.1) (*DIRECT* ანდერძის) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობის)
- (5.2) (*ADVERBIAL* მემკვიდრედ და მესაკუთრედ) (*ACTOR* ცნობის შესახებ)
- (6.2) (*ADVERBIAL* მემკვიდრედ) (*ACTOR* ცნობა)
- (7) (*ATTR* არასაცხოვრებელი) (*DIRECT* ფართის) (*ADVERBIAL* არაკანონიერად) (*ACTOR* მითვისების შესახებ)
- (8.1) (*ATTR* არასრულწლოვანი) (*DIRECT* შვილის) (*ADVERBIAL* მშობლებთან) (*ACTOR* დაბრუნების)
- (8.2) (*ADVERBIAL* შეცვლისათვის) (*OBJECT2* დედის) (*DIRECT* გვარის) (*ACTOR* მიკუთვნების შესახებ)
- (9.1) (*DIRECT* აქტის) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა)
- (10) (*ADVERBIAL* ახლად აღმოჩენილ გარემოებათა გამო) (*DIRECT* საქმის) (*ACTOR* განახლება)
- (11) (*ADVERBIAL* ახლად აღმოჩენილ გარემოებათა გამო) (*DIRECT* საქმის) (*ACTOR* განახლების შესახებ)
- (13.1) (*ADVERBIAL* ბინიდან) (*ACTOR* გამოსახლება)
- (14) (*OBJECT2* ბინის) (*OBJECT3* ნასყიდობის) (*DIRECT* ხელშეკრულების) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა)
- (15) (*OBJECT2* ბინის) (*OBJECT3* ყიდვა-გაყიდვის) (*DIRECT* ხელშეკრულების) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა)
- (16.1) (*DIRECT* გადაწყვეტილების) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა)
- (16.2) (*ADVERBIAL* ახლად აღმოჩენილ გარემოებათა გამო) (*OBJECT2* საქმის) (*DIRECT* წარმოების) (*ACTOR* განახლება)

4.12. Маркирование атрибутов внутри адвербиативных конструкций (обстоятельств или их списков) в тексте юридических формул гражданского иска, описывающих атрибутивные свойства, относящиеся непосредственно к адвербу.

Атрибуты внутри адвербиативных конструкций синтаксически подчинены только выделенному адвербу и являются синтаксическим объектом более

низкого уровня, чем атрибуты, подчиненные непосредственно прямому дополнению или актору. Они играют иерархически подчиненную семантическую роль предикативного уточнения условия или обстоятельства применения действия, представленного актором к прямому дополнению (оба из которых находятся на иерархически более высоком уровне, чем уточняемое обстоятельство).

Как показал статистический анализ текстов юридических формул гражданского иска, содержащимися в базе экспериментального исследования атрибутами низкого уровня могут являться следующие части речи:

- обстоятельная форма прилагательного, применяемая для образования дополнения в т.н. линейном падеже (например, ადმინისტრაციულ);
- прилагательные в именительном падеже (например, გამოწვეული, გამოყენებული);
- существительные в обстоятельном падеже, неотличимые от наречий (например, ახლად).

Эта задача решается модулем LOW_ATTR. Алгоритм решения основан на том, что расположение атрибутов низкого уровня фиксировано: они должны находиться либо слева, либо справа от подчиняющего адверба (чаще всего – непосредственно).

Для маркирования атрибутов низшего уровня используется специальная нотация, представляющая собой вложенную скобочную конструкцию внутри подчиняющей ей адвербиативной конструкции, со специальным ключевым словом «**LOWATTR**» (отделяющих их от атрибутов высокого уровня, подчиненных прямому дополнению или актору).

Синтаксическая неоднозначность при решении указанной задачи также играет важную роль, так как позиция атрибутов низкого уровня и его грамматические свойства допускают широкую вариацию, препятствующую формализации синтаксической оценки. Разрешение возникающей в этом случае синтаксической неоднозначности может быть достигнуто как путем использования эвристических правил, так и диалоговыми средствами, причем, сложность решения задачи заключается в возможном появлении многовариантности эвристических решений.

В этом случае, аналогично предыдущим, использование диалогового режима является необходимой мерой, подтверждающей правильность решения

или правильность выбора из нескольких вариантов решения те, которые предложены на основании эвристических критериев.

Ниже приводятся только те примеры из базы экспериментального исследования семантики юридических формул гражданского иска, в которых выделяются атрибуты низшего уровня, находящиеся внутри адвербиативных конструкций.

(1.1) (*ADVERBIAL* (*LOWATTR* აღმინისტრაციულ) აქტში) (*DIRECT* შეცდომის) (*ACTOR* გასწორებისა)

(1.2) (*ADVERBIAL* მისგან (*LOWATTR* გამოწვეული)) (*ATTR* სამართლებრივი) (*DIRECT* შედეგების) (*ACTOR* შეცვლის შესახებ)

(2.1) (*OBJECT2* ავტორის) (*OBJECT3* თანხმობის გარეშე (*LOWATTR* გამოყენებული)) (*DIRECT* ნაწარმოების) (*ACTOR* აკრძალვისა)

(10) (*ADVERBIAL* (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილ) გარემოებათა გამო) (*DIRECT* საქმის) (*ACTOR* განახლება)

(11) (*ADVERBIAL* (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილ) გარემოებათა გამო) (*DIRECT* საქმის) (*ACTOR* განახლების შესახებ)

(16.2) (*ADVERBIAL* (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილ) გარემოებათა გამო) (*OBJECT2* საქმის) (*DIRECT* წარმოების) (*ACTOR* განახლება)

4.13. Построение грамматически-независимого представления текстов юридических формул гражданского иска.

Для решения поставленных в гл.2-3 задач, использующих формализованное описание семантики текстов юридических формул гражданского иска, необходимо такое преобразование морфологических и синтаксических особенностей грамматических конструкций, при которых в наибольшей возможной степени были бы нивелированы все грамматические свойства языка, на которых они были первоначально записаны, и тем самым были бы в наиболее обнаженной форме представлены особенности глубинных

структур языка, имеющие непосредственное отношение к семантике текста, а с другой стороны, были бы устранены те случайные для представления семантики черты, которые определяются эволюцией и традициями естественного языка.

Для рассматриваемой языковой среды, отражающей как особенности естественного грузинского языка, так и специфику сублингвистической среды, в которой существует предметная область, порождающая генерацию юридических формул гражданского иска и связанные с ней коммуникативные процессы, преобразования морфологических и синтаксических особенностей грамматических конструкций, согласно построенным в гл.2-3 моделям, сводятся к следующим подзадачам.

4.13.1. Приведение всех лексических единиц, встречающихся в тексте, к нормализованной форме, лишенной морфологических признаков, в частности для существительных, означает преобразование словоформы к именительному падежу единственного числа. Это правило не распространяется на существительные в обстоятельном падеже, которые с формальной точки зрения неотличимы от наречий, а также неприменимо к самим наречиям и союзам, которые сохраняют свою нормализованную форму, фактически использованную в тексте.

Что касается отделяемых послелогов, то они в нормализованной форме текста опускаются, поскольку синтаксическая модель полностью раскрывает их морфо-синтаксические функции, и их текстуальное представление является излишним для глубинных структур языка, непосредственно представляющих семантику текста.

4.13.2. Перестановка (транспозиция) всех встречающихся в тексте последовательностей развития причинных связей, в большинстве случаев не совпадающей с последовательностью слов в естественном языке, в последовательность, отражающую развитие семантического подчинения, которая является базой для представления модели семантики текста, практически не зависит от особенностей конкретного языка-носителя текста (в данном случае грузинского). [80]

Для рассматриваемой предметной области, с учетом принятых в гл.2-3 модельных ограничений, эта последовательность формализуется следующим образом (в порядке семантического подчинения сверху-вниз):

– актер (единственный в принятом представлении текста);

- прямое дополнение (единственный в принятом представлении текста);
- косвенное дополнение (не обязательный элемент текста);
- атрибуты, подчиняющиеся прямому дополнению (не обязательный элемент текста);
- адвербы или адвербиативные конструкции, подчиняющиеся прямому дополнению (не обязательный элемент текста);
- атрибуты внутри адвербиативных конструкций, подчиняющиеся непосредственно адвербу (не обязательный элемент текста).

Эта задача решается модулем DEL_GRAM. Алгоритм решения основан на анализе списковых структур, полученных на предыдущих шагах построения синтаксического описания текста, на основании правил 4.13.1-4.13.2. Синтаксическая неоднозначность при решении указанной задачи не возникает.

Ниже приводятся грамматически-независимое представление текстов юридических формул гражданского иска, выбранных в качестве базы экспериментального исследования. Для удобства, синтаксические элементы в преобразованном тексте помещены с новой строки, если они отнесены к уровню главного – актора.

(1.1) (*ADVERBIAL* (*LOWATTR* ადმინისტრაციულ) აქტში) (*DIRECT* შეცდომის) (*ACTOR* გასწორებისა)

(*ACTOR* გასწორება)

(*DIRECT* შეცდომა)

(*ADVERBIAL* აქტი (*LOWATTR* ადმინისტრაციული))

(1.2) (*ADVERBIAL* მისგან (*LOWATTR* გამოწვეული)) (*ATTR* სამართლებრივი) (*DIRECT* შედეგების) (*ACTOR* შეცვლის შესახებ)

(*ACTOR* შეცვლა)

(*DIRECT* შედეგი)

(*ATTR* სამართლებრივი)

(*ADVERBIAL* მისი (*LOWATTR* გამოწვეული))

(2.1) (*OBJECT2* ავტორის) (*OBJECT3* თანხმობის გარეშე (*LOWATTR* გამოყენებული)) (*DIRECT* ნაწარმოების) (*ACTOR* აკრძალვისა)

(**ACTOR** აკრძალვა)

(**DIRECT** ნაწარმოები)

(**OBJECT2** ავტორი)

(**OBJECT3** თანხმობა (**LOWATTR** გამოყენებული))

(2.2) (**ATTR** შესაბამისი) (**DIRECT** კომპენსაციის) (**ACTOR** გადახდის შესახებ)

(**ACTOR** გადახდა)

(**DIRECT** კომპენსაცია)

(**ATTR** შესაბამისი)

(3.1) (**DIRECT** ალიმენტის) (**ACTOR** დაკისრების)

(**ACTOR** დაკისრებაი)

(**DIRECT** ალიმენტი)

(3.2) შვილთან (**OBJECT2** ურთიერთობის) (**DIRECT** დღეების) (**ACTOR** დადგენის შესახებ)

(**ACTOR** დადგენა)

(**DIRECT** დღე)

(**OBJECT2** ურთიერთობა)

(**ADVERBIAL** შვილი)

(4.1) (**DIRECT** ალიმენტის) (**ACTOR** დანიშვნის)

(**ACTOR** დანიშვნა)

(**DIRECT** ალიმენტი)

(4.2) (**ATTR** საცხოვრებელი) (**DIRECT** ფართით) (**ACTOR** უზრუნველყოფის შესახებ)

(**ACTOR** უზრუნველყოფა)

(**DIRECT** ფართი)

(**ATTR** საცხოვრებელი)

(5.1) (**DIRECT** ანდერძის) (**ADVERBIAL** ბათილად) (**ACTOR** ცნობის)

(*ACTOR* ცნობა)
(*DIRECT* ანდერძი)
(*ADVERBIAL* ბათილად)

(5.2) (*ADVERBIAL* მემკვიდრედ და მესაკუთრედ) (*ACTOR* ცნობის შესახებ)

(*ACTOR* ცნობა)
(*ADVERBIAL* მემკვიდრედ)
(*ADVERBIAL* მესაკუთრედ)

(6.1) (*DIRECT* ანდერძის) (*ATTR* ნაწილობრივ) (*ACTOR* ბათილობის)
(*ACTOR* ბათილობა)
(*DIRECT* ანდერძი)
(*ATTR* ნაწილობრივი)

(6.2) (*ADVERBIAL* მემკვიდრედ) (*ACTOR* ცნობა)
(*ACTOR* ცნობა)
(*ADVERBIAL* მემკვიდრედ)

(7) (*ATTR* არასაცხოვრებელი) (*DIRECT* ფართის) (*ADVERBIAL* არაკანონიერად) (*ACTOR* მითვისების შესახებ)

(*ACTOR* მითვისება)
(*DIRECT* ფართი)
(*ATTR* არასაცხოვრებელი)
(*ADVERBIAL* არაკანონიერად)

(8.1) (*ATTR* არასრულწლოვანი) (*DIRECT* შვილის) (*ADVERBIAL* მშობლებთან) (*ACTOR* დაბრუნების)

(*ACTOR* დაბრუნებაი)
(*DIRECT* შვილი)
(*ATTR* არასრულწლოვანი)
(*ADVERBIAL* მშობელი)

(8.2) (*ADVERBIAL* შეცვლისათვის) (*OBJECT2* დედის) (*DIRECT* გვარის) (*ACTOR* მიკუთვნების შესახებ)

(*ACTOR* მიკუთვნება)

(*DIRECT* გვარი)

(*OBJECT2* დედა)

(*ADVERBIAL* შეცვლა)

(9.1) (*DIRECT* აქტის) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა)

(*ACTOR* ცნობა)

(*DIRECT* აქტი)

(*ADVERBIAL* ბათილად)

(9.2) (*ATTR* მიყენებული) (*ATTR* მორალური და მატერიალური) (*DIRECT* ზიანის) (*ACTOR* ანაზრაურების შესახებ)

(*ACTOR* ანაზღაურება)

(*DIRECT* ზიანი)

(*ATTR* მიყენებული)

(*ATTR* მორალური)

(*ATTR* მატერიალური)

(10) (*ADVERBIAL* (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილ) გარემოებათა გამო) (*DIRECT* საქმის) (*ACTOR* განახლება)

(*ACTOR* განახლება)

(*DIRECT* საქმე)

(*ADVERBIAL* გარემოება (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილი))

(11) (*ADVERBIAL* (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილ) გარემოებათა გამო) (*DIRECT* საქმის) (*ACTOR* განახლების შესახებ)

(*ACTOR* განახლება)

(*DIRECT* საქმე)

(*ADVERBIAL* გარემოება (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილი))

(12.1) (**OBJECT2** ბავშვის) (**ATTR** საცხოვრებელი) (**DIRECT** ადგილის) (**ACTOR** განსაზღვრა)

(**ACTOR** განსაზღვრა)

(**DIRECT** ადგილი)

(**OBJECT2** ბავშვი)

(**ATTR** საცხოვრებელი)

(12.2) (**DIRECT** ალიმენტის) (**ACTOR** დაკისრება)

(**ACTOR** დაკისრება)

(**DIRECT** ალიმენტი)

(13.1) (**ADVERBIAL** ბინიდან) (**ACTOR** გამოსახლება)

(**ACTOR** გამოსახლება)

(**ADVERBIAL** ბინა)

(13.2) (**ATTR** მატერიალური) (**DIRECT** ზარალის) (**ACTOR** ანაზღაურება)

(**ACTOR** ანაზღაურება)

(**DIRECT** ზარალი)

(**ATTR** მატერიალური)

(14) (**OBJECT2** ბინის) (**OBJECT3** ნასყიდობის) (**DIRECT** ხელშეკრულების) (**ADVERBIAL** ბათილად) (**ACTOR** ცნობა)

(**ACTOR** ცნობა)

(**DIRECT** ხელშეკრულება)

(**OBJECT2** ბინა)

(**OBJECT3** ნასყიდობა)

(**ADVERBIAL** ბათილად)

(15) (**OBJECT2** ბინის) (**OBJECT3** ყიდვა-გაყიდვის) (**DIRECT** ხელშეკრულების) (**ADVERBIAL** ბათილად) (**ACTOR** ცნობა)

(**ACTOR** ცნობა)

(*DIRECT* ხელშეკრულება)

(*OBJECT2* ბინა)

(*OBJECT3* ყიდვა-გაყიდვა)

(*ADVERBIAL* ბათილად)

(16.1) (*DIRECT* გადაწყვეტილების) (*ADVERBIAL* ბათილად) (*ACTOR* ცნობა)

(*ACTOR* ცნობა)

(*DIRECT* გადაწყვეტილება)

(*ADVERBIAL* ბათილად)

(16.2) (*ADVERBIAL* (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილ) გარემოებათა გამო) (*OBJECT2* საქმის) (*DIRECT* წარმოების) (*ACTOR* განახლება)

(*ACTOR* განახლება)

(*DIRECT* წარმოება)

(*OBJECT2* საქმე)

(*ADVERBIAL* გარემოება (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილი))

4.14. Маркирование ключевых семантических элементов формализованного представления содержания текстов юридических формул гражданского иска (актора и агента).

После преобразования текстов юридических формул гражданского иска, в котором нивелированы грамматические свойства языка, на котором они были первоначально записаны, с целью получения возможно более точного представления особенностей глубинных структур языка, имеющих непосредственное отношение к семантике изучаемых текстов, следует маркирование ключевых семантических элементов формализованного представления содержания текста – актора и агента.

Семантическое понятие актора является полным аналогом одноименного синтаксического понятия, но для того, чтобы оно было однозначно формализовано, что этот элемент принадлежит не внешней структуре текста, а к его внутренней, глубинной структуре, он обозначается специальным маркером «*action*». Это обозначение подчеркивает, что в данном случае имеет место не синтаксическое подчинение всех остальных слов текста (через иерархическую структуру), которое выявляется при вариации грамматических конструкций или при окружении текста другими текстами, а маркирование ведущего семантического элемента ситуации, которую описывает данный текст и которая может быть однозначно переведена на другие языки или с которой может быть связана другая вариация описания семантики текста.

Примеры синтаксической вариации:

ანდერძის ბათილად ცნობა >> ანდერძი გაბათილდა

ანდერძის ბათილად ცნობა >> მოხდა ანდერძის გაბათილება.

Примеры семантической вариации:

ანდერძის ბათილად ცნობა >> განცხადების ბათილად ცნობა.

ანდერძის ბათილად ცნობა >> признание завещания недействительным.

Семантическое понятие агента является аналогом синтаксического понятия прямого дополнения, но отличается от него принципиально другим пониманием управления между членами предложения. Если с синтаксической точки зрения прямое дополнение грамматически подчинено актору и должно находиться, в большинстве случаев, в родительном падеже (в грузинском языке), то связь между актором и агентом имеет отстраненный от синтаксиса семантический характер и при ее описании вообще не рассматриваются падежи, в которых стоят соответствующие словоформы. С точки зрения описания семантики, в данном случае имеет место главная причинная связь сцены, которую описывает текст синтагмы, причем эта связь направлена от актора (представляющего действие сцены) к семантически подчиненному ему агенту (на который направлено действие сцены). Среди всех других объектов, участвующих в сцене, агент является объектом первого уровня, на который непосредственно направлено действие актора, тогда как на все другие объекты действие передается через агента, вследствие чего все они (если представлены в сцене)

имеют подчиненное, второстепенное значение, разве что лишь уточняют роль агента как приемника действия.

Эта задача решается модулем MAIN_SEM. Алгоритм решения основан на анализе списковых структур, полученных на предыдущих шагах построения синтаксического описания текста. Синтаксическая неоднозначность при решении указанной задачи не возникает.

Ниже приводится начальная часть формализованного представления содержания текстов юридических формул гражданского иска, в котором маркируются только актер и агент, выявленные по ее синтаксической структуре. Для удобства анализа приводятся три представления текста: исходное на естественном языке, грамматически-независимое с маркированием синтаксических элементов и семантическое, с указанием только актора и агента.

(1) ადმინისტრაციულ აქტში შეცდომის გასწორებისა და მისგან გამომწვეული სამართლებრივი შედეგების შეცვლის შესახებ

(1.1) (**ACTOR** გასწორება)

(**DIRECT** შეცდომა)

(**ADVERBIAL** აქტი (**LOWATTR** ადმინისტრაციული))

(**action** გასწორება) (**agent** შეცდომა)

(1.2) (**ACTOR** შეცვლა)

(**DIRECT** შედეგი)

(**ATTR** სამართლებრივი)

(**ADVERBIAL** მისი (**LOWATTR** გამომწვეული))

(**action** შეცვლა) (**agent** შედეგი)

(2) ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების აკრძალვისა და შესაბამისი კომპენსაციის გადახდის შესახებ

(2.1) (**ACTOR** აკრძალვა)

(**DIRECT** ნაწარმოები)

(**OBJECT2** ავტორი)

(**OBJECT3** თანხმობა (**LOWATTR** გამოყენებული))

(**action** აკრძალვა) (**agent** ნაწარმოები)

(2.2) (**ACTOR** გადახდა)

(**DIRECT** კომპენსაცია)

(**ATTR** შესაბამისი)

(**action** გადახდა) (**agent** კომპენსაცია)

(3) ალიმენტის დაკისრების და შვილთან ურთიერთობის დღეების დადგენის შესახებ

(3.1) (**ACTOR** დაკისრება)

(**DIRECT** ალიმენტი)

(**action** დაკისრება) (**agent** ალიმენტი)

(3.2) (**ACTOR** დადგენა)

(**DIRECT** დღე)

(**OBJECT2** ურთიერთობა)

(**ADVERBIAL** შვილი)

(**action** დადგენა) (**agent** დღე ურთიერთობა)

(4) ალიმენტის დანიშვნის და საცხოვრებელი ფართით უზრუნველყოფის შესახებ

(4.1) (**ACTOR** დანიშვნა)

(**DIRECT** ალიმენტი)

(**action** დანიშვნა) (**agent** ალიმენტი)

(4.2) (**ACTOR** უზრუნველყოფა)

(**DIRECT** ფართი)

(**ATTR** საცხოვრებელი)

(**action** უზრუნველყოფა) (**agent** ფართი)

(5) ანდერძის ბათილად ცნობის, მემკვიდრედ და მესაკუთრედ ცნობის შესახებ

(5.1) (**ACTOR** ცნობა)

(**DIRECT** ანდერძი)

(**ADVERBIAL** ბათილად)

(**action** ცნობა) (**agent** ანდერძი)

(5.2) (**ACTOR** ცნობა)

(**ADVERBIAL** მემკვიდრედ)

(*ADVERBIAL* მესაკუთრედ)

(*action* ცნობა) (*agent* -)

(6) ანდერძის ნაწილობრივ ბათილობის და მემკვიდრედ ცნობა

(6.1) (*ACTOR* ბათილობა)

(*DIRECT* ანდერძი)

(*ATTR* ნაწილობრივი)

(*action* ბათილობა) (*agent* ანდერძი)

(6.2) (*ACTOR* ცნობა)

(*ADVERBIAL* მემკვიდრედ)

(*action* ცნობა) (*agent* -)

(7) არასაცხოვრებელი ფართის არაკანონიერად მითვისების შესახებ

(*ACTOR* მითვისება)

(*DIRECT* ფართი)

(*ATTR* არასაცხოვრებელი)

(*ADVERBIAL* არაკანონიერად)

(*action* მითვისება) (*agent* ფართი)

(8) არასრულწლოვანი შვილის მშობლებთან დაბრუნების, შეცვლისათვის დედის გვარის მიკუთვნების შესახებ

(8.1) (*ACTOR* დაბრუნებაი)

(*DIRECT* შვილი)

(*ATTR* არასრულწლოვანი)

(*ADVERBIAL* მშობელი)

(*action* დაბრუნებაი) (*agent* შვილი)

(8.2) (*ACTOR* მიკუთვნება)

(*DIRECT* გვარი)

(*OBJECT2* დედა)

(*ADVERBIAL* შეცვლა)

(*action* მიკუთვნება) (*agent* გვარი დედა)

(9) აქტის ბათილად ცნობა და მიყენებული მორალური და
მატერიალური ზიანის ანაზღაურების შესახებ

(9.1) (**ACTOR** ცნობა)

(**DIRECT** აქტი)

(**ADVERBIAL** ბათილად)

(**action** ცნობა) (**agent** აქტი)

(9.2) (**ACTOR** ანაზღაურება)

(**DIRECT** ზიანი)

(**ATTR** მიყენებული)

(**ATTR** მორალური)

(**ATTR** მატერიალური)

(**action** ანაზღაურება) (**agent** ზიანი)

(10) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება

(**ACTOR** განახლება)

(**DIRECT** საქმე)

(**ADVERBIAL** გარემოება (**LOWATTR** ახლად) (**LOWATTR**
აღმოჩენილი))

(**action** განახლება) (**agent** საქმე)

(11) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლების
შესახებ

(**ACTOR** განახლება)

(**DIRECT** საქმე)

(**ADVERBIAL** გარემოება (**LOWATTR** ახლად) (**LOWATTR**
აღმოჩენილი))

(**action** განახლება) (**agent** საქმე)

(12) ბავშვის საცხოვრებელი ადგილის განსაზღვრა, აღიშენების
დაკისრება

(12.1) (**ACTOR** განსაზღვრა)

(**DIRECT** ადგილი)

(**OBJECT2** ბავშვი)

(*ATTR* საცხოვრებელი)

(*action* განსაზღვრა) (*agent* ადგილი ბავშვი)

(12.2) (*ACTOR* დაკისრება)

(*DIRECT* ალიმენტი)

(*action* დაკისრება) (*agent* ალიმენტი)

(13) ბინიდან გამოსახლება და მატერიალური ზარალის
ანაზღაურება

(13.1) (*ACTOR* გამოსახლება)

(*ADVERBIAL* ბინა)

(*action* გამოსახლება) (*agent* -)

(13.2) (*ACTOR* ანაზღაურება)

(*DIRECT* ზარალი)

(*ATTR* მატერიალური)

(*action* ანაზღაურება) (*agent* ზარალი)

(14) ბინის ნასყიდობის ხელშეკრულების ბათილად ცნობა

(*ACTOR* ცნობა)

(*DIRECT* ხელშეკრულება)

(*OBJECT2* ბინა)

(*OBJECT3* ნასყიდობა)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* ხელშეკრულება)

(15) ბინის ყიდვა-გაყიდვის ხელშეკრულების ბათილად ცნობა

(*ACTOR* ცნობა)

(*DIRECT* ხელშეკრულება)

(*OBJECT2* ბინა)

(*OBJECT3* ყიდვა-გაყიდვა)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* ხელშეკრულება)

(16) გადაწვევების ბათილად ცნობა, ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების განახლება

(16.1) (*ACTOR* ცნობა)

(*DIRECT* გადაწვევება)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* გადაწვევება)

(16.2) (*ACTOR* განახლება)

(*DIRECT* წარმოება)

(*OBJECT2* საქმე)

(*ADVERBIAL* გარემოება (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილი))

(*action* განახლება) (*agent* წარმოება)

4.15. Маркирование адвербиативных семантических элементов формализованного представления содержания текстов юридических формул гражданского иска.

Адвербиативные элементы формализованного представления содержания текста являются семантически одним из наиболее нагруженных элементов текста, раскрывающим содержание как описанного актором действия, так и особенности его применения к агенту. Можно считать, что семантика текстов на естественных языках в большинстве случаев формализуется с помощью триады «актор – агент – адверб», тогда как другие элементы текстов на естественных языках играют, большей частью, иерархически подчиненную роль, уточняющую семантические свойства основной триады. Возможна также и другая точка зрения, рассматривающая триаду как сложно организованную, дифференцированную форму основной семантической пары «актор – агент», для которой важнейшие параметры, определяющие условия или обстоятельства применения действия, направленного от актора к агенту, для удобства их классификации и сравнения выделены в отдельное семантическое понятие, маркируемое особой словоформой.

В связи с этими интерпретациями в понятии адвербов большое значение приобретает их семантическая классификация, сравнительно слабо связанная с их синтаксической ролью, поскольку она в первую очередь определяет условия или обстоятельства применения действия, представленного актором. Эта классификация существенным образом определяется особенностями предметной области юридических формул гражданского иска, которая отражена в базе экспериментального исследования. Для рассмотренных текстов мы принимаем следующую систему классификации адвербов, естественно, не претендующую на полноту в случае необходимости анализа текстов широкого профиля (например, общих корпусов естественного языка).

4.15.1. Адverb юридического статуса, например, представляемый наречием «ბათილად, არაკანონიერად»; маркируется ключевым словом «*adv_status*».

4.15.2. Адverb имущественного статуса, например, представляемый наречиями «მესაკუთრედ, მემკვიდრედ»; маркируется ключевым словом «*adv_having*».

4.15.3. Адverb направления действия логической принадлежности или каузативной связи, например, представляемый послелоговой (постпозиционной) формой существительных или местоимений «მისგან, ბიბიდან, შვილთან, მშობლებთან»; маркируется ключевым словом «*adv_direction*».

4.15.4. Адverb места выполнения действия, например, представляемый послелоговой (постпозиционной) формой существительного «აქგში»; маркируется ключевым словом «*adv_location*».

4.15.5. Адverb логической связи, каузирующей действие, например, представляемый отделенной послелоговой (постпозиционной) формой существительного «გარემოებათა გამო»; маркируется ключевым словом «*adv_sequence*».

Эта задача решается модулем ADVERB_SEM. Алгоритм решения основан на анализе списковых структур, полученных на предыдущих шагах построения синтаксического описания текста и на приведенную выше классификацию адвербов. Синтаксическая неоднозначность при решении указанной задачи возникает практически во всех случаях, когда словоформа, представляющая адverb, не включена в классификацию адвербов.

Ниже приводится фрагмент формализованного представления содержания текстов юридических формул гражданского иска, в котором дополнительно маркируются адвербы, выявленные по принятому выше алгоритму.

4.16. Маркирование составных семантических элементов – вложенных агентов формализованного представления содержания текстов юридических формул гражданского иска.

Предметная область текстов юридических формул гражданского иска отражает весь богатый инструментарий естественного языка (в частности, грузинского), а именно, такую распространенную синтаксическую конструкцию, как вложенные (распространенные) дополнения. С их помощью семантика агента может быть существенно уточнена дополнительными словоформами, являющимися существительными в родительном падеже, синтаксически подчиненными агенту (прямому дополнению). Благодаря этому, строится сложный агент, имеющий иерархическую структуру, развивающую семантику основного агента, непосредственно подчиненного актору.

В связи с этими интерпретациями понятия агентов, имеющих вложенную структуру, важна их семантическая классификация, отражающая их синтаксическую роль, поскольку она определяет подчиненность объектов, к которым применяются действия, представленные актором. Эта классификация существенным образом определяется особенностями предметной области юридических формул гражданского иска, которая отражена в базе экспериментального исследования. Для рассмотренных текстов мы принимаем следующую систему классификации вложенных агентов, также не претендующую на полноту, в случае необходимости анализа текстов широкого профиля (например, общих корпусов естественного языка).

4.16.1. Подчиненный агент принадлежности или уточнения основного агента, образуемый родительным падежом одного существительного, который маркируется ключевым словом «*subagent*» и оформляется в виде однократной вложенной списковой структуры, помещенной внутри списка основного агента. Например:

– сочетание дополнений «დედის გვარის» формализуется как (*agent* გვარი (*subagent* დედა));

– сочетание дополнений «საქმის წარმოების» формализуется как (*agent* წარმოება (*subagent* საქმე));

– сочетание дополнений «ურთიერთობის დღეების» формализуется как (*agent* დღე (*subagent* ურთიერთობა));

4.16.2. Подчиненный агент принадлежности или уточнения основного агента, образуемый цепочкой родительных падежей нескольких существительных, также маркируется ключевым словом «*subagent*» и оформляется в виде многократно вложенной списковой структуры, помещенной внутри списка основного агента, причем, подчиненность дополнений однозначно отображается во вложенность списковых структур. Например:

– сочетание дополнений «ბინის ნასყიდობის ხელშეკრულების» формализуется как (*agent* ხელშეკრულება (*subagent* ბინა (*subagent* ნასყიდობა)))

4.16.3. Подчиненный агент, оппозиционный основному агенту, образуемый родительным падежом существительного в сочетании со специальным союзом «გარეშე», маркируется ключевым словом «*contragent*» и оформляется в виде вложенной списковой структуры, помещенной внутри списка основного агента, причем эта структура может произвольно сочетаться с формализацией рассмотренного выше подчиненного агента принадлежности или уточнения основного агента. Например:

– сочетание дополнений «ავტორის თანხმობის გარეშე ნაწარმოების» формализуется как (ნაწარმოები (*subagent* ავტორი (*contragent* თანხმობა)))

Эта задача решается модулем SUB_AGENT. Алгоритм решения основан на анализе списковых структур, полученных на предыдущих шагах построения синтаксического описания текста и на приведенной выше классификации агентов, а также на их взаимном расположении. Имеет место эвристическое правило, основанное на синтаксических традициях грузинского языка, что цепочка подчиненных агентов, являющихся существительными в родительном падеже, расположена в тексте в последовательности, обратной их синтаксическому подчинению.

Синтаксическая неоднозначность при решении указанной задачи возникает практически во всех случаях, когда словоформа, представляющая агента, нарушает указанное выше эвристическое правило.

Ниже приводится фрагмент формализованного представления содержания текстов юридических формул гражданского иска, в котором дополнительно маркируются вложенные агенты, выявленные по принятому выше алгоритму.

(1) ადმინისტრაციულ აქტში შეცდომის გასწორებისა და მისგან გამომწვეული სამართლებრივი შედეგების შეცვლის შესახებ

(1.1) (**ACTOR** გასწორება)

(**DIRECT** შეცდომა)

(**ADVERBIAL** აქტი (**LOWATTR** ადმინისტრაციული))

(**action** გასწორება) (**agent** შეცდომა) (**adv_location** აქტი)

(1.2) (**ACTOR** შეცვლა)

(**DIRECT** შედეგი)

(**ATTR** სამართლებრივი)

(**ADVERBIAL** მისი (**LOWATTR** გამომწვეული))

(**action** შეცვლა) (**agent** შედეგი) (**adv_direction** მისი)

(2) ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების აკრძალვისა და შესაბამისი კომპენსაციის გადახდის შესახებ

(2.1) (**ACTOR** აკრძალვა)

(**DIRECT** ნაწარმოები)

(**OBJECT2** ავტორი)

(**OBJECT3** თანხმობა (**LOWATTR** გამოყენებული))

(**action** აკრძალვა) (**agent** ნაწარმოები (**subagent** ავტორი (**contragent** თანხმობა)))

(2.2) (**ACTOR** გადახდა)

(**DIRECT** კომპენსაცია)

(**ATTR** შესაბამისი)

(**action** გადახდა) (**agent** კომპენსაცია)

(3) ალიმენტის დაკისრების და შვილთან ურთიერთობის დღეების დადგენის შესახებ

(3.1) (**ACTOR** დაკისრება)

(**DIRECT** ალიმენტი)

(**action** დაკისრება) (**agent** ალიმენტი)

(3.2) (**ACTOR** დადგენა)

(**DIRECT** დღე)

(**OBJECT2** ურთიერთობა)

(**ADVERBIAL** შვილი)

(**action** დადგენა) (**agent** დღე (**subagent** ურთიერთობა)) (**adv_direction** შვილი)

(4) ალიმენტის დანიშვნის და საცხოვრებელი ფართით უზრუნველყოფის შესახებ

(4.1) (**ACTOR** დანიშვნა)

(**DIRECT** ალიმენტი)

(**action** დანიშვნა) (**agent** ალიმენტი)

(4.2) (**ACTOR** უზრუნველყოფა)

(**DIRECT** ფართი)

(**ATTR** საცხოვრებელი)

(**action** უზრუნველყოფა) (**agent** ფართი)

(5) ანდერძის ბათილად ცნობის, მემკვიდრედ და მესაკუთრედ ცნობის შესახებ

(5.1) (**ACTOR** ცნობა)

(**DIRECT** ანდერძი)

(**ADVERBIAL** ბათილად)

(**action** ცნობა) (**agent** ანდერძი) (**adv_status** ბათილად)

(5.2) (**ACTOR** ცნობა)

(**ADVERBIAL** მემკვიდრედ)

(**ADVERBIAL** მესაკუთრედ)

(**action** ცნობა) (**agent** -) (**adv_having** მესაკუთრედ) (**adv_having** მემკვიდრედ)

(6) ანდერძის ნაწილობრივ ბათილობის და მემკვიდრედ ცნობა

(6.1) (**ACTOR** ბათილობა)

(**DIRECT** ანდერძი)

(**ATTR** ნაწილობრივი)

(**action** ბათილობა) (**agent** ანდერძი)

(6.2) (**ACTOR** ცნობა)

(**ADVERBIAL** მემკვიდრედ)

(**action** ცნობა) (**agent** -) (**adv_having** მემკვიდრედ)

(7) არასაცხოვრებელი ფართის არაკანონიერად მითვისების შესახებ

(**ACTOR** მითვისება)

(**DIRECT** ფართი)

(**ATTR** არასაცხოვრებელი)

(**ADVERBIAL** არაკანონიერად)

(**action** მითვისება) (**agent** ფართი) (**adv_status** არაკანონიერად)

(8) არასრულწლოვანი შვილის მშობლებთან დაბრუნების, შეცვლისათვის დედის გვარის მიკუთვნების შესახებ

(8.1) (**ACTOR** დაბრუნება)

(**DIRECT** შვილი)

(**ATTR** არასრულწლოვანი)

(**ADVERBIAL** მშობელი)

(**action** დაბრუნება) (**agent** შვილი) (**adv_direction** მშობელი)

(8.2) (**ACTOR** მიკუთვნება)

(**DIRECT** გვარი)

(**OBJECT2** დედა)

(**ADVERBIAL** შეცვლა)

(**action** მიკუთვნება) (**agent** გვარი (**subagent** დედა)) (**adv_direction**

შეცვლა)

(9) აქტის ბათილად ცნობა და მიყენებული მორალური და
მატერიალური ზიანის ანაზღაურების შესახებ

(9.1) (**ACTOR** ცნობა)

(**DIRECT** აქტი)

(**ADVERBIAL** ბათილად)

(**action** ცნობა) (**agent** აქტი) (**adv_status** ბათილად)

(9.2) (**ACTOR** ანაზღაურება)

(**DIRECT** ზიანი)

(**ATTR** მიყენებული)

(**ATTR** მორალური)

(**ATTR** მატერიალური)

(**action** ანაზღაურება) (**agent** ზიანი)

(10) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება

(**ACTOR** განახლება)

(**DIRECT** საქმე)

(**ADVERBIAL** გარემოება (**LOWATTR** ახლად) (**LOWATTR**
აღმოჩენილი))

(**action** განახლება) (**agent** საქმე) (**adv_sequence** გარემოება)

(11) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლების
შესახებ

(**ACTOR** განახლება)

(**DIRECT** საქმე)

(**ADVERBIAL** გარემოება (**LOWATTR** ახლად) (**LOWATTR**
აღმოჩენილი))

(**action** განახლება) (**agent** საქმე) (**adv_sequence** გარემოება)

(12) ბავშვის საცხოვრებელი ადგილის განსაზღვრა, აღიმენტის
დაკისრება

(12.1) (**ACTOR** განსაზღვრა)

(**DIRECT** ადგილი)

(**OBJECT2** ბავშვი)

(*ATTR* საცხოვრებელი)

(*action* განსაზღვრა) (*agent* ადგილი (*subagent* ბავშვი))

(12.2) (*ACTOR* დაკისრება)

(*DIRECT* ალიმენტი)

(*action* დაკისრება) (*agent* ალიმენტი)

(13) ბინიდან გამოსახლება და მატერიალური ზარალის
ანაზღაურება

(13.1) (*ACTOR* გამოსახლება)

(*ADVERBIAL* ბინა)

(*action* გამოსახლება) (*agent -*) (*adv_direction* ბინა)

(13.2) (*ACTOR* ანაზღაურება)

(*DIRECT* ზარალი)

(*ATTR* მატერიალური)

(*action* ანაზღაურება) (*agent* ზარალი)

(14) ბინის ნასყიდობის ხელშეკრულების ბათილად ცნობა

(*ACTOR* ცნობა)

(*DIRECT* ხელშეკრულება)

(*OBJECT2* ბინა)

(*OBJECT3* ნასყიდობა)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* ხელშეკრულება (*subagent* ბინა (*subagent*
ნასყიდობა))) (*adv_status* ბათილად)

(15) ბინის ყიდვა-გაყიდვის ხელშეკრულების ბათილად ცნობა

(*ACTOR* ცნობა)

(*DIRECT* ხელშეკრულება)

(*OBJECT2* ბინა)

(*OBJECT3* ყიდვა-გაყიდვა)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* ხელშეკრულება (*subagent* ბინა (*subagent*
ყიდვა-გაყიდვა))) (*adv_status* ბათილად)

(16) გადაწვევების ბათილად ცნობა, ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების განახლება

(16.1) (*ACTOR* ცნობა)

(*DIRECT* გადაწვევება)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* გადაწვევება) (*adv_status* ბათილად)

(16.2) (*ACTOR* განახლება)

(*DIRECT* წარმოება)

(*OBJECT2* საქმე)

(*ADVERBIAL* გარემოება (*LOWATTR* ახლად) (*LOWATTR* აღმოჩენილი))

(*action* განახლება) (*agent* წარმოება (*subagent* საქმე)) (*adv_sequence* გარემოება)

4.17. Маркирование семантических элементов – атрибутов в тексте формализованного представления содержания юридических формул гражданского иска.

Важной частью предметной области текстов юридических формул гражданского иска являются такие распространенные синтаксические конструкции, как определения (атрибуты), которые могут быть простыми и вложенными.

Синтаксическая сторона атрибутивных конструкций, в частности, их синтаксическая маркировка, была подробно рассмотрена выше, в 4.10 и 4.12. Маркирование порождаемых ими семантических элементов, играющих семантическую роль предикативного уточнения свойств объекта, на которое направлено действие актора, является достаточно хорошо формализуемой задачей, поскольку она опирается на выявленные на указанных шагах одноименные синтаксические элементы. С их помощью семантика агента может быть существенно уточнена дополнительными словоформами, являющимися прилагательными в номинативном (არასრულწლოვანი, მატერიალური, მორალური, საცხოვრებელი, სამართლებრივი, შესაბამისი и т.д.) или в линейном падеже (ნაწილობრივ, можно рассматривать и как наречие), а также

наречия (ახლად), синтаксически подчиненными агенту (прямому дополнению). Благодаря этому, уточняется иерархическая структура семантики основного агента, непосредственно подчиненного актору.

В связи с этими интерпретациями понятия атрибутов, порождаемых подчиненными или вложенными структурами, важна их семантическая классификация, отражающая их синтаксическую роль, поскольку она определяет подчиненность объектов, к которым применяются действия, уточняемые атрибутами. Эта классификация существенным образом определяется особенностями предметной области юридических формул гражданского иска, которая отражена в базе экспериментального исследования. Для рассмотренных текстов мы принимаем следующую систему классификации атрибутов, аналогично рассмотренным выше классификациям, не претендующую на полноту в случае необходимости анализа текстов широкого профиля (например, общих корпусов естественного языка).

4.17.1. Атрибуты, непосредственно подчиненные актору или агенту. Например:

საცხოვრებელი ფართით უზრუნველყოფა (4.2);

მატერიალური ზარალის ანაზღაურება (13.2).

Для маркирования подобных атрибутов используется ключевое слово «*attrib*», записываемое в виде добавочного списка в конце спискового представления семантики анализируемого текста (на уровне актора и агента).

4.17.2. Иерархически лежащие ниже атрибуты, подчиненные уже подчиненным агентам принадлежности или уточнения основного агента. Например:

ადმინისტრაციულ აქტში შეცდომის გასწორება (1.1);

ახლად აღმოჩენილი გარემოება (10).

Для маркирования подобных атрибутов используется ключевое слово «*exactly*», записываемое в виде добавочного списка в конце спискового представления семантики подчиненного агента (на уровне списка, определяющего подчиненный агент).

Эта задача решается модулем ATTRIBUTES. Алгоритм решения основан на анализе списковых структур, полученных на предыдущих шагах построения синтаксического описания текста и на приведенной выше классификации атрибутов, а также на их взаимном расположении. Имеет место эвристическое

правило, основанное на синтаксических традициях грузинского языка, что цепочка атрибутов, удовлетворяющих рассмотренным выше морфологическим условиям, расположена в тексте непосредственно перед словом, семантические свойства которых определяются (или уточняются) этим атрибутом.

Синтаксическая неоднозначность при решении указанной задачи возникает практически во всех случаях, когда словоформа, представляющая атрибут, нарушает указанное выше эвристическое правило.

Ниже приводится фрагмент формализованного представления содержания текстов юридических формул гражданского иска, в котором дополнительно маркируются как непосредственные, так и вложенные атрибуты, выявленные по принятому выше алгоритму.

(1) ადმინისტრაციულ აქტში შეცდომის გასწორებისა და მისგან გამომწვეული სამართლებრივი შედეგების შეცვლის შესახებ

(1.1) (**ACTOR** გასწორება)

(**DIRECT** შეცდომა)

(**ADVERBIAL** აქტი (**LOWATTR** ადმინისტრაციული))

(**action** გასწორება) (**agent** შეცდომა) (**adv_location** აქტი (**exactly** ადმინისტრაციული))

(1.2) (**ACTOR** შეცვლა)

(**DIRECT** შედეგი)

(**ATTR** სამართლებრივი)

(**ADVERBIAL** მისი (**LOWATTR** გამომწვეული))

(**action** შეცვლა) (**agent** შედეგი) (**adv_direction** მისი (**exactly** გამომწვეული)) (**attrib** სამართლებრივი)

(2) ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების აკრძალვისა და შესაბამისი კომპენსაციის გადახდის შესახებ

(2.1) (**ACTOR** აკრძალვა)

(**DIRECT** ნაწარმოები)

(**OBJECT2** ავტორი)

(**OBJECT3** თანხმობა (**LOWATTR** გამოყენებული))

(*action* აკრძალვა) (*agent* ნაწარმოები (*subagent* ავტორი (*contragent* თანხმობა (*exactly* თანხმობა)))

(2.2) (*ACTOR* გადახდა)

(*DIRECT* კომპენსაცია)

(*ATTR* შესაბამისი)

(*action* გადახდა) (*agent* კომპენსაცია) (*attrib* შესაბამისი)

(3) ალიმენტის დაკისრების და შვილთან ურთიერთობის დღეების დადგენის შესახებ

(3.1) (*ACTOR* დაკისრება)

(*DIRECT* ალიმენტი)

(*action* დაკისრება) (*agent* ალიმენტი)

(3.2) (*ACTOR* დადგენა)

(*DIRECT* დღე)

(*OBJECT2* ურთიერთობა)

(*ADVERBIAL* შვილი)

(*action* დადგენა) (*agent* დღე (*subagent* ურთიერთობა)) (*adv_direction* შვილი)

(4) ალიმენტის დანიშვნის და საცხოვრებელი ფართით უზრუნველყოფის შესახებ

(4.1) (*ACTOR* დანიშვნა)

(*DIRECT* ალიმენტი)

(*action* დანიშვნა) (*agent* ალიმენტი)

(4.2) (*ACTOR* უზრუნველყოფა)

(*DIRECT* ფართი)

(*ATTR* საცხოვრებელი)

(*action* უზრუნველყოფა) (*agent* ფართი) (*attrib* საცხოვრებელი)

(5) ანდერძის ბათილად ცნობის, მემკვიდრედ და მესაკუთრედ ცნობის შესახებ

(5.1) (*ACTOR* ცნობა)

(*DIRECT* ანდერძი)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* ანდერძი) (*adv_status* ბათილად)

(5.2) (*ACTOR* ცნობა)

(*ADVERBIAL* მემკვიდრედ)

(*ADVERBIAL* მესაკუთრედ)

(*action* ცნობა) (*agent -*) (*adv_having* მესაკუთრედ) (*adv_having* მემკვიდრედ)

(6) ანდერძის ნაწილობრივ ბათილობის და მემკვიდრედ ცნობა

(6.1) (*ACTOR* ბათილობა)

(*DIRECT* ანდერძი)

(*ATTR* ნაწილობრივი)

(*action* ბათილობა) (*agent* ანდერძი) (*attrib* ნაწილობრივი)

(6.2) (*ACTOR* ცნობა)

(*ADVERBIAL* მემკვიდრედ)

(*action* ცნობა) (*agent -*) (*adv_having* მემკვიდრედ)

(7) არასაცხოვრებელი ფართის არაკანონიერად მითვისების შესახებ

(*ACTOR* მითვისება)

(*DIRECT* ფართი)

(*ATTR* არასაცხოვრებელი)

(*ADVERBIAL* არაკანონიერად)

(*action* მითვისება) (*agent* ფართი) (*adv_status* არაკანონიერად) (*attrib* არასაცხოვრებელი)

(8) არასრულწლოვანი შვილის მშობლებთან დაბრუნების, შეცვლისათვის დედის გვარის მიკუთვნების შესახებ

(8.1) (*ACTOR* დაბრუნება)

(*DIRECT* შვილი)

(*ATTR* არასრულწლოვანი)

(*ADVERBIAL* მშობელი)

(*action* დაბრუნება) (*agent* შვილი) (*adv_direction* მშობელი) (*attrib* არასრულწლოვანი)

(8.2) (*ACTOR* მიკუთვნება)

(*DIRECT* გვარი)

(*OBJECT2* დედა)

(*ADVERBIAL* შეცვლა)

(*action* მიკუთვნება) (*agent* გვარი (*subagent* დედა)) (*adv_direction* შეცვლა)

(9) აქტის ბათილად ცნობა და მიყენებული მორალური და მატერიალური ზიანის ანაზღაურების შესახებ

(9.1) (*ACTOR* ცნობა)

(*DIRECT* აქტი)

(*ADVERBIAL* ბათილად)

(*action* ცნობა) (*agent* აქტი) (*adv_status* ბათილად)

(9.2) (*ACTOR* ანაზღაურება)

(*DIRECT* ზიანი)

(*ATTR* მიყენებული)

(*ATTR* მორალური)

(*ATTR* მატერიალური)

(*action* ანაზღაურება) (*agent* ზიანი) (*attrib* მიყენებული) (*attrib* მორალური) (*attrib* მატერიალური)

(10) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება

(*ACTOR* განახლება)

(*DIRECT* საქმე)

(*ADVERBIAL* გარემოება (*LOWATTR* ახლად) (*LOWATTR*

აღმოჩენილი))

(*action* განახლება) (*agent* საქმე) (*adv_sequence* გარემოება (*exactly* ახლად) (*exactly* აღმოჩენილი))

(11) ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლების შესახებ

(**ACTOR** განახლება)

(**DIRECT** საქმე)

(**ADVERBIAL** გარემოება (**LOWATTR** ახლად) (**LOWATTR** აღმოჩენილი))

(**action** განახლება) (**agent** საქმე) (**adv_sequence** გარემოება (**exactly** ახლად) (**exactly** აღმოჩენილი))

(12) ბავშვის საცხოვრებელი ადგილის განსაზღვრა, ალიმენტის დაკისრება

(12.1) (**ACTOR** განსაზღვრა)

(**DIRECT** ადგილი)

(**OBJECT2** ბავშვი)

(**ATTR** საცხოვრებელი)

(**action** განსაზღვრა) (**agent** ადგილი (**subagent** ბავშვი)) (**attrib** საცხოვრებელი)

(12.2) (**ACTOR** დაკისრება)

(**DIRECT** ალიმენტი)

(**action** დაკისრება) (**agent** ალიმენტი)

(13) ბინიდან გამოსახლება და მატერიალური ზარალის ანაზღაურება

(13.1) (**ACTOR** გამოსახლება)

(**ADVERBIAL** ბინა)

(**action** გამოსახლება) (**agent -**) (**adv_direction** ბინა)

(13.2) (**ACTOR** ანაზღაურება)

(**DIRECT** ზარალი)

(**ATTR** მატერიალური)

(**action** ანაზღაურება) (**agent** ზარალი) (**attrib** მატერიალური)

(14) ბინის ნასყიდობის ხელშეკრულების ბათილად ცნობა

(**ACTOR** ცნობა)

(**DIRECT** ხელშეკრულება)

(**OBJECT2** ბინა)

(**OBJECT3** ნასყიდობა)
(**ADVERBIAL** ბათილად)
(**action** ცნობა) (**agent** ხელშეკრულება (**subagent** ბინა (**subagent** ნასყიდობა))) (**adv_status** ბათილად)

(15) ბინის ყიდვა-გაყიდვის ხელშეკრულების ბათილად ცნობა
(**ACTOR** ცნობა)
(**DIRECT** ხელშეკრულება)
(**OBJECT2** ბინა)
(**OBJECT3** ყიდვა-გაყიდვა)
(**ADVERBIAL** ბათილად)
(**action** ცნობა) (**agent** ხელშეკრულება (**subagent** ბინა (**subagent** ყიდვა-გაყიდვა))) (**adv_status** ბათილად)

(16) გადაწყვეტილების ბათილად ცნობა, ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების განახლება

(16.1) (**ACTOR** ცნობა)
(**DIRECT** გადაწყვეტილება)
(**ADVERBIAL** ბათილად)
(**action** ცნობა) (**agent** გადაწყვეტილება) (**adv_status** ბათილად)
(16.2) (**ACTOR** განახლება)
(**DIRECT** წარმოება)
(**OBJECT2** საქმე)
(**ADVERBIAL** გარემოება (**LOWATTR** ახლად) (**LOWATTR** აღმოჩენილი))

(**action** განახლება) (**agent** წარმოება (**subagent** საქმე)) (**adv_sequence** გარემოება (**exactly** ახლად) (**exactly** აღმოჩენილი))

4.18. Пример решения задачи определения принадлежности текста судебного решения к статье классификатора судебной отчетности, путем установления релевантности двух текстов и сравнения маркированных представлений их семантики.

4.18.1. Даны два текста на ограниченном естественном грузинском языке, принадлежащие предметной области судебного делопроизводства по гражданским делам. Первый текст является формулировкой содержания конкретного гражданского иска, ранее рассмотренного как пример (3) в п.5.1.

(3) აღიძენტი დაკისრების და შვილთან ურთიერთობის დღეების დადგენის შესახებ

Второй текст является одной из статей действующего в настоящее время кодификатора форм статистической отчетности по гражданским делам судебных органов Грузии:

(35) ბავშვის შესახებ აღიძენტი გადახდევინების შესახებ

Требуется на основании изложенной в п.1.5 модели вычисления релевантности двух текстов подтвердить гипотезу об их релевантности или опровергнуть ее. При положительном решении, первый текст принимается относящимся к статье кодификатора форм статистической отчетности, представленной вторым текстом.

4.18.2. Результаты морфологического и синтаксического анализа обоих текстов по изложенным в п.4.1-4.13 алгоритмам, представлены в нормализованной форме.

(3.1) (*ACTOR* დაკისრება)

(*DIRECT* აღიძენტი)

(35) (*ACTOR* გადახდევინება)

(*DIRECT* აღიძენტი)

(*OBJECT2* ბავშვი)

(*ADVERBIAL* შესახებ)

4.18.3. Результаты построения модели семантики обоих текстов путем маркирования входящих в них семантических элементов.

(3.1) (*action* დაკისრება) (*agent* ალიმენტი)

(35) (*action* გადახდევინება) (*agent* ალიმენტი (*subagent* ბავშვი))

(*adv_having* შესანახად)

4.18.4. Экспертные правила эквивалентности между сущностями моделей семантики текстов.

При сравнении сущностей, входящих в модели семантики сравниваемых текстов, используется имеющее содержательную обоснованность экспертное правило, устанавливающее равнозначность двух акторов:

(3.1) (*action* დაკისრება)

(35) (*action* გადახდევინება)

Семантическая близость масдаров «დაკისრება» (применительно к оплате алимента) и «გადახდევინება» очевидна для носителей грузинского языка. Она, в частности, подтверждается лексикологическим материалом Большого грузино-английского словаря, в котором приводятся переводы обеих слов на английский язык, удостоверяющие близость семантического содержания этих масдаров:

(3.1) obliging (to do sth); charging (with task); giving responsibility

(35) making sb pay

4.18.5. Результаты сравнения списковых структур, представляющих модели семантики текстов.

С учетом экспертного правила эквивалентности 4.18.4, семантическое совпадение акторов можно считать установленным.

Семантическое совпадение агентов подтверждается точным текстуальным совпадением.

Согласно правилу установления релевантности 1.5.1, необходимо подтверждение правомерности игнорирования в (35) семантических элементов

(*subagent* ბავშვი)

(*adv_having* შესანახად)

Эта задача по своему содержанию близка к проблеме разрешения синтаксической и семантической неоднозначности и должна решаться в той же самой технологии, в которой снимается неоднозначность при установлении синтаксической и семантической структур текстов. В данном случае у эксперта-оператора, принимающего решение по релевантности сравниваемых списковых

структур, имеются вполне обоснованные аргументы для принятия решения по игнорированию подчиненных семантических элементов текста, являющихся субагентом (уточняющим не обязательный элемент текста – косвенное дополнение, указывающее принадлежность агента) и соответственно адвербом (уточняющим не обязательный элемент текста – имущественную принадлежность результата действия актора).

4.18.6. На основании приведенных выше формализованных рассуждений (с использованием экспертных оценок и интерактивного аппарата разрешения неоднозначности), можно установить справедливость гипотезы относительно релевантности текстов (3.1) и (35). Исходя из этого, можно утверждать, что текст судебного решения (3.1) может быть отнесен к статье (3) кодификатора форм статистической отчетности по гражданским делам судебных органов Грузии.

Классификатор судебных решений, принятый постановлением правительства Грузии от 05.10.2006.

<i>1</i>	<i>სანივთო სამართლებრივი დავები</i>	
	1	შამეზობლო
	2	დაკავშირებული უძრავ ქონებაზე საკუთრებასთან
	3	აღნაგობა
	4	ზუფრუქტი
	5	შერვიტუტი
	6	ირავნობა
	7	იპოთეკა
	8	სხვა სახის დავები
2	<i>ვალდებულებითი სამართლებრივი დავები</i>	
	9	ნასყიდობა, გაცვლა, ჩუქება
	10	ირავნობა
	11	იჯარა, სასოფლო-სამეურნეო მიწის იჯარა
	12	ლიზინგი, ფრენშიზინგი
	13	შესხი
	14	არღობა

	15	ტურისტული მომსახურება
	16	გადაზიდვა-გადაყვანა
	17	ავალება
	18	საკუთრების მინდობა
	19	სატრანსპორტო ექსპედიცია
	20	შუამავლობა
	21	ჯილდოს საჯაროდ დაპირება, კონკურსი
	22	მიბარება, მიბარება სასაქონლო საწყობში
	23	დაზღვევა
	24	საბანკო მოსახურება
	25	ფასიანი ქაღალდების ბრუნვიდან წარმოშობილი ვალდებულებანი
	26	ერთობლივი საქმიანობა (ამხანაგობა)
	27	სამისდღეშო რჩენა
	28	სხვა სახის დაგები
	29	კანონისმიერი ვლდებულებითი ურთიერთობანი
	30	დელიქტური ვალდებულებანი
3	ინტელექტუალური საკუთრება	
	31	საავტორო და მომიჯნავე უფლებებთან დაკავშირებული დაგები
	32	სამრეწველო საკუთრებასთან დაკავშირებული დაგები
4	საოჯახო სამართლებრივი დაგები	
	33	მეუღლეთა განქორწინების თაობაზე
	34	ქორწინების ბათილად ცნობის შესახებ
	35	ბავშვის შესანახად აღიშენების გადახდევინების შესახებ
	35	მამობის დადგენის შესახებ
	37	მშობლის უფლების ჩამორთმევის შესახებ
	38	სხვა სახის დაგები
5	მემკვიდრეობითი დაგები	
	39	მემკვიდრეობითი დაგები
6	სამეწარმეო სამართლებრივი დაგები	
	40	კერძო სამართლის იურიდიული პირების (კავშირები, ფონდები და ა.შ.) რეგისტრაციასთან დაკავშირებული დაგები
	41	სამეწარმეო საქმიანობასთან დაკავშირებული სხვა დაგები
	42	გაკოტრების საქმეებთან დაკავშირებული დაგები
7	შრომის სამართლებრივი დაგები	

	43	სამუშაოზე ადღგენისა და იძულებითი განაცდურის ანაზღაურების შესახებ
	44	მუშაკის სიცოცხლისა და ჯანმრთელობისათვის ზიანის მიყენებასთან დაკავშირებული დავები
	45	სხვა სახის დავები
	46	სულ შრომის სამართლებრივი დავებიდან ქალთა და არასრულწლოვანთა შრომასტან დაკავშირებული დავები
8	სარჩელები პირადი არაქონებრივი უფლების დაცვის თაობაზე	
	47	სიტყვისა და გამოსატვის თავისუფლებისა და ცილისწამების თაობაზე
	48	პატივისა და ღირსების დაცვის შესახებ
	49	პატივისა და ღირსების დაცვის შესახებ დაკავშირებული ქონებრივი ზიანის ანაზღაურებასთან
	50	სხვა სახის დავები
9	დეკრეტი	
	51	წარმოება რეკრუტი ქონების ჩამორთმევისა და სახელმწიფოსათვის გადაცემასთან დაკავშირებით
13	შხვა სასარჩელო საქმეები	
	52	სხვა სასარჩელო საქმეები
	53	სულ სასარჩელო წარმოების საქმეები
10	გამარტივებული წარმოება	
	54	თამასუქისა და ჩეკის თაობაზე
	55	დავალიანების გადახდის შესახებ
	56	სულ გამარტივებული წარმოების საქმეები
	57	მ.შ. საქმის განხილვა დავის სასარჩელო წარმოების წესით მოპასუხის შესაგებელის საფუძველზე
11	უდავო წარმოების საქმეები	
	58	საქმეები იურიდიული მნიშვნელობის ფაქტის დადგენის შესახებ
	59	მამობის ფაქტის დადგენა
	60	მოქალაქის უზოუკვლოდ დაკარგულად აღიარებისა და მოქალაქის გარდაცვლილად გამოცხადების თაობაზე
	61	მოქალაქის შეზღუდულ ქმედუნარიანად და ქმედუნაროდ აღიარების თაობაზე
	62	დაკარგულ ან განადგურებულ საწარმომადგენლო ფასიან ქაღალდსა და საორდერო ფასიან ქაღალდზე უფლების ადღგენის თაობაზე
	63	ქონების უპატრონოდ ცნობის თაობაზე

	64	შვილად აყვანის თაობაზე
	65	სულ უდავო წარმოების საქმეები
13	სხვა სასარჩელო საქმეები	
	66	სულ სამოქალაქო საქმეები
	67	მ.შ. უცხოური ინვესტორების მონაწილეობით
12	საქმის წარმოების განახლება	
	68	განცხადება გადაწყვეტილების (განჩინების) ბათილად ცნობის შესახებ
	69	განცხადება ახლად აღმოჩენილ გარემოებათა გამო

Табл. 4.18.6.1

4.19. Примеры установления релеванности текстов судебного решения и текстов классификатора судебной отчетности путем сравнения маркированных представлений их семантики (тексты всех примеров заимствованы из 4.1)

4.19.1. Текст формулировки содержания гражданского иска из примера (3).

ალიმენტის დაკისრების და შვილთან ურთიერთობის დღეების დადგენის შესახებ

Текст статьи (35) кодификатора форм статистической отчетности по гражданским делам:

ბავშვის შესანახად ალიმენტის გადახდევინების შესახებ

Гипотеза о релевантности этих двух текстов подтверждена в 4.18.

4.19.2. Текст формулировки содержания конкретного гражданского иска из примера (4).

ალიმენტის დანიშვნის და საცხოვრებელი ფართით უზრუნველყოფის შესახებ

Подтверждение гипотезы о релевантности этого текста тексту статьи (35) кодификатора форм статистической отчетности отличается от процедуры подтверждения гипотезы 4.18 (4.19.1) разве что только модификацией экспертного правила, устанавливающего равнозначность двух акторов:

(*action* დანიშვნა)

(*action* გადახდევინება)

4.19.3. Текст формулировки содержания конкретного гражданского иска из примера (12).

ბავშვის საცხოვრებელი ადგილის განსაზღვრა, აღიშენების დაკისრება

Подтверждение гипотезы о релевантности этого текста тексту статьи (35) кодификатора форм статистической отчетности полностью идентично процедуре подтверждения гипотезы 4.18 (4.19.1).

4.19.4. Текст формулировки содержания конкретного гражданского иска из примера (10).

ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლება

Маркированное представление семантики этого текста согласно принятой модели имеет вид:

(*action* განახლება) (*agent* საქმე (*adv_sequence* გარემოება (*exactly* ახლად) (*exactly* აღმოჩენილი))

Текст статьи (69) кодификатора форм статистической отчетности по гражданским делам:

განცხადება ახლად აღმოჩენილ გარემოებათა გამო

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид:

(*action* განცხადება) (*agent* გარემოება (*exactly* ახლად) (*exactly* აღმოჩენილი))

Сравнение списковых структур, представляющих модели семантики текстов, в данном случае сталкивается с существенными трудностями, связанными с тем, что в одном случае описываемые текстом сущности имеют характер конкретной цели «*საქმის განახლება*», а в другом фиксируется только побуждение к действию *განცხადება*. Экспертные правила в данном случае должны увязывать акт побуждения (*action* განცხადება), отражаемый классификатором, и факт реализации этого акта «*საქმის განახლება*», отражаемый в маркерной конструкции (*action* განახლება) (*agent* საქმე).

Корректность введения этого правила подтверждается близостью подчиненных конструкций:

(*adv_sequence* გარემოება (*exactly* ახლად) (*exactly* აღმოჩენილი))

(*agent* გარემოება (*exactly* ახლად) (*exactly* აღმოჩენილი)),

которые отличаются в данном случае только синтаксико-семантической интерпретацией слова «გარემოება». В первом случае оно подчинено агенту «საქმე», во втором является самостоятельным агентом, который, однако, элиминируется при отмеченной выше эквиваленции акта побуждения и реализации этого акта. Это обстоятельство позволяет, в данном случае, обосновать экспертную эквивалентность адверба логической связи (*adv_sequence* გარემოება), каузирующего действие находящегося на более высоком уровне синтаксической иерархии актора, и непосредственно заданного актора (*agent* გარემოება).

Совокупность указанных выше экспертных решений (подтверждаемых в диалоге с оператором) обосновывает подтверждение гипотезы о релевантности рассмотренных текстов.

4.19.5. Текст формулировки содержания конкретного гражданского иска из примера (11)

ახლად აღმოჩენილ გარემოებათა გამო საქმის განახლების შესახებ

Подтверждение гипотезы о релевантности этого текста тексту статьи (69) кодификатора форм статистической отчетности отличается от процедуры подтверждения гипотезы 4.19.4 только на морфологическом уровне представления актора либо с помощью прямого указания масдара «განახლება», либо с помощью синтаксической конструкции, связывающей этот масдар с помощью отделяемого послелога «განახლების შესახებ».

4.19.6. Текст формулировки содержания конкретного гражданского иска из примера (1).

აღმინისტრაციულ აქტში შეცდომის გასწორებისა და მისგან გამოწვეული სამართლებრივი შედეგების შეცვლის შესახებ

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид, состоящий из двух частей, вследствие наличия в тексте объединительного союза «და»

(*action* გასწორება) (*agent* შეცდომა) (*adv_location* აქტი (*exactly* აღმინისტრაციული))

(*action* შეცვლა) (*agent* შედეგი) (*adv_direction* მისი (*exactly* გამოწვეული)) (*attrib* სამართლებრივი)

Данная маркированная конструкция является контрпримером, демонстрирующим невозможность формального подтверждения гипотезы о релевантности этого текста тексту статьи (69) кодификатора форм статистической отчетности только путем использования рассмотренных выше простых экспертных правил, хотя с точки зрения интуитивной оценки, опирающейся на юридическую практику, их релеванность вполне убедительна. Создание необходимой для этого существенно более развитой системы экспертных правил, учитывающих, в частности, более тонкую природу взаимоотношений между побуждением и его реализацией в виде синтагмы «სამართლებრივი შედეგების შეცვლა», представляется делом будущего развития модели.

4.19.7. Текст формулировки содержания конкретного гражданского иска из примера (16).

გადაწყვეტილების ბათილად ცნობა, ახლად აღმოჩენილ გარემოებათა გამო საქმის წარმოების განახლება.

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид, состоящий из двух частей, вследствие наличия в тексте объединительного союза «და»

(*action* ცნობა) (*agent* გადაწყვეტილება) (*adv_status* ბათილად)

(*action* განახლება) (*agent* წარმოება (*subagent* საქმე)) (*adv_sequence* გარემოება (*exactly* ახლად) (*exactly* აღმოჩენილი)).

Текст статьи (68) кодификатора форм статистической отчетности по гражданским делам может быть представлен в двух равноценных вариантах:

განცხადება გადაწყვეტილების (განჩინების) ბათილად ცნობის შესახებ.

Маркированное представление семантики этого текста, согласно принятой модели, имеет две равнозначные формы:

(*action* განცხადება) (*agent* ცნობა) (*subagent* გადაწყვეტილება)
(*adv_status* ბათილად)

(*action* განცხადება) (*agent* ცნობა) (*subagent* განხილება) (*adv_status*
ბათილად)

Экспертные правила для подтверждения релевантности текстов сталкиваются с трудностями, отмеченными в 4.19.4. В одном случае, описываемые текстом сущности имеют характер конкретной цели «გადაწყვეტილების ბათილად ცნობა», а в другом, фиксируется только побуждение к действию განცხადება. Экспертные правила в данном случае должны увязывать акт побуждения (*action* განცხადება), отражаемый классификатором, и факт реализации этого акта «საქმის განახლება გადაწყვეტილების ბათილად ცნობის შესახებ», отражаемый в маркерной конструкции (*action* განახლება). Корректность введения этого правила подтверждается совпадением подчиненных синтаксических конструкции при элиминировании актора, описывающего акт побуждения, что приводит к следующему сопоставлению списковых структур:

(*action* ცნობა) (*agent* გადაწყვეტილება) (*adv_status* ბათილად)

(*agent* ცნობა) (*subagent* გადაწყვეტილება) (*adv_status* ბათილად)

Это обстоятельство позволяет обосновать экспертную эквивалентность, в данном случае, непосредственно заданного актора юридического действия (*action* ცნობა) и одноименного агента (*agent* ცნობა), каузирующего действие находящегося на более высоком уровне синтаксической иерархии актора побуждения (*action* განცხადება).

Совокупность указанных выше экспертных решений (подтверждаемых в диалоге с оператором) обосновывает подтверждение гипотезы о релевантности рассмотренных текстов.

4.19.8. Текст формулировки содержания конкретного гражданского иска из примера (5).

ანდერძის ბათილად ცნობის, მემკვიდრედ და მესაკუთრედ ცნობის შესახებ

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид (его первая часть, связанная со второй запятой, играет роль объединительного союза)

(*action* ბათილობა) (*agent* ანდერძი) (*attrib* ნაწილობრივი)

Подтверждение гипотезы об релевантности этого текста тексту статьи (68) кодификатора форм статистической отчетности по существу не отличается от рассмотренного выше примера 4.19.7, если принять следующие дополнительные экспертные правила, устанавливающие равнозначность синтаксико-семантических конструкций:

(*action* ბათილობა)

(*agent* ცნობა) (*adv_status* ბათილად),

а также равнозначность в данном юридическом контексте следующих двух агентов:

(*agent* გადაწყვეტილება)

(*agent* ანდერძი)

4.19.9. Текст формулировки содержания конкретного гражданского иска из примера (6).

ანდერძის ნაწილობრივ ბათილობის და მეტკვიდრედ ცნობა

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид (его первая часть связана со второй объединительным союзом და)

(*action* ბათილობა) (*agent* ანდერძი) (*attrib* ნაწილობრივი)

Подтверждение гипотезы о релевантности этого текста тексту статьи (68) кодификатора форм статистической отчетности также не отличается от рассмотренного выше примера 4.19.7, если элиминировать имеющий для классификации мало значительную юридическую нагрузку атрибут (*attrib* ნაწილობრივი)

4.19.10. Текст формулировки содержания конкретного гражданского иска из примера (9).

აქტის ბათილად ცნობა და მიყენებული მორალური და მატერიალური ზიანის ანაზრაურების შესახებ

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид (его первая часть связана со второй объединительным союзом *და*)

(*action* ცნობა) (*agent* აქტი) (*adv_status* ბათილად)

Подтверждение гипотезы о релевантности этого текста тексту статьи (68) кодификатора форм статистической отчетности также не отличается от рассмотренного выше примера 4.19.7, если ввести не вызывающее сомнений в юридической обоснованности эксперное правило равнозначности агентов

(*agent* გადაწვეტილება)

(*agent* აქტი)

4.19.11. Текст формулировки содержания конкретного гражданского иска из примера (14).

ბიზის ნასყიდობის ხელშეკრულების ბათილად ცნობა

Маркированное представление семантики этого текста, согласно принятой модели имеет вид:

(*action* ცნობა) (*agent* ხელშეკრულება (*subagent* ბიზა (*subagent* ნასყიდობა))) (*adv_status* ბათილად)

Подтверждение гипотезы о релевантности этого текста тексту статьи (68) кодификатора форм статистической отчетности возможно полностью аналогично предыдущим примерам, если ввести не вызывающее сомнений в юридической обоснованности эксперное правило равнозначности агентов

(*agent* გადაწვეტილება)

(*agent* ხელშეკრულება),

а также элиминировать имеющие для классификации мало значительную юридическую нагрузку субагенты (связанные в цепочку логической подчиненности)

(*subagent* ბიზა (*subagent* ნასყიდობა))

4.19.12. Текст формулировки содержания конкретного гражданского иска из примера (15).

ბიზის ყიდვა-გაყიდვის ხელშეკრულების ბათილად ცნობა

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид:

(*action* ცნობა) (*agent* ხელშეკრულება) (*subagent* ბინა) (*subagent* ყიდვა-გაყიდვა)) (*adv_status* ბათილად)

Подтверждение гипотезы о релевантности этого текста тексту статьи (68) кодификатора форм статистической отчетности возможно полностью аналогично примеру 4.19.11, если ввести не вызывающее сомнений в юридической обоснованности экспертное правило равнозначности субагентов, действующее в контексте установления релевантности

(*subagent* ყიდვა-გაყიდვა)

(*subagent* ნასყიდობა)

4.19.13. Текст формулировки содержания конкретного гражданского иска из примера (2).

ავტორის თანხმობის გარეშე გამოყენებული ნაწარმოების აკრძალვისა და შესაბამისი კომპენსაციის გადახდის შესახებ

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид, состоящий из двух частей, вследствие наличия в тексте объединительного союза «და»:

(*action* აკრძალვა) (*agent* ნაწარმოები) (*attrib* გამოყენებული) (*subagent* ავტორი) (*contragent* თანხმობა))

(*action* გადახდა) (*agent* კომპენსაცია) (*attrib* შესაბამისი)

Текст статьи (31) кодификатора форм статистической отчетности по гражданским делам:

საავტორო და მომიჯნავე უფლებებთან დაკავშირებული დავები

Маркированное представление семантики этого текста, согласно принятой модели, имеет вид (с некоторым упрощением, опуская альтернативу «მომიჯნავე»):

(*action* დავა) (*agent* უფლება) (*subagent* საავტორო)

Данная маркированная конструкция является также контрпримером, демонстрирующим невозможность формального подтверждения гипотезы о релевантности этого текста тексту статьи (69) кодификатора форм статистической отчетности только путем использования рассмотренных выше простых экспертных правил, хотя с точки зрения интуитивной оценки, опирающейся на юридическую практику, их релеванность тоже вполне

убедительна. Создание необходимой для этого существенно более развитой системы экспертных правил, учитывающих, в частности, более тонкую природу взаимоотношений между юридической акцией, представляемой синтагмой «საავტორო უფლებებთან დაკავშირებული დავა» и его реализацией в виде синтагмы «ავტორის თანხმობის გარეშე ნაწარმოების აკრძალვა», представляется делом будущего развития модели. К сожалению, в данном случае в тексте служебного решения отсутствует явное упоминание понятия «საავტორო უფლებები», равнозначность которого понятию «ავტორის თანხმობის გარეშე აკრძალვა» требует многошагового логического вывода об окружающем данное понятие миру авторских прав.

4.19.14. Текст формулировки содержания конкретного гражданского иска из примера (7).

არასაცხოვრებელი ფართის არაკანონიერად მითვისების შესახებ

Маркированное представление семантики этого текста: согласно принятой модели, имеет вид:

(*action* მითვისება) (*agent* ფართი) (*adv_status* არაკანონიერად) (*attrib* არასაცხოვრებელი).

Текст статьи (2) кодификатора форм статистической отчетности по гражданским делам:

დავები დაკავშირებული უძრავ ქონებაზე საკუთრებასთან

Данная маркированная конструкция является еще одним контрпримером, демонстрирующим невозможность формального подтверждения гипотезы о релевантности этого текста тексту статьи (2) кодификатора форм статистической отчетности только путем использования рассмотренных выше простых экспертных правил, хотя, с точки зрения интуитивной оценки, опирающейся на юридическую практику, и в данном случае их релеванность тоже вполне убедительна. В данном случае необходимо создание еще одной существенно более развитой системы экспертных правил, опирающейся на более развитый аппарат дедукивного вывода, который должна позволить установить логическую связь между использованным в формулировке содержания иска понятием «ფართის არაკანონიერი მითვისება» и имеющимся в классификаторе судебных решений понятием «უძრავ ქონებაზე საკუთრება». В данном случае, аналогично предыдущему, равнозначность этих

понятий требует многошагового логического вывода, опирающегося на развитые знания об окружающем данные понятия мирах авторских прав и на более обширную систему правил дедуктивного вывода.

4.19.15. Текст формулировки содержания конкретного гражданского иска из примера (8).

არასრულწლოვანი შვილის მშობლებთან დაბრუნების, შეცვლისათვის დედის გვარის მიკუთვნების შესახებ

Данный текст не может быть приписан ни к одной статье существующего классификатора судебных решений, поскольку ни в одной из них не охвачено ни юридическое понятие «შვილის მშობლებთან დაბრუნების», ни юридическое понятие «შეცვლისათვის გვარის მიკუთვნება». Это ограничение порождено не принятой в диссертации моделью, а является логическим следствием недостаточно развитой системы понятий самого классификатора, возникшем, скорее всего, из - за ограниченности числа статей, принципиально свойственным всем вручную обрабатываемым базам знаний. В подобных случаях эксперту-оператору, заполняющему документы судопроизводства, предлагается отнести это решение к статье «Прочее» (в данном случае к статье 38 «სხვა სახის დავები», входящей в группу статей «საოჯახო სამართლებრივი დავები»). Система «*Фемида*» в подобных случаях предлагает эксперту-оператору принять решение путем явного указания в диалоге статьи классификатора 38.

4.19.16. Текст формулировки содержания конкретного гражданского иска из примера (13).

ბიბიდან გამოსახლება და მატერიალური ზარალის ანაზღაურება

Данный текст также не может быть приписан ни к одной статье существующего классификатора судебных решений, поскольку ни в одной из них не охвачено ни юридическое понятие «ბიბიდან გამოსახლება», ни юридическое понятие «მატერიალური ზარალის ანაზღაურება». Это серьезное ограничения классификатора преодолимо только путем указания эксперта-оператора вручную статьи классификатора 8 «სხვა სახის დავები», входящей в группу статей «სახიფათო სამართლებრივი დავები». Указанные примеры являются подтверждением известного из теории АСУ положения, что

большинство используемых при описании реалий окружающего мира классификаторов не являются застывшими схемами, а требуют постоянного совершенствования и все более полного учета живой практики формализации отношений внутри человеческого общества, тем более в такой сложной и социально ответственной среде, как система знаний о юридическом делопроизводстве.

Глава 5. Автоматизированная система обработки документов гражданского судопроизводства Грузии «Фемида»

5.1. Общее описание автоматизированной системы «Фемида»

Автоматизированная система обработки документов гражданского судопроизводства Грузии «Фемида» является объектом внедрения рассмотренной в основной части диссертационной работы автоматизированной системы классификации юридическо-правовой документации, для которой внедряемая система является одной из информационно взаимодействующих подсистем интеллектуальной обработки данных, реализованной в виде специального программно-информационного модуля, предусматривающего диалоговое общение с пользователем.

Автоматизированная система обработки документов гражданского судопроизводства «Фемида» в целом предназначена для информационной и технологической поддержки судопроизводства, предоставления информационных услуг работникам районных судов и верховного суда Грузии.

Система «Фемида» обеспечивает автоматизацию процессов сбора, обработки, накопления, хранения, выдачи средств отображения информации и передачи по каналам связи следующих данных:

- о персональном составе органов судебной власти, действующих на территории Грузии;
- о сроках и состоянии хода рассмотрения гражданских дел в судах различных уровней;
- об оформлении судебных документов гражданских дел в процессе регистрации, ознакомления, подготовки, рассмотрения дел, вынесения постановлений, определений, приостановлении производства по делу, прекращении дела, передачи в архив и представления материалов хода судебного заседания;
- о результатах обработки судебной статистики с обеспечением объективного анализа;
- о состоянии завершенных гражданских дел;

- о нерассмотренных гражданских жалобах;
- о составе гражданских коллегий;
- о нагрузке судей гражданских коллегий;
- о количественных показателях судебной практики, включая использование базы данных (БД) судебных решений, а также информационно-аналитическую поддержку судебного делопроизводства;
- о делопроизводстве и документообороте судов по гражданским делам Грузии, включая вопросы контроля исполнения документов и поручений, организации оперативного и архивного хранения.

При этом достигается полная автоматизация следующих функции:

- судебного документооборота (коллективная разработка документов, поиск документов, обмен документами, контроль исполнения);
- правового обеспечения деятельности судов;
- обеспечения эксплуатации и сервисного обслуживания системы «Фемида»;
- формирования протокола хода судебного заседания;
- обеспечения защиты информации.

Защищённый информационный контур системы «Фемида» обеспечивает выполнение следующих функций:

- защиту электронного документооборота и делопроизводства с использованием различных средств защиты информации от несанкционированного доступа, методов и средств криптографической защиты информации для обеспечения конфиденциальности в сети;
- защиту передачи организации данных защищенного электронного документооборота между объектами;

Публичный информационный контур системы «Фемида» обеспечивает доступ к публичным сервисам сети Интернет, а также предоставляет услуги по ведению Интернет-сайта виртуальной приемной обращений граждан и т.д.

Система «Фемида» имеет многоуровневую иерархическую структуру, соответствующую уровням иерархии судов общей юрисдикции и системы судебного департамента.

Программная реализация системы «Фемида» представляет совокупность следующих систем обработки информации:

- система обработки информации, предназначенная для автоматизации Судебного департамента при Верховном Суде;

- система обработки информации, предназначенная для автоматизации районных судов;

- система обработки информации, предназначенная для автоматизации апелляционного суда;

Каждая из систем обработки информации представляет собой совокупность взаимосвязанных программно-технических средств, предназначенных для сбора, обработки, накопления, хранения, поиска и передачи информации.

В системе «Фемида» обеспечено как непрерывное круглосуточное функционирование всех комплексов средств автоматизации (режим работы OnLine), так и возможность автономного функционирования каждого комплекса средств автоматизации (режим работы OffLine).

Отдельные системы автоматизации всех уровней должны объединяться в информационную сеть с использованием государственных и коммерческих каналов, сетей связи и передачи данных.

Системы обработки информации, входящие в состав системы автоматизации «Фемида», строятся на базе структурированных локальных Ethernet-сетей, позволяющих динамично и оперативно наращивать вычислительные и информационные ресурсы, число пользователей сети, в том числе с учетом территориальной распределённости, логического объединения пользователей в сегменты, распределения доступа пользователей к ресурсам сети, организации технологических коммуникационных помещений, стоечного размещения технических средств.

Каждому уровню автоматизации судов и иерархической структуры судебного департамента соответствует свой комплекс средств автоматизации, единичный для отдельно выделенных объектов и тиражируемый для неоднократно повторяющихся. В структуре каждого реализуются типовые технические решения для однотипных функциональных узлов.

В число типовых функциональных узлов входят:

- серверное оборудование;

- автоматизированные рабочие места пользователей;

- автоматизированные рабочие места административного персонала;

- средства системы обеспечения безопасности информации;
- средства подсистем связи и передачи данных;
- средства отображения информации (видеопроекторы, плазменные панели, телевизоры, информационные киоски);
- активное сетевое оборудование.

Состав и характеристики технических средств, входящих в типовые функциональные узлы, в зависимости от назначения определённой автоматизированной системы, могут в некоторой степени количественно и по производительности отличаться, сохраняя при этом свое основное функциональное назначение.

Требования по защите информации и защищенности корпоративной сети «Фемида», в том числе и по техническим каналам связи, отражаются в ограничениях на организацию взаимодействия абонентов с сетью Интернет. Для гарантированной защиты сети Интранет от несанкционированных действий и вирусных атак, в том числе со стороны сети Интернет, аппаратные решения по телекоммуникационному оборудованию предусматривают физическую развязку средств хранения и обработки информации сетей системы «Фемида» и Интернет.

При разработке структур систем автоматизации приняты одинаковые для всех технические решения:

- состав серверного оборудования каждой системы укомплектовывается сервером базы данных, совмещенным с сервером электронной почты и сервером приложений;
- в состав каждого рабочего места, помимо системного блока, клавиатуры, манипулятора «мышь», монитора, входят лазерный принтер формата А4 и источник бесперебойного питания (UPS);
- серверное и активное сетевое оборудование собираются в стойки, укомплектованные соответствующими по мощности источниками бесперебойного питания;
- рабочие места всех администраторов комплектуются приводом DVD+RW Drive.

Автоматизированная система спроектирована с учетом обеспечения высокого реагирования, высокой защищенности и надежности, централизованного хранения данных. Ядром обмена данными между серверным

оборудованием и рабочими местами пользователей являются магистральные высокоскоростные коммутаторы со скоростью портов 100 Мбит/сек и 1 Гбит/сек. По этим портам с использованием сетевых интерфейсов 1000Base-T и 100Base-TX производится подключение серверов, обеспечивающих функционирование подсистем. Повышенная надежность достигается дублированием магистральных коммутаторов и наличием двух сетевых карт в серверах, одновременно подключенных к портам обоих коммутаторов. Для создания гибкой сетевой инфраструктуры в качестве коммутаторов используются коммутаторы, поддерживающие стандарт IEEE 802.1Q по организации виртуальных сетей (VLAN-сети).

Изолирование отдельных узлов локальной сети на канальном уровне с использованием технологии виртуальных сетей позволяет решить несколько задач, связанных с эффективностью сетевого обмена:

- повышение производительности локальных сетей путем локализации соответствующего трафика в пределах виртуальной сети;
- пересылка соответствующих пакетов, предназначенных для определенной группы пользователей внутри виртуальной сети, а не между виртуальными сетями;
- повышение безопасности сети за счет обеспечения недоступности части ресурсов сети для определенных категорий пользователей. Для централизованного хранения данных используется выделенное хранилище данных, построенное по принципу RAID50, что обеспечивает высокую надежность хранения и доступность данных. Хранилище данных подключается к серверам через два параллельно работающих коммутатора, один из которых работает в качестве основного, второй – дополнительного, подключаясь при выходе из строя основного коммутатора. К коммутаторам также подключена библиотека резервного копирования Базы Данных, развернутой на сервере. Данная конфигурация обеспечивает высокоскоростной доступ к данным и гарантирует, что при выходе любого сервера из строя, данные не будут потеряны.

5.2. Описание функционирования автоматизированной системы «Фемида»

Автоматизированная система обработки документов судопроизводства Грузии **«Фемида»**, подобно всем современным компьютеризованным системам, создана в помощь ведения документооборота судебных работников для осуществления судебного процесса без использования (или с минимальным использованием) бумажных документов. Основная задача системы **«Фемида»** – облегчить работу судебных работников, избавив их от излишней бумажной волокиты. Одновременно с этим система существенно облегчает встречающиеся в судебной практике задачи поиска как конкретных дел, так и фигурантов, участвующих в процессе. Отдельной задачей системы **«Фемида»** является автоматический контроль соблюдения сроков рассмотрения дел, исключая механические ошибки персонала, а также подсчета статистических данных по завершившимся и происходящим судебным делам.

С точки зрения эргономического эффекта работы системы **«Фемида»**, ее важнейшей положительной стороной является освобождение служащих судов от очень большой части рутинной работы, поскольку многочисленные документы, регламентирующие отчетность по работе гражданских судов, заполняются автоматически почти без участия персонала. Другим важным фактором является снятие контроля сроков рассмотрения дел со служащих судов, что в значительной степени повышает авторитет судебного производства. Наконец, использование системы **«Фемида»** позволяет предотвратить возможность случайной утери важных документов в процессе перехода дела от одного юридического состояния к другому, что также важно для повышения социального авторитета судов.

Для примера в качестве иллюстрации приведем типичный пример автоматизированного документооборота в процессе работы персонала суда с конкретным делом, которое обеспечивается поддержкой автоматизированной системы **«Фемида»**.

5.2.1. Рассматривается дело о назначении выплаты алиментов. Истец Ирма Ахалкаци (имя условное) подает в суд на своего бывшего супруга с требованием назначения выплаты алиментов и определения дней свиданий отца с ребенком.

Первым этапом в системе является создание искового заявления (рис. 5.2.1.1), в ходе которого работник канцелярии заполняет приведенную ниже форму **სარჩელის შეტანა**.

დასრულება

პირითადი მონაცემები	დამატებითი მონაცემები						
1. წარმოების წესი	საერთო წესი						
1. სარჩელის ნომერი	7655						
2. შემოსვლის თარიღი	02/09/2005						
3. შემოსვლის წესი	პირველად						
4. სადავო ღირებულება	0						
5. პაის ოდენობა	0						
6. მოსარჩელე	<table border="1"> <tr> <td>გვარი, სახელი ან ორგანიზაციის დასახელება</td> <td>მისამართი</td> <td>ტ</td> </tr> <tr> <td>ახალკაცი ირმა</td> <td>თბილისი, გლდანი 3 მ. 45კ. 3ბ</td> <td></td> </tr> </table>	გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ	ახალკაცი ირმა	თბილისი, გლდანი 3 მ. 45კ. 3ბ	
გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ					
ახალკაცი ირმა	თბილისი, გლდანი 3 მ. 45კ. 3ბ						
7. წარმომადგენელი	<table border="1"> <tr> <td>გვარი, სახელი ან ორგანიზაციის დასახელება</td> <td>მისამართი</td> <td>ტ</td> </tr> <tr> <td></td> <td></td> <td></td> </tr> </table>	გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ			
გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ					
8. პოპასუხე	<table border="1"> <tr> <td>გვარი, სახელი ან ორგანიზაციის დასახელება</td> <td>მისამართი</td> <td>ტ</td> </tr> <tr> <td>სონდულაშვილი უშანგი</td> <td>თბილისი, ლილო 5კ. 15ბ.</td> <td></td> </tr> </table>	გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ	სონდულაშვილი უშანგი	თბილისი, ლილო 5კ. 15ბ.	
გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ					
სონდულაშვილი უშანგი	თბილისი, ლილო 5კ. 15ბ.						
9. წარმომადგენელი	<table border="1"> <tr> <td>გვარი, სახელი ან ორგანიზაციის დასახელება</td> <td>მისამართი</td> <td>ტ</td> </tr> <tr> <td></td> <td></td> <td></td> </tr> </table>	გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ			
გვარი, სახელი ან ორგანიზაციის დასახელება	მისამართი	ტ					
10. სარჩელის არსი (მოსარჩელის მოთხოვნა)	ალიმენტის დაკისრების და შეიღთან ურთიერთობის დღეების დადგენის შესახებ						

Рис. 5.2.1.1

В данную форму вводится подробная информация о сторонах дела и об исковых требованиях истца. Специальные диалоги помогают пройти процесс создания иска максимально просто. На рисунке 5.2.1.2 представлена ключевая паспортная информация об истце:

- имя,
- фамилия,
- адрес места проживания,
- телефон,
- поле для дополнительной информации.

The screenshot shows a web form titled "მოსარჩელე" (Plaintiff) in a blue header. The form contains several input fields with the following labels and values:

- Label: გვარი, სახელი ან ორგანიზაციის დასახელება (Surname, name or organization name). Value: ახალკაცი ორმა (Ahalkatsi Orma).
- Label: მისამართი (Address). Value: თბილისი, გლდანის რაიონი 3 მ. 45კ. 34ბ (Tbilisi, Gldani district, 3 m. 45k. 34b).
- Label: ტელეფონი (Phone). Value: 675843.
- Label: სხვა (Other). Value: -.

At the bottom of the form, there are two buttons: "ნაიწეროს" (Save) with a floppy disk icon and "დასრულება" (Finish) with a checkmark icon.

Рис. 5.2.1.2

Аналогичные паспортные данные вводятся и для других участников процесса:

ответчика,
представителей сторон,
третьих лиц,
например:

Рис. 5.2.1.3

Для каждой группы вводимой паспортной информации после ее ввода оператору предоставляются две альтернативы: либо (после визуальной проверки и, при необходимости, корректировки) подтвердить введенные данные и передать их в область анкеты, либо отказаться от их ввода и вернуться к основному меню создания искового заявления.

Особо следует отметить поле искового требования. В нем работник суда, на которого возложена ответственность за оформление и принятие дела в

производство, записывает на естественном грузинском языке составленную им формулировку содержания иска, который предъявляет истец ответчику. Как было отмечено в гл.1, текст этой формулировки имеет в значительной степени субъективный характер и в значительной степени основан на опыте, квалификации и юридической культуре ответственного за создание искового заявления лица. С другой стороны, этот текст не может рассматриваться просто как комментарий к содержанию оформляемого иска, не имеющий значения для его юридической квалификации, поскольку на основании текста формулировки содержания иска ответственное лицо определяет важнейший параметр судебного делопроизводства – юридически формализованную категорию дела, выбираемую из утвержденных вышестоящими органами классификаторов, которое в дальнейшем служит ключевым реквизитом как для задач поиска, так и при формировании статистической отчетности. Трудности, возникающие в этом процессе, по существу являющимся поиском формального образа классифицируемого текста на ограниченном естественном языке, хорошо известны в практике судебного делопроизводства и были подробно обсуждены в гл.1 при обосновании актуальности настоящей диссертационной работы.

В рассматриваемом примере приводится образец квалифицированного выполнения записи формулировки содержания иска для данного конкретного дела

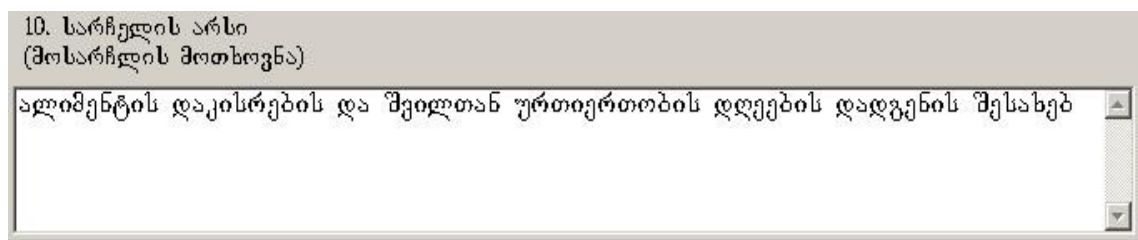


Рис. 5.2.1.4

5.2.2. Каждое из вводимых при создании искового заявления (рис. 5.2.1.1) данное обрабатывается в произвольном порядке по единому сценарию: ввод – контроль – подтверждение их корректности – занесение в анкету.

Проверка корректности каждого из данных производится специализированными «прозрачными» программными модулями, вступающими в диалог с оператором только в случае, когда они обнаруживают ошибку в данных и предлагают выполнить повторный ввод или корректировку данного.

Алгоритмы коррекции основаны на известных из практики создания программного обеспечения автоматизированных систем управления принципах:

- ограничения на класс символов, содержащихся в данном (только буквы, только цифры, дефис и т.д.);
- ограничения на минимальное и максимальное значение данного (даты, значение судебных издержек);
- ограничения на принадлежность данного определенному множеству (множество всех фамилий, множество всех имен);
- ограничения на структуру данного (отчество);
- ограничения на контекстную зависимость данного (сравнение текущей даты с датой рождения участников делопроизводства и т.д.)

После подтверждения всех требуемых анкетой исходных данных по создаваемому делу и их сохранения в базе данных система переходит в режим выдачи списка всех принятых заявлений (рис 5.2.2.1), в котором курсор установлен на только что введенное дело. Каждый элемент списка занимает одну строку и содержит краткое описание дела и текст его содержания.

Навигация по списку осуществляется стандартным набором клавиш продвижения по списку вверх и вниз как на одну строку, так и на страницу, а также с помощью вспомогательной системы поиска.

После выборки интересующего оператора дела возможно возвращение к корректировке некоторых его параметров (в пределах ограничений, установленной юридическими процедурами, запрещающими изменение существенных реквизитов иска после его принятия в делопроизводство), а также переход к дальнейшим регламентированным процедурам судебного производства по данному делу.

მონიშნული სარჩელის ნახვა სარჩელის რედაქტირება				
დასრულება				
ნომერი	შემოვიდა	მოსარჩელე	მომასუსე	დავის არსი
7648	03/10/2005	საღუჭებამ თამარი	მაისურამე რამაზი	განკორწინება
7649	04/10/2005	იანშიელი გურამი	მთაწმინდა-კრწანისის	პოლიტიკური რეპრესიის მსხვერპლად აღიარება
7650	04/10/2005	კვშულარია ამირანი	ვარლიანშვილი ივანე	თანხის დაკისრება
7651	04/10/2005	ბრიუკოვა ალა	უნიკაშვილი დევიზა - ნ	ნათესაური კავშირის დადგენა, მემკვიდრედ აღიარება
7652	29/08/2005	სარიტონაშვილი დარეჯა	გეფხვაძე დავით	გადახდის ბრძანების მიღების შესახებ
7653	13/09/2005	ოსამე ელაზა	ინწკორელი დ. - ნოტა	მესაკუთრედ ცნობა
7654	05/09/2005	თეთრაძე სოსო	აბულაძე ბაღირი	უკანონო მშენებლობის შეჩერების, ნაგებობის დასრულების
7655	02/09/2005	ახალკაცი ირმა	სონღულაშვილი უმანგი	ადამიანის დაკისრება
7656	23/09/2005	პურტენაგა ანზორი		იურიდიული ფაქტი
7657	21/09/2005	მაყიშვილი გიორგი	გაზეთი "ახალი ვერსია	მორალური ზიანის ანაზღაურება
7658	05/10/2005	საქ თავდაცვის სამინ	თვარაძე კახაბერი	ზიანის ანაზღაურება
7659	20/09/2005	გავაშვილი ნატალია		იურიდიული ფაქტი
7660	09/09/2005	შპს "ჯორჯია ონლაინი"	ხარშილაძე	გადახდის ბრძანება
7661	05/10/2005	ცქვიტინიძე ალექსანდრე		რეპრესიის მსხვერპლად აღიარება
7662	05/10/2005	ჯაფარიშვილი ვაჟა	აბესაძე ამირანი	გადახდის ბრძანება
7663	05/10/2005	ბილიხოძე-შენგულია ნ.		რეპრესიის მსხვერპლად აღიარება
7664	27/09/2005	მანდრიცკია კსენია	მეულია ლ.	მინდობილობისა და ნასყიდობის ხელშეკრულების დასრულება
7665	05/10/2005	კვნიანიშვილი გენერა		რეპრესიის მსხვერპლად აღიარება
7666	22/09/2005	ორმოცაძე სოფიო		მამობის დადგენა
7667	05/10/2005	სს "პროკრედიტბანკი"	ი/ს "თ. მერკელიაშვილი	დავალიანების გადახდის დაკისრების, იპოთეკის დასრულების
7668	09/09/2005	ფონდი "კონსტანტა"	ჩინჩილაძე ნადია	გადახდის ბრძანება
7669	05/10/2005	ზაზუნაშვილი ვახტანგი		პოლიტიკური რეპრესიის მსხვერპლად აღიარება
7670	08/09/2005	საქ თავდაცვის სამინ	ნინიაშვილი ზაზა	ზიანის ანაზღაურება
7671	08/09/2005	საქართველოს თანდაცვი	ბაბუნაშვილი ზვიადი	ზიანის ანაზღაურება
7672	06/09/2005	საქ თავდაცვის სამინ	მაცასარაშვილი მურაბი	ზიანის ანაზღაურება
7673	08/09/2005	საქ თავდაცვის სამინ	გულაშვილი ნიკოლოზი	ზიანის ანაზღაურება
7674	06/09/2005	საქ თავდაცვის სამინ	რუსაძე შაველი	ზიანის ანაზღაურება
7675	06/09/2005	საქ თავდაცვის სამინ	გორგილაძე ვიტალი	ზიანის ანაზღაურება
7676	04/08/2005	სახ. ქონების აღრიცხვ	ჯანელიძე თემური	საიჯარო ხელშეკრულების მოშლის, ძირის დასრულების შესახებ
7677	24/08/2005	სს "საქართველოს ელექტროენერჯიტი"	შპს "ტე-პენტერა და კ	თანხის დაბრუნების შესახებ
7678	28/07/2005	ჯანელიძე გია	სს "თბილგაზი"	ზიანის ანაზღაურება
7679	28/07/2005	შპს "გური"	ბაქო-თბილისი-ჯეიჰანი	მატერიალური ზიანის ანაზღაურება
7680	15/07/2005	ჩხიკვაძე ვახტანგი	ჯაფარი ნოდარი	უკანონო ავტოფარგის მშენებლობის აკრძივების დასრულება
7681	20/06/2005	ზომერიკი ქეთევანი	სამშკომპ"კავშირბენი	სადასრულებო ფურცლის ბათილად ცნობა,

Рис 5.2.2.1

Важной частью данных формализованного образа дела, помещенного в базу данных, является приписанный ему уникальный номер, который является уникальным ключевым реквизитом (идентификатором), однозначно определяющим соответствующую конкретному делу запись в базе данных. Существуют различные алгоритмы формирования номера дела, обычно опирающиеся на традиции судебного делопроизводства без использования средств автоматизации; в настоящей версии системы «Фемида» принята последовательная цифровая нумерация дел в порядке времени их поступления и создания их образа в базе данных.

5.2.3. После создания дела по исковому заявлению следующим юридическим этапом этого рассмотрения является назначение судьи, ответственного за ведение судебных заседаний по рассмотрению этого дела, в том числе обязанного в течение трех дней принять иск к рассмотрению, либо отклонить его.

Назначение судьи по рассмотрению данного иска производится Председателем суда, что является его важнейшей функцией в судебном производстве.

В системе «Фемида» назначение судьи не решается средствами самой системы, а производится путем диалога с ответственным работником суда, отвечающим за ввод в систему решения Председателя суда (по мере повышения квалификации судебного персонала – самим Председателем суда). При этом система обеспечивает поддержку принятия этого важнейшего решения, предоставляя в режиме диалога как список всех судей данного судебного органа, имеющих право исполнять эти функции, так и текущую занятость каждого из судей. Предусматривается также вспомогательное сервисное программное обеспечение для корректировки списка судей (изменения, добавления и удаления записей).

კოდი	მოსამართლის გვარი, სახელი	კოდი	მოსამართლის გვარი, სახელი	საკმ.რაოდენობა	ბოლო საქმე
1	ღავეთ ღიბრაძე	4	ია შარიაშიძე	336	11/12/2006
2	თამაზ ჯაღლიაშვილი	3	თინათინ ფეიქრიშვილი	339	21/12/2006
3	თინათინ ფეიქრიშვილი	1	ღავეთ ღიბრაძე	947	20/03/2007
4	ია შარიაშიძე	28	მაჭავარიანი სოფიკო	32	20/03/2007
5	ნონა ზარქუა	14	ცისკარიძე ეკატერინე	877	02/04/2007
6	არსენ შაშვილი დედოფა	12	ზამბახიძე თამარი	426	02/04/2007
7	ირაკლი ადგიშვილი	19	ჩინლაძე თამარი	536	03/04/2007
8	პაატა ქათამაძე	21	ჭელია შაკა	589	05/04/2007
9	აბუქანდაძე დავი	8	პაატა ქათამაძე	155	11/04/2007
10	ავალიშვილი ლია	16	სოსხვილი-ნიკოლაიშვილი თეა	711	13/04/2007
11	ბურჯაშვილი დიანა	13	კონრეძე ასპათი	1061	16/04/2007
12	ზამბახიძე თამარი	7	ირაკლი ადგიშვილი	223	16/04/2007
13	კონრეძე ასპათი	10	ავალიშვილი ლია	986	17/04/2007
14	ცისკარიძე ეკატერინე	27	სულდაკაძე გიორგი	343	17/04/2007
15	კუჭავა ქეთევანი	5	ნონა ზარქუა	1227	18/04/2007
16	სოსხვილი-ნიკოლაიშვილი თეა	17	ტუცუაძე ელისო	463	20/04/2007
		18	წიკლაური ვანო	517	22/04/2007

Рис. 5.2.3.1

5.2.4. Непосредственно после назначения судьи, он принимает решение о возможности приема иска в судебное производство (или отклоняет его). При положительном решении ответственный специалист канцелярии суда выполняет в системе «Фемида» технические действия, вводящие в базу данных информацию об открытии дела, в результате чего выдается окно открытия дела (рис. 5.2.4.1).

განხილვის განმარტება და ნაწილაკი დახარულება

საწესი განჩინების გამოტანა

განჩინების ტექსტი

სარჩელის წარმოებაში მიღება

სარჩელის შეცვლის შედეგად წარმოებაში მიღება

კომპ. კოდი: 0007654

საქმის ნომერი: 2/3283-06

გადაცემულია: 11/11/2005

სარჩელში არსებული მისარჩელის მოთხოვნა

აღიშენების დაკისრების და შეიღთან ურთიერთობის დღეების დადგენის შესახებ

დავის კატეგორია: 35

დავის სახეობა: ბაქმის შესახებ აღიშენების გადახდების შესახებ

წარმოებაშია: 07/09/2005

განხილვის დაგეგმილი ვადა: 07/10/2005

უცხოური ინვესტორი შემოცმებულია

სარჩელის ნომერი	შემოსვლის თარიღი	მისარჩელე	მოპასუხე	დავის არსი
7655	02/09/2005	ახალკაცი ირმა	სონღულაშვილი უზუნგი	აღიშენის დაკისრება

Рис. 5.2.4.1

Значительная часть информации в окне открытия дела повторяет данные, уже введенные или сформированные в окне создания искового заявления. Однако в новом окне имеются важные данные, вводимые в режиме диалога, среди которых особое значение имеет код категории дела, являющийся формальным образом его содержания, записанного ранее в поле искового требования в виде текста на естественном грузинском языке (см. 5.2.1). Это поле определяет юридически формализованную категорию дела, выбираемую из утвержденных вышестоящими органами классификаторов, которое в дальнейшем служит ключевым реквизитом как для задач поиска, так и при формировании статистической отчетности. В первоначальном варианте системы «Фемида» поле кода категории дела заполнялось ответственным работником суда, отвечающим за правильность поисковой и статистической информации по

всей совокупности хранящихся в базе данных дел, для чего в текущем окне предусматривалось всплывающееся меню, строки которого являлись классификатором содержания гражданских дел. Однако, как отмечалось в гл.1, принимаемое в этих случаях решение оператора по целому ряду субъективных и объективных причин недопустимо, часто является неадекватным фактическому содержанию дела, что влечет за собой заметное снижение качества судебного производства со всеми вытекающими отсюда социальными последствиями.

В настоящей диссертации, результаты которой легли в основу расширенной версии системы «Фемида», предлагается механизм контроля соответствия между текстом формулировки содержания дела и его формальным классификационным кодом. Практическое использование разработанных в диссертации алгоритмов проверки корректности кода классификации содержания дела изложено в п. 4.19.

После принятия решения судьи по рассматриваемому делу (удовлетворить иск, отклонить или удовлетворить частично) автоматически генерируется текст решения, который может быть выдан в виде печатной формы. Благодаря рассмотренной выше структуре окон ввода, отражающей базу данных судебного делопроизводства по гражданским делам, выдача в печать выходного документа – первичного постановления суда, устанвливающего юридический факт открытия дела, производится вводом соответствующей директивы.

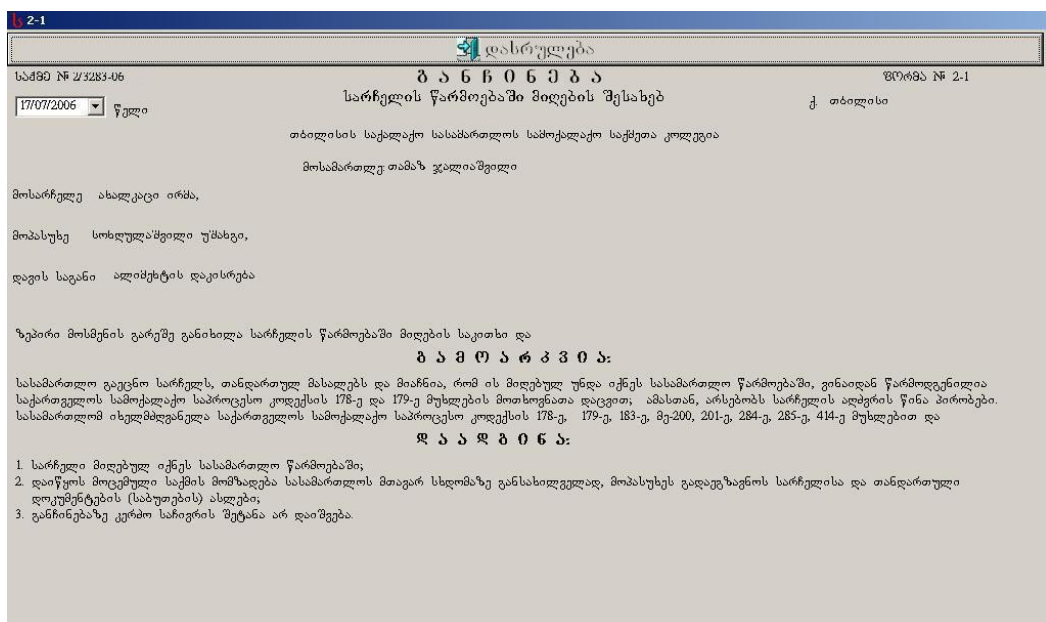


Рис. 5.2.4.2

5.2.5. После выхода из режима диалога по созданию дела запись появляется в списке всех обрабатываемых автоматизированной системой «Фемида» дел, изображенном на главной странице системы.

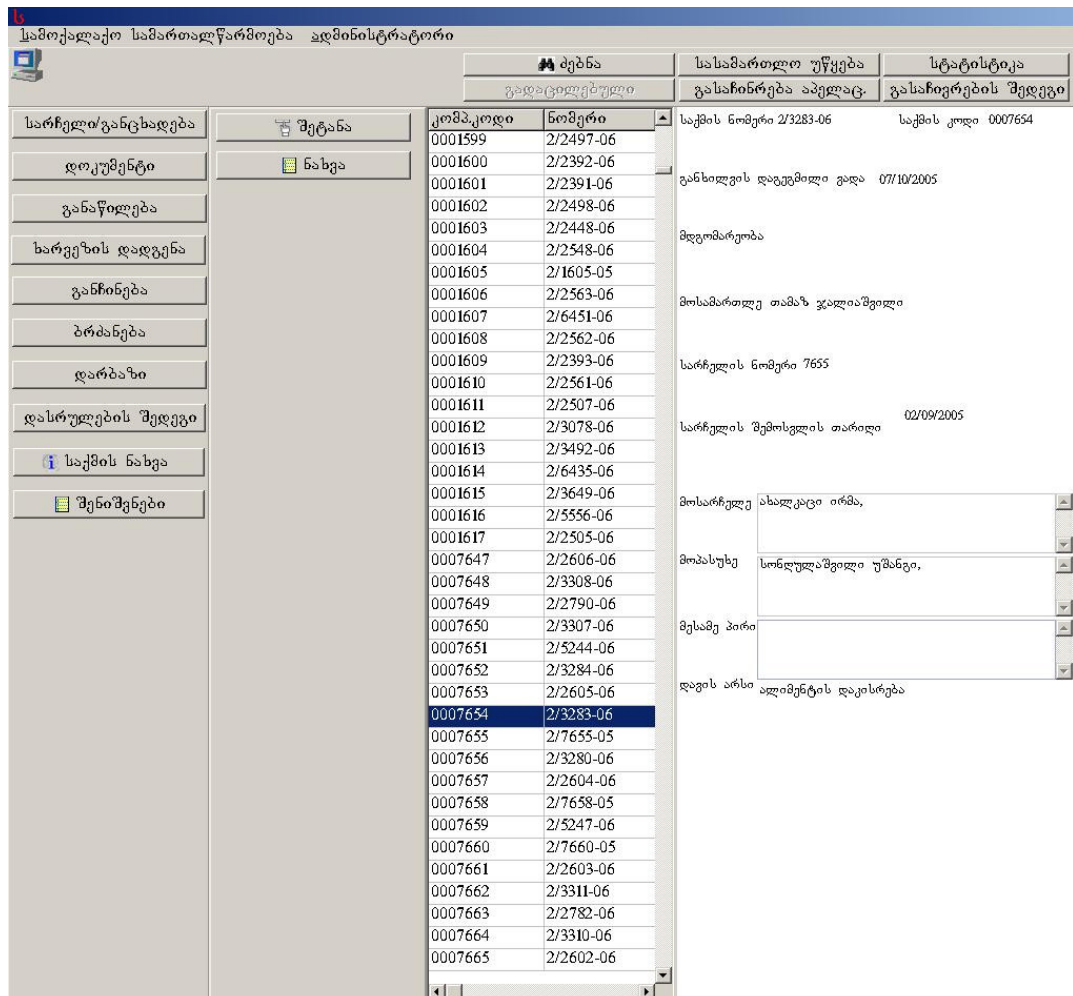


Рис 5.2.5.1

Навигация по списку всех находящихся в базе данных системы дел осуществляется стандартными средствами, общепринятыми для современных баз данных.

При установке курсора на строку, соответствующую конкретному делу, в правой колонке окна выдается подробная информация о фигурантах этого дела. После нажатия на левую клавишу мыши выводится окно выдачи информации о текущем состоянии дела (рис. 5.2.5.2)

The screenshot shows a software window titled "დასრულება" (Completed). It contains several input fields and dropdown menus for case information:

- საქმის კოდი: 0007654
- საქმის ნომერი: 2/3283-06
- გადაცემულია: 02/09/2005
- სარჩელის თარიღი: 02/09/2005
- დავის სახეობა: საოჯახო სამართლებრივი დავები
- დავის სახეობა: 35
- წარმოებაშია: 07/09/2005
- განხილვის დაგეგმილი ვადა: 07/10/2005
- დაკისრებული ბაგის ოდენობა: 45.00ლ
- წარმოებაშია: 0

Below the form is a table with the following data:

სარჩელის ნომერი	შემოსვლის თარიღი	მოსარჩელე	მოხასუნე	შესამე პირი
7655	02/09/2005	ახალკაცი ირმა	სონღულაშვილი უმანგი	

Below the table is another table with the following data:

განჩინების ნომ	საქმის ნომ	განჩინების შინაარსი	განჩინების თარიღი	ეცისიონ- უმბერ
11499	0007654	სასამართლოს გადაწყვეტილება	14/03/2006	
5226	0007654	სარჩელის წარმოებაში მიღების შესახებ	07/09/2005	

At the bottom, there is a table with the following data:

№	დოკუმენტის სახეობა	აღრუსატი	ავტორი	თარიღი

Рис 5.2.5.2

Каждому юридическому действию, совершаемому по текущему делу, соответствует формирование выходного документа – постановления суда, который помещается в базу данных системы «Фемида» и может быть произвольным образом выдан на печать или послан по локальным сетям связи в другие автоматизированные системы (в пределах установленного права доступа). Типичная форма постановления суда (рис. 5.2.5.3) включает в себя ряд реквизитов, выбранных из записи базы данных системы, соответствующей рассматриваемому делу, а также определенные реквизиты, заполняемые специально для формируемого постановления суда. Большинство из этих реквизитов имеют текстовый характер и их формулировка является прерогативой судьи, ведущего данный процесс, хотя технически она осуществляется ответственным техническим сотрудником суда, несущим полную ответственность за их адекватность решению суда.

2-37

განხილვის გამოტანა დასრულება

სამომხმადებელი № 2/3283-06

შ(რ)მპ № 2-37

09/03/2005 წელი

კ. თბილისი

სარჩელის მდგომარეობა

დაკმაყოფილდა

უარი ეთქვა

თბილისის საქალაქო სასამართლოს სამოქალაქო საქმეთა კოლეგია

მოსამართლე თამაზ ჯაღიანიშვილი

სსდომის მდივანი

მოსარჩელე ახალკაცი ირმა,

მოპასუხე სოხლულაშვილი უმბახვი,

დავის საგანი ალიმქიტის დაკისრება

ბ ა მ ტ ა რ კ ვ ი ა :

მოსარჩელე მოთხოვს

მხარეებს ეცნობათ სსდომის დრო და ადგილი, საქართველოს სსკ-ის 70-78-ე მუხლებით დადგენილი წესის შესაბამისად, მაგრამ, მიუხედავად ამისა, მოპასუხე სასამართლოს სსდომაზე არ გამოცხადდა და სასამართლოსათვის არც გამოცხადებულობის მიზეზების შესახებ უცნობებია. მოსარჩელემ იმუამდგომლა სასამართლოს წინაშე დაუსწრებელი გადაწყვეტილების გამოტანისა და სარჩელის დაკმაყოფილების შესახებ. საქართველოს სსკ-ის 230-ე მუხლის პირველი ნაწილის შესაბამისად, თუ სასამართლოს მთავარ სსდომაზე არ გამოცხადდება მოპასუხე, რომელსაც გაეგზავნა შეტყობინება სსკ-ის 70-78-ე მუხლებით დადგენილი წესით და მოსარჩელე უამდგომლობს დაუსწრებელი გადაწყვეტილების გამოტანაზე, მაშინ სარჩელში მითითებული გარემოებები დამტკიცებულად ითვლება. აღნიშნულიდან გამომდინარე, საქართველოს სსკ-ის 230-ე მუხლის პირველი ნაწილის შესაბამისად, სარჩელში მითითებული ფაქტობრივი გარემოებები სასამართლომ დამტკიცებულად ჩათვალა. ამავე მუხლის მეორე ნაწილის მოთხოვნათა მიხედვით, სამართლებრივი თვალსაზრისით, სასამართლომ შეაფასა სარჩელში მითითებული და დამტკიცებულად ჩათვლილი ფაქტობრივი გარემოებები და მიანია, რომ:

სასამართლოს მიანია, რომ აღნიშნული გარემოებები ნაწილობრივ ამართლებს (ასაბუთებს) მოსარჩელის მოთხოვნას, გინაიდან

სასამართლომ იხელმძღვანელა საქართველოს სამოქალაქო საპროცესო კოდექსის 49-ე, 53-ე, 230-ე, 234-ე, 235-ე, 236-ე, 237-ე, 369-ე მუხლებით და

ბ ა დ ა ყ ვ ი ა :

1. სასარჩელო მოთხოვნა დაკმაყოფილდეს ნაწილობრივ მოპასუხის გამოცხადებულობის გამო,

2.

3. მოსარჩელის სასარჩელო მოთხოვნის ის ნაწილი, რომლითაც

არ დაკმაყოფილდეს, გინაიდან სარჩელში მითითებული და დამტკიცებულად ჩათვლილი გარემოებანი იურიდიულად ვერ ამართლებენ მოსარჩელის მოთხოვნას ამ ნაწილში;

4.

5. განემატოს მოპასუხეს, რომ ამ გადაწყვეტილებაზე საჩივრის შემოტანა შეუძლია ამავე სასამართლოში, დაუსწრებელი გადაწყვეტილების ასლის ჩაბარებიდან 10 (ათი) დღის განმავლობაში. ამ ვადის გასვლის შემდეგ დაუსწრებელი გადაწყვეტილება შედის კანონიერ ძალაში და აღარ გასაჩივრდება;

6. გადაწყვეტილების ის ნაწილი, რომლითაც არ დაკმაყოფილდა მოსარჩელის მოთხოვნა, შეიძლება გასაჩივრდეს

საბაუდაციო სასამართლოში გამოცხადებიდან (მხარისათვის ამ გადაწყვეტილების ასლის გადაცემიდან) 14 (თოთხმეტი) დღის ვადაში.

Рис 5.2.5.3

5.3. Описание функционирования автоматизированной системы «Фемида» в режиме проверки релевантности текста формулировки судебного решения и текста статьи классификатора судебной отчетности

В 4.18. был описан пример решения задачи определения соответствия текста судебного решения статье классификатора судебной отчетности, путем установления релевантности двух текстов и сравнения маркированных представлений их семантики. Актуализация данных этого примера в режиме диалога с пользователем системы «Фемида» была проиллюстрирована выше, в 5.2. Поставленная в диссертационном исследовании научная проблема является важной, хотя не единственной частью технологического процесса функционирования автоматизированной системы «Фемида», и состоит в проверке взаимного соответствия содержания указанных текстов с помощью расширения диалога на этапе 5.2.4, когда ответственный специалист канцелярии суда выполняет в системе «Фемида» технические действия, вводящие в базу данных информацию об открытии дела и выдается окно открытия дела.

Исходной информацией для проверки релевантности является содержимое двух окон открытия дела:

– поле искового требования, в котором работник суда, отвечающий за оформление и принятие дела в производство, записывает на естественном грузинском языке составленную им формулировку содержания иска, который предъявляет истец ответчику

– поле текста статьи классификатора содержания уголовных и гражданских дел на естественном грузинском языке, однозначно определяемое задаваемым в другом поле цифровым кодом категории дела, являющимся формальным образом его содержания и определяющим юридически формализованную категорию дела, который в дальнейшем служит ключевым реквизитом как для задач поиска, так и при формировании статистической отчетности. Поле кода категории дела заполняется ответственным работником суда, отвечающим за правильность поисковой и статистической информации по всей совокупности хранящихся в базе данных дел, тем самым автоматически определяется текст статьи классификатора. Формирование поля кода производится с помощью

всплывающего меню, альтернативы которого содержат как значение кода, так и соответствующий ему текст классификатора содержания гражданских дел.

Процесс установления релевантности двух текстов, имеющий своей целью построения маркированных представлений их семантики и их последующего сравнения происходит согласно описанному в 4.18 алгоритму. Поскольку целый ряд шагов алгоритма, прежде всего разрешение неоднозначностей и проверка релевантности входящих в тексты сущностей осуществляются в диалоговом режиме оператором системы, ниже приводится типичный сценарий на примере рассмотренных в 4.18 текстов.

5.3.1. Разрешение лексическо-грамматической неоднозначности в тексте формулировки судебного решения при морфологическом анализе входящих в текст словоформ.

Для рассматриваемого предложения эта задача возникает при морфологическом разборе слова «და», которое может иметь две существенно различающиеся интерпретации, выбор между которыми производится путем положительного ответа оператора на вопросы следующего диалогового окна:

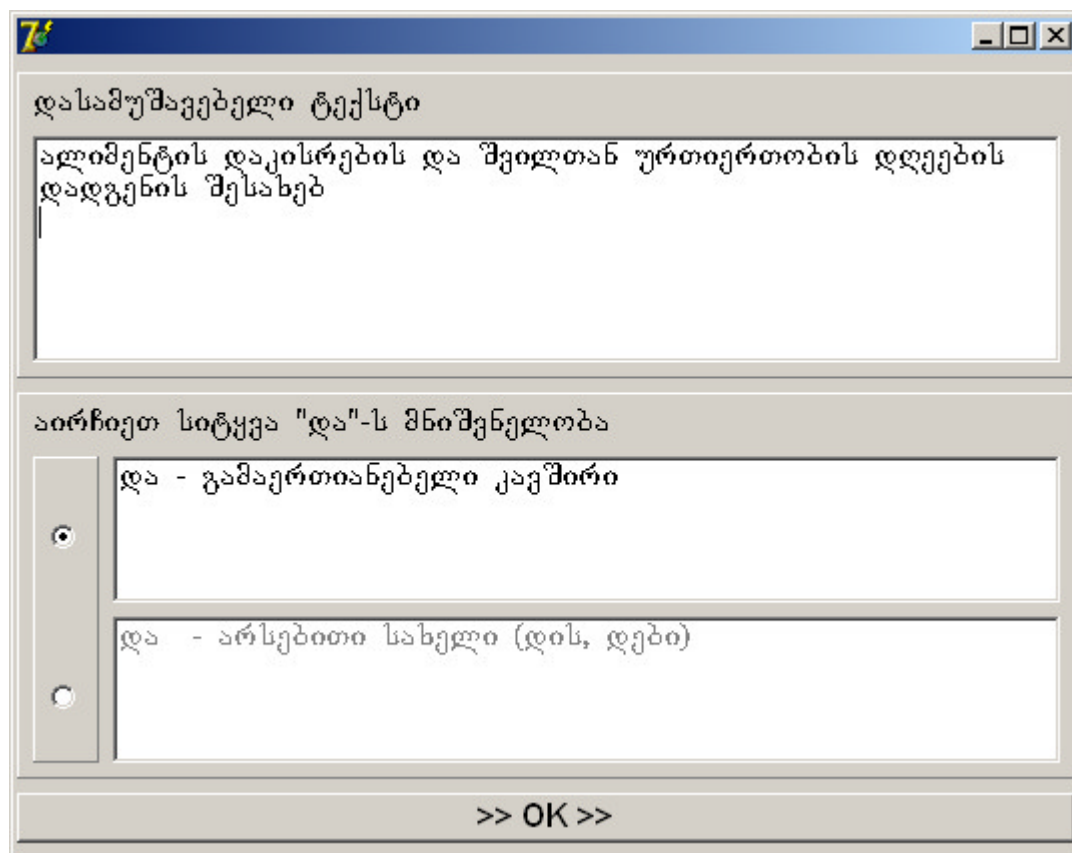
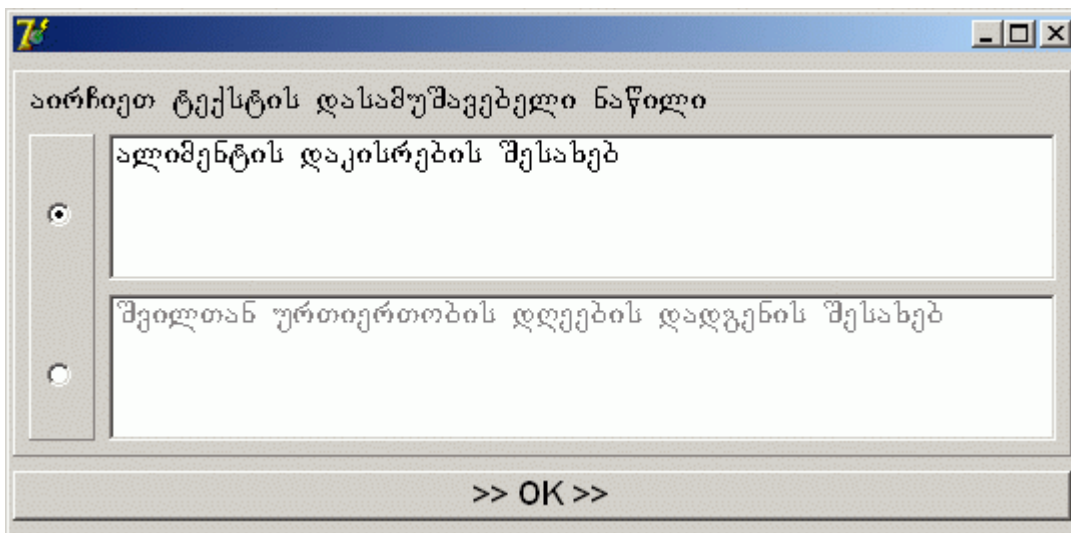


Рис 5.3.1.1

В результате выбора альтернативы №1 словоформа «და» приобретает интерпретацию соединительного союза, а исходный текст разбивается на две семантически независимые части; построение маркированных представлений



семантики будет производиться для первой из них, что решается путем выбора альтернативы №1 на вопросы следующего диалогового окна:

Рис 5.3.1.2

5.3.2. Подтверждение результатов синтаксического анализа текста по изложенным в п.4.1-4.13 алгоритмам и построения модели семантики текста путем маркирования входящих в них сущностей.

Результаты синтаксического анализа текста и построения модели его семантики изображаются в следующем диалоговом окне, тексты которого адаптированы для оператора, владеющего юридической терминологией, но не имеющего специальной подготовки в области информационных технологий

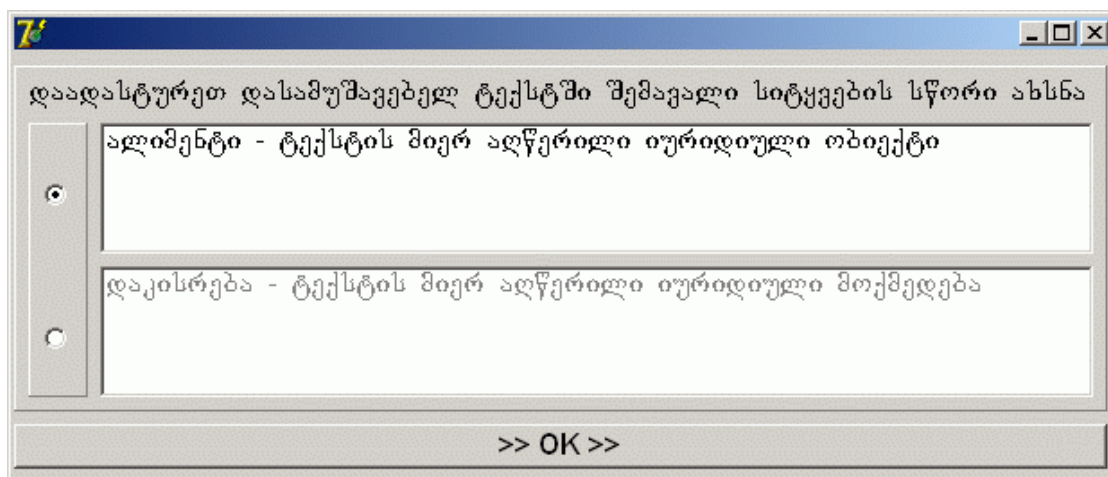


Рис 5.3.2.1

Положительный ответ означает согласие оператора на продолжение параллельного семантического анализа текстов и определения их релевантности.

5.3.3. Сопоставление моделей семантики текстов формулировки судебного решения и текста статьи классификатора судебной отчетности по числу входящих в них маркированных семантических элементов и по их приоритетности для описания юридического содержания текстов

Параллельное описание текстов по входящим в них сущностям необходимо для выявления тех словоформ, которые изображают сущности, являющиеся второстепенными с точки зрения установления релевантности и могут быть игнорированы при сопоставлении юридического содержания текстов. Ими являются, в наибольшей степени, подчиненные семантические элементы текста, являющиеся, например, субагентом (уточняющим не обязательный элемент текста – косвенное дополнение, указывающее принадлежность агента), а также адвербом (уточняющим не обязательный элемент текста – имущественную принадлежность результата действия актора). Решение об их игнорировании принимается оператором на основании рекомендации, выдаваемой системой в диалоговое окно. Процесс игнорирования продолжается до тех пор, пока число сравниваемых элементов в текстах и их семантические роли не станут одинаковыми.

На данном этапе установления релевантности текстов не имеет значение, каким именно словом обозначена та или иная входящая в них сущность, важна только их роль при их взаимном сопоставлении.

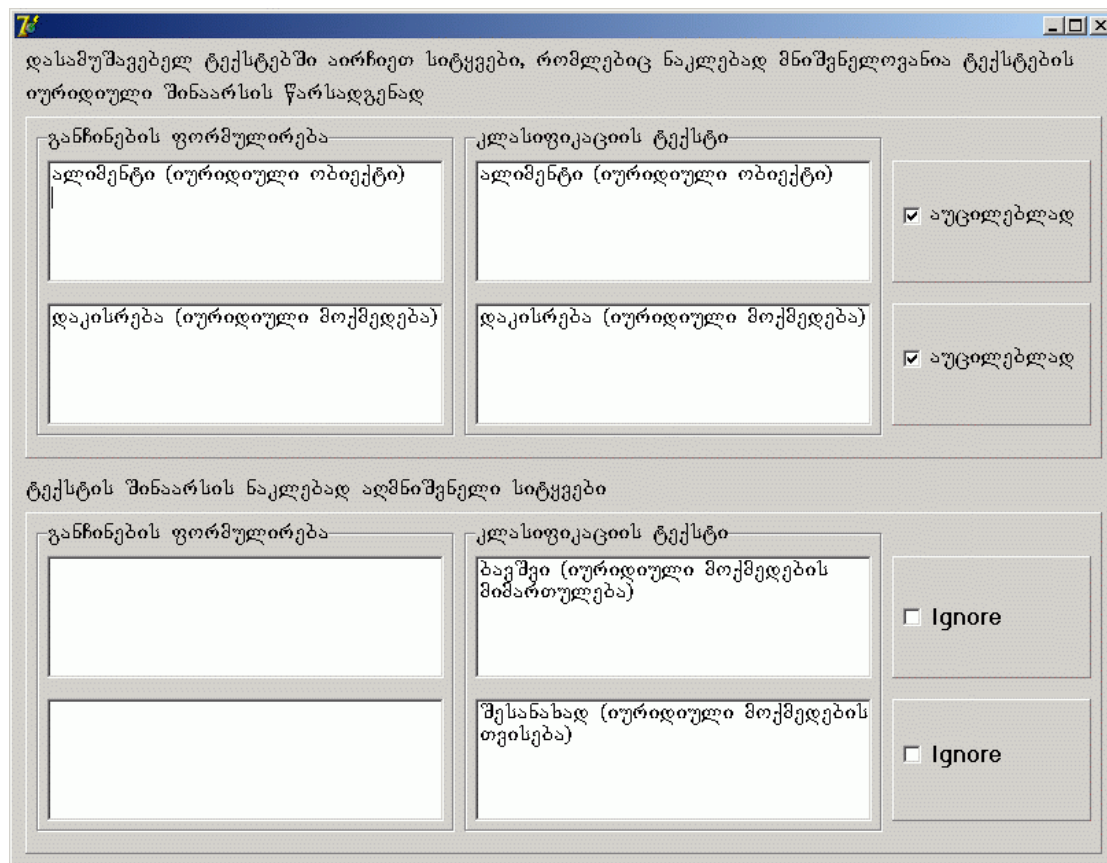


Рис 5.3.3.1

5.3.4. Сопоставление моделей семантики текстов формулировки судебного решения и текста статьи классификатора судебной отчетности по семантической близости наименования сущностей, имеющих наибольший приоритет для описания юридического содержания текстов

Сопоставление текстов по наименованиям входящим в них сущностей необходимо для выявления тех словоформ, которые изображают одни и те же сущности, разница между именами которых имеет лексикологический или чисто синонимический характер, но которые с точки зрения описания юридического содержания текстов могут быть признаны равнозначными с точки зрения тех экспертных знаний, которыми располагает квалифицированный оператор. Оператору предстоит принять решение по семантической равнозначности этих словоформ, причем, это решение может распространяться только на

наименования сущностей, имеющих наибольший приоритет для описания юридического содержания текстов.

Результаты экспертного сопоставления семантики сущностей с различными изображающими их лексическими элементами отображаются в следующем диалоговом окне, тексты которого адаптированы для оператора, владеющего юридической терминологией, но не имеющего специальной подготовки в области информационных технологий

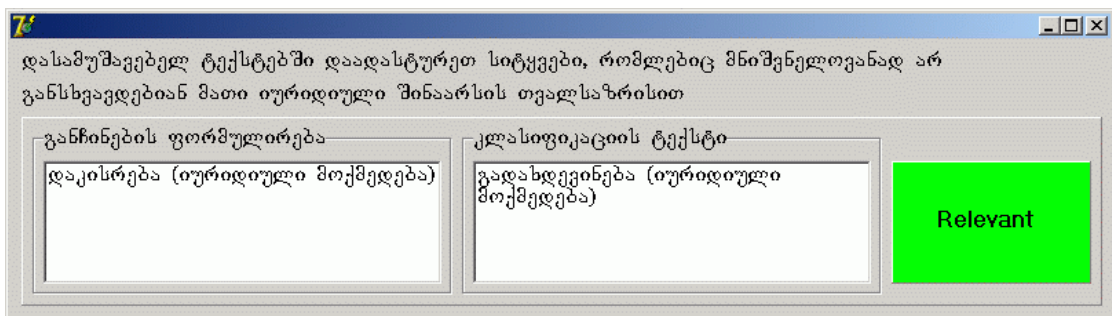


Рис 5.3.4.1

5.3.5. Принятие окончательного решения по релевантности текста формулировки судебного решения и текста статьи классификатора судебной отчетности

На основании всех проведенных выше шагов анализа релевантности текста формулировки судебного решения и текста статьи классификатора судебной отчетности, часть из которых выполнена чисто алгоритмическими средствами на основании принятых моделей, а часть нуждается в подтверждении экспертом-оператором в диалоговом режиме, окончательное решение выносится на подтверждение оператору. При его положительном решении формулировка судебного решения, записанная на начальном этапе оформления судебной документации, и избранная позже статья классификатора содержания судебного решения, считаются адекватными друг другу с точностью, вытекающей из принятых моделей и единых традиций оформления судебной документации.

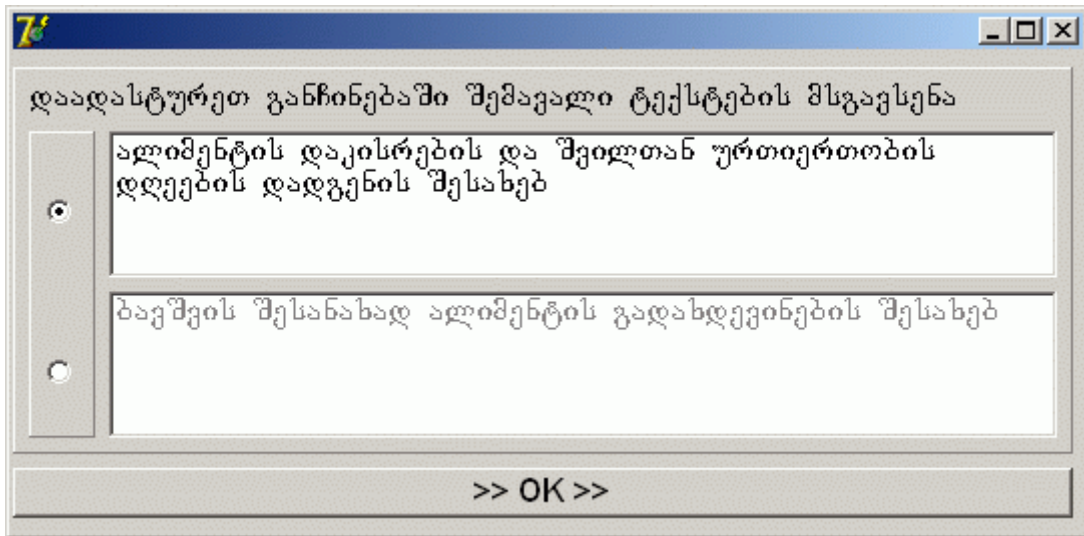


Рис 5.3.5.1

5.3.6. Отрицательное решение оператора на каждом этапе установления релевантности текстов означает невозможность решения задачи автоматизированными средствами (в рамках принятых моделей) и необходимость принятия решения чисто экспертным путем, с возложением всей ответственности на уполномоченного на это решение специалиста.

Выводы

На основании результатов проведенных теоретических и экспериментальных исследований можно сформулировать следующие выводы.

1. Настоящая диссертационная работа посвящена автоматизированным системам управления, основанным на обработке документов, содержащих информационно значимые тексты на ограниченной версии естественного языка, принятой в проблемной сфере автоматизации. В работе рассматривается одна из задач этого класса, возникающая при создании систем автоматизированного ведения судебного делопроизводства по гражданским и уголовным делам.

2. Прикладная актуальность темы диссертационной работы определяется ее ролью в повышении объективности и независимости судебных органов, достижения в работе уровня мировых стандартов судопроизводства. Этому, как показало изучение практики судебного делопроизводства, препятствуют объективные и субъективные трудности, вызванные особенностями традиционного уровня формализации создаваемых многочисленных документов, удовлетворительных с точки зрения немашинного делопроизводства, но являющиеся заметным препятствием в процессе внедрения автоматизации.

3. Конкретным объектом прикладного исследования диссертационной работы является несоответствие между текстовым образом формулировки содержания гражданского иска с ее формализованным кодом для принятого в Грузии классификатора судебной отчетности. Эта задача играет очень заметную роль как в оценке работы суда через регламентированные формы статистической отчетности, так и при решении различных поисковых задач, связанных с такой социально ответственной деятельностью, как контроль над решением судебных заседаний, оценка различных сторон функционирования судебной системы, поиск конкретных дел и т.д.

4. Предлагаемый в диссертационной работе научный подход для радикального решения этой задачи заключается в создании автоматизированных систем, обладающих расширенными «интеллектуальными» возможностями, с достаточной для практики эффективностью обеспечивающих обнаружение и исправление возможных противоречий в судебной документации.

5. Научная новизна диссертационного исследования заключается в использовании формальных моделей смысла текстов на естественном языке, в частности, естественном грузинском языке, который является формулировкой содержания конкретного гражданского иска, записанного имеющим на это юридическое действие право лицом, входящим в состав судебного органа.

6. Поставленная в диссертационной работе научная проблема известна из теории поисковых систем, ставящих своей целью поиск семантически адекватных документов. Успешность решения этой задачи затрудняется сложностью описания грамматики естественного языка, а также возникновением порождаемых при этом неоднозначностей. Наибольшие успехи в этой области достигнуты для естественного английского языка с его минимальными морфологическими средствами и однозначными правилами синтаксиса, в частности, жестко фиксированной последовательностью слов в предложении. Для таких флективных языков со сравнительно свободной последовательностью слов в предложении и сложной морфологией, как немецкий или русский, даже задача синтаксического анализа далека от своего полного решения, не говоря уже о трудностях в моделировании семантики текста для лексически достаточно развитых предметных областей. Для естественного грузинского языка, с его важнейшими чертами – эргативностью и полиперсоналией глагола, невозможно использование даже ограниченного аппарата морфо-синтаксического анализа, развитого для индо-европейских языков и требуется создание оригинальных научных подходов, сочетающихся со значительными ограничениями на язык, действующими в избранной предметной области. На сегодняшний день прецедента решения аналогичных задач для естественного грузинского языка не существует.

7. В данном диссертационном исследовании наиболее важным из принятых ограничений на естественный грузинский язык является подтверждаемая в юридической практике составления формул гражданского иска допустимая грамматическая форма принятых текстов. Они могут являться только сингамами грузинского языка, не имеющими сказуемого, т.е. не содержащими личных глагольных форм, а только - существительные, прилагательные и наречия, объединенные в специальную структуру, установленную сложившимися в этой области традициями описания содержания документов. Принятый в работе оригинальный научный подход заключается в построении

специальной модели семантики имеющих синтагматическую форму текстов судебной документации, записанных на ограниченном естественном грузинском языке.

8. Важной стороной принятого научного подхода является использование современных принципов т.н. «грамматической разметки текстов» для разрешения их синтаксической и семантической неоднозначности. Построение системы разрешения неоднозначности существенным образом опирается на грамматические особенности грузинского языка и ограничения, присущие языку конкретной предметной области формулировок содержания гражданского иска. Реализация системы разрешения неоднозначности основана на принципе генерации и отбора синтаксических, лексических и семантических гипотез путем интерактивно-диалогового взаимодействия с квалифицированным экспертом, а также путем использования экспертных знаний о текстовых корпусах судебной документации.

9. Решаемая в диссертационном исследовании проблема формулируется следующим образом.

Даны два текста на естественном грузинском языке, ограниченном предметной областью судебного делопроизводства по гражданским делам. Первый текст является формулировкой содержания конкретного гражданского иска. Вторым текстом является текст одной из статей принятого в судопроизводстве Грузии кодификатора форм статистической отчетности

Задача заключается в построении формализованной модели указанных двух текстов, позволяющей автоматически определить формальные образы их содержания и получить строго формальное определение их релевантности, после чего установление степени релевантности конкретных текстов становится вычисляемой процедурой специального вида. При положительном решении задачи заданный в классификаторе код статьи кодификатора объявляется полностью формализованным образом содержания данного гражданского иска и используется в дальнейшем во всех статистических и поисковых задачах автоматизированной системы.

10. Созданная в диссертационном исследовании формализованная модель текстов формулировки содержания гражданского иска, в свою очередь, разбивается на следующие компоненты:

– модель синтаксиса текста гражданского иска на ограниченном естественном грузинском языке;

– модель семантики текста гражданского иска на ограниченном естественном грузинском языке;

– модель установления релевантности формальных образов двух формул гражданского иска, преобразованных соответственно в две схемы причинно-квалитативных связей между выявленными в формулах сущностями.

Каждая из указанных моделей, в свою очередь, разбивается на частные модели, реализующие отдельные этапы морфологического, синтаксического и семантического анализа текстов предметной области, а также установление релевантности между их формальными образами. Эти модели реализованы в виде программного комплекса на языке программирования «Delphi».

11. Работоспособность разработанной в диссертационном исследовании модели подтверждается на решении конкретных примеров из практики судебного делопроизводства по гражданским делам.

Использованная литература

1. ლ. ჩხაიძე გ.კვარაცხელია. ქართული ენის კომპიუტერიზაციის ამოცანები. ნაბეჭდი. "ჩონფერენცე" "ჩონცეპტიონალ ანდ ჩომპუტერალ". ჩჰიკობავა ს ინსტიტუტე ოფ ინგუისტიც. თბილისი, 1996. 29-31.
2. ლ. ჩხაიძე თ.უთურგაიძე, მ.თანდაშვილი და სხვა. ქართული ენის მორფოლოგიის ცოდნის ბაზის შექმნა. " ოდელს ოფ ანგუაგე". 43-44.
3. ლ. ჩხაიძე მ.თანდაშვილი. ქართული ზმნის უღველილების ფრეიმული მოდელი. 22-27.
4. ლ. ჩხაიძე ქ.ბოლქვაძე, ი.ქერქაძე. ქართველური ენების ფუძეთა მორფო-ფონოლოგიური შესატყვისობათა მონაცემთა ბაზის შექმნა. 11-12.
5. ლ. ჩხაიძე თ.უთურგაიძე, მ.ჩაჩანიძე. ქართული ენის ორპირიან ზმნათა პარა-დიგმის აგების სრულიად ფორმალიზებულ-ლი მოდელი (მოქმედებითი გვარისათვის). ჩონფერენცე "ჩონცეპტიონალ ანდ ჩომპუტერალ ". ჩჰიკობავა ს ინსტიტუტე ოფ ინგუისტიც. თბილისი, 1997. 53-58.
6. ლ. ჩხაიძე თ.უთურგაიძე, ზ.ზურაბიშვილი, ივ.რამიშვილი. "ქართული ენის მართლწერის შემოწმების ავტომატიზებული სისტემის (სპელჩეკერის) შექმნის საკითხისათვის. " ოდელს ოფ ანგუაგე -97. 59-63.
7. ლ. ჩხაიძე ც.კვანტალიანი, თ.კვინიკაძე. ქართული ენის ზმნური ფუძეების ელექტრონული ლექსიკონი. 77-78.
8. ლ. ჩხაიძე ე.გიუნაშვილი. "ირანულ-ქართულ ენობრივ ურთიერთობათა კომპიუტერული მოდელირების საკითხისათვის." 11-12. .
9. ლ. ჩხაიძე ქ.ბოლქვაძე, ი.ქერქაძე, ე.ჭკუასელი. ქართველურ ენათა ძირების მორფონოლოგიურ შესატყვისობათა მანქანური მოდელი. თბილისი, 1998. 21-23.

10. ლ. ჩხაიძე მ.ჩაჩანიძე, ლ.ჩხაიძე, ი.ჯანაშია. "ქართული დინამიკური ზმნის უღლების მოდელის გამოცნობის ევრისტიკული სისტემა". თბილისი, 1998. 99-100.
11. ლ. ჩხაიძე გ.გოგიჩაიშიაილი, თ.უთურგაიძე. "შეზღუდული ბუნებრივი ქართული ენა, როგორც ურთიერთობის საშუალება მართვის ეტომოტიზებული სისტემებისათვის." 24-26.
12. ლ. ჩხაიძე გ.გოგლაშვილი, ც.კვანტალიანი. "ქართული ენის ზმნური ფუძეების ელექტრონული ლექსიკონი". 27-29.
13. ლ. ჩხაიძე რ.ლანდია, გ.ჩოგოვაძე. "ლიტერატურულ ნაწარმოებთა და პრესის ტექსტების გამოყენება ლექსიკონთა შევსების ევრისტიკული სისტემის შექმნისათვის". 67-69.
14. ლ. ჩხაიძე მ.მანჯგალაძე. "ქართული ენის ზმნური ფუძეების ელექტრონული ლექსიკონის გაფართოება პირიანობის მოდელით." 70-73.
15. ლ. ჩხაიძე თ.უთურგაიძე. "ქართული ზმნის ფორმალიზებული მოდელის აგებისათვის". 82-84.
16. ლ. ჩხაიძე ც.კვანტალიანი, თ.კვინიკაძე. ქართული ზმნური ფუძეების ელექტრონული ლექსიკონი". *იბტ.შემპოსიუმ ანგუაგე, ოგიც ანდ ჩომპუტაციონ. თბილისი, 1997. 19-20. . . .*
17. Л.Чхаидзе Г.Ш.Кварацхелиа. "Обзор работ Института языкознания по машинному моделированию грузинского языка и компьютерной лингвистике". стр.185-190.
18. Л.Чхаидзе Ф.Г.Утургаидзе. "Полностью формализованная модель построения парадигмы грузинского глагола". *Societas Cauca-sologica Europea. The 9th International Colloquium of Caucasian Studies. Makhachkala, 1998. 5-7.*
19. Л.Чхаидзе Г.Б.Гоголашвили, Ц.З.Кванталиани. "Электронный словарь глагольных основ грузинского языка". 11-13.
20. Л.Чхаидзе И.К.Керкадзе, К.З.Болквадзе. "Компьютерные средства в разработке историко-этимологического словаря грузинского языка". 13-15.
21. ლ. ჩხაიძე გ.გოგიჩაიშიაილი, მ.ვეკუა. "მარტივი წინადადების სინტაქსური ანალიზატორი მართვის ავტომატიზებულ სისტემებთან დიალოგისათვის შეზღუდულ ქართულ ენაზე".

- ეორგიან თეცჰნიცაღ ნივერსიტყ. თრანსაცტიონს 4 (437).
თბილისი, 2000. 70-74.
22. ლ. ჩხაიძე გ. გოგიჩაიშვილი, მ. ვეკუა. "მარტივი წინადადების
სინტაქსური ანალიზატორი შეზღუდული ქართული ენისათვის".
ქართული ენა და კომპიუტერ-ული ტექნოლო-გიები. კონფე-
რენციის მასა-ლედი.
23. ლ. ჩხაიძე შ. დავითელაშვილი. "საიფორმაციონ სისტემა ნახური
ენის ლექსიკონის ავტომატი-ზებული დაპროექტებისათვის".
კომისია, ჩჰიკობავა ს ინსტიტუტე ოფ ინგუისტიც. თბილისი,
2003. 13-15.
24. ლ. ჩხაიძე მ. კრავჩენკო, ც. კვანტალიანი, ც. ჯანჯღავა. "ქართული
ზმნის ცოდნის ბაზის აქტუ-ალიზაციისა და სრულყოფის
ავტომა-ტიზებული სისტემის შექმნის შესახებ". 24-25.
25. ლ. ჩხაიძე რ. ღანდია, თ. უთურგაიძე, ი. გიპერტი. "ძველი ქართული
ენის მანქანური ლექსიკონის ავტომატიზებული გენერაციის
სისტემა (სალიტერატურო ძეგლთა ტექსტების ბაზაზე)". 26-27.
26. ლ. ჩხაიძე თ. უთურგაიძე, მ. ჩაჩანიძე, ი. ჯანაშია, კ. მოირერი..
ქართული ზმნის პარადიგმების კომპიუტერული ტაქსონომია". 34-
36.
27. ლ. ჩხაიძე დ. ჯაში, ი. იაკობაშვილი. "ქართულენოვან ტექსტებში
არამკაფიო ძებნის მეთოდების დამუშავება და მათი გამოყენება
ქართული პრესის ელექტრო-ნული არქივის საძიებო
სისტემისათვის". 37-38.
28. ლ. ჩხაიძე ც. კვანტალიანი, მ. კრავჩენკო. ქართული ზმნის ცოდნის
ბაზის შექმნის პრინციპები". ინტელექტი, ნომ. 2 (16). თბილისი,
2003. 92-93.
29. ლ. ჩხაიძე. ქართული პარადოქსული ტექსტის მაგალითების
მიმოხილვა. ხელნაწ. რ. ჩჰიკობავა ს ინსტიტუტე ოფ
ინგუისტიც. ონდს. თბილისი, 2005.
30. გ. გოგოლაშვილი გ., კვანტალიანი ც., შენგელია დ. ქართული
ენის ზმნური ფუძეების ლექსიკონი. თბ., 1989.
31. გ. გოგოლაშვილი გ., კვანტალიანი ც., შენგელია დ. ქართული
ენის სახელზმნური ფუძეების ლექსიკონი. თბ., 1991.

32. გოგიჩაიშვილი გ., უთურგაიძე თ., ჩხაიძე ლ. შეზღუდული ბუნებრივი ქართული ენა, როგორც ურთიერთობის საშუალება მართვის ავტომატიზებული სისტემებისათვის. თბ., 1998.
33. გოგლაშვილი გ., კვანტალიანი ც., ჩხაიძე ლ. ქართული ენის ზმნური ფუძეების ელექტრონული ლექსიკონი. თბ., 1998.
34. გოგოლაშვილი გ. ქართული ზმნის უღვლილების სისტემა. თბ., 1988.
35. გოგიჩაიშვილი მ., ჩხაიძე ლ., ვეკუა მ. მარტივი წინადადების სინტაქსური ანალიზატორი მართვის ავტომატიზებულ სისტემებთან დიალოგისათვის შეზღუდულ ქართულ ენაზე. თ ს თრანსაქტიონს 4 (437), 1999.
36. დათუკიშვილი ქ. პირის ნიშანთა სისტემა ქართულში. თბ., 1997.
37. დათუკიშვილი ქ. უღლების პარადიგმათა ტიპები თანამედროვე ქართულში. თბ., 1997.
38. დამენია მ. ქართული ზმნური მორფემების სტრუქტურული მოდელები. თბ., 1982.
39. თანდაშვილი მ., ჩხაიძე ლ. ქართული ზმნის უღვლილების ფრეიმული მოდელი. თბ., 1996.
40. კვანტალიანი ც. ზმნისწინის სიტყვაწარმოებითი დანიშნულება ახალ ქართულში. სადისერტაციო ნაშრომი. 2001.
41. კვარაცხელია გ., ჩხაიძე ლ. ქართული ენის კომპიუტერიზაციის ამოცანები. თბ., 1996.
42. კრავჩენკო მ., ჩხაიძე ლ., კვანტალიანი ც., ჯანჯღავა ც. ქართული ზმნის ცოდნის ბაზის აქტუალიზაციისა და სრულყოფის ავტომატიზებული სისტემის შექმნის შესახებ. თბ., 2003.
43. მანჯგალაძე მ., ჩხაიძე ლ. ქართული ენის ზმნური ფუძეების ელექტრონული ლექსიკონის გაფართოება პირიანობის მოდელით. თბ., 1998.
44. მარგველანი ლ. ქართული სიტყვა-ფორმების კომპიუტერული სინთეზი. თბ., 1996.
45. მაჭავარიანი მ. ქართული ზმნის მოდელირების ძირითადი პრინციპები. თბ., 1996.

46. მელიქიშვილი დ. მოქმედებითი გვარის ზმნიჭ უღლების სისტემა თანამედროვე ქართულში. მაცნე, თბ. 1979.
47. მელიქიშვილი დ. ქართული ზმნის უღლების სისტემა. თბ., 2001.
48. უთურგაიძე თ. გრამატიკული კატეგორიებისა და მათი ურთიერთ-მიმართებისათვის ქართულ ზმნაში. თბ., 2002.
49. უთურგაიძე თ. ზმნურ კატეგორიათა იერარქიული მიმართებისათვის ქართულში. ქუთაისი, 1993.
50. უთურგაიძე თ., ჩაჩანიძე მ., ჩხაიძე ლ., ჯანაშია ი., მოირერი პ. ქართული ზმნის პარადიგმების კომპიუტერული ტაქსონომია. თბ., 2003.
51. უთურგაიძე თ., ჩხაიძე ლ. ქართული ზმნის ფორმალიზებული მოდელის აგებისათვის. თბ., 1998.
52. უთურგაიძე თ., ჩხაიძე ლ., ზურაბიშვილი ზ., რამიშვილი ივ. ქართული ენის მართლწერის შემოწმების ავტომატიზებული სისტემის (სპელჩეკერის) შექმნის საკითხისათვის. თბ., 1996.
53. უთურგაიძე თ., ჩხაიძე ლ., თანდაშვილი მ., ჩაჩანიძე მ., ყიფშიძე შ., კვინიკაძე თ., ჟვანია ნ. ქართული ენის მორფოლოგიის ცოდნის ბაზის შექმნა. თბ., 1996.
54. უთურგაიძე თ., ჩხაიძე ლ., ჩაჩანიძე მ. ქართული ენის ორპირიან ზმნათა პარადიგმის აგების სრულიად ფორმალიზებული მოდელი (მოქმედებითი გვარისათვის). თბ., 1996.
55. ფეიქრიშვილი ჟ. ქართული ენის მორფოლოგია. ქუთაისი, 2002.
56. ქართული ენის განმარტებითი ლექსიკონი. რვატომეული, არნ. ჩიქობავას საერთო რედაქციით. თბ., 1950-1964.
57. შანიძე აკ. თხზულებანი. შვიდტომეული, თბ., 1975-1980.
58. შანიძე აკ. ქართული გრამატიკის საფუძვლები. თბ. 1973.
59. ჩაჩანიძე მ., ჩხაიძე ლ., ჯანაშია ი. ქართული დინამიკური ზმნის უღლების მოდელის გამოცნობის ევრისტიკული სისტემა. თბ., 1998.
60. ჩხაიძე ლ., კავნტალიანი ც., კრავჩენკო მ. ქართული ზმნის ცოდნის ბაზის შექმნის პრინციპები. თ ს თრანსაქციონს, 2003.

61. ჩხაიძე ლ., კვანტალიანი ც., კვინიკაძე თ. ქართული ენის ზმნური ფუძეების ელექტრონული ლექსიკონი. თბ., 1996.
62. ჩხაიძე ლ., კვანტალიანი ც., კვინიკაძე თ. ქართული ზმნური ფუძეების ელექტრონული ლექსიკონი. " ანგუაგე, ოგიც ანდ ჩომპუტაციონ", თბ., 1997.
63. ჯორბენაძე ბ. ზმნის გვარის ფორმათა წარმოებისა და ფუნქციის საკითხები ქართულში. თბ., 1975.
64. ივლიშვილი ჩ. თქვენი ჩასვლა ოფისში, 1968.
65. Karanadze O. Development of German-Georgian dictionary for Multilingual Knowledge-Based Machine Translation & Saarbruecken, 1997.
66. Woods W. Transition Network Grammar for Natural Language Analysis. CSM, 1970.
67. Апресян Ю.Д. Лексическая семантика. Синонимические средства языка. М., 1995.
68. Бенвенист Э. Общая лингвистика. М., 1974.
69. Гоголашвили Г.Б., Кванталиани Ц.З., Чхаидзе Л.Л. Электронный словарь глагольных основ грузинского языка. Махачкала, 1998.
70. Зализняк А.А. Грамматический словарь русского языка. М., 1987.
71. Искусственный интеллект. Справочник в 3 томах. М., 1980.
72. Кварацхелия Г.Ш., Чхаидзе Л.Л. Обзор работ Института языкознания по машинному моделированию грузинского языка и компьютерной лингвистике. Тбилиси, 1998.
73. Маркус С. Теоретико-множественные модели языков. М., 1970.
74. Мельчук И.А. Опыт теории лингвистических моделей "Смысл-текст". М., 1974.
75. Минский М. Фреймы для представления знаний. М., 1975.
76. Попов Э.В., Преображенский А.В. Особенности реализации ЕЯ-систем. М., 1990.
77. Теньер Л. Основы структурного синтаксиса. М., 1988.
78. Утургаидзе Ф.Г., Чхаидзе Л.Л. Полностью формализованная модель построения парадигмы грузинского глагола. Махачкала, 1998.
79. Хомский Н. Аспекты теории синтаксиса. М., 1972.
80. Хомский Н. Язык и мышление. М., 1972.
81. Шенк Р., Абельсон В. Сценарий, планы и знание. М., 1976.

82. Ю. Д. Апресян, И. М. Богуславский, Л. Л. Иомдин, А. В. Лазурский, Н. В. Перцов, В. З. Санников, Л. Л. Цинман. «Лингвистическое обеспечение системы» ЭТАП-2. Москва, Наука, 1989 г.
83. Ю. Д. Апресян, И. М. Богуславский, Л. Л. Иомдин, А. В. Лазурский, Л. Г. Митюшин, В. З. Санников, Л. Л. Цинман. «Лингвистический процессор для сложных информационных систем». Москва, Наука, 1992г.
84. И. М. Богуславский, Л. Л. Иомдин, В. Г. Сизов, И. С. Чардин. «Использование размеченного корпуса текстов при автоматическом синтаксическом анализе». Труды международной конференции «Когнитивное моделирование в лингвистике-2003». Варна, 2003г.
85. И. М. Богуславский, Л. Л. Иомдин, А. В. Лазурский, Л. Г. Митюшин, А. С. Бердичевский, Л. Г. Крейдлин, В. Г. Сизов. «Интерактивное разрешение неоднозначности различных типов в машинном переводе». Труды международной конференции Диалог 2005. Москва, 2005г.
86. И. М. Богуславский, Л. Л. Иомдин, Л. Г. Крейдлин, Н. Е. Фрид, И. Л. Сагалова, В. Г. Сизов. «Модуль универсального сетевого языка (UNL) в составе системы ЭТАП-3». Труды международной конференции Диалог 2000. Москва, 2000г.
87. Лайонз Дж. «Введение в теоретическую лингвистику». Москва, 1977г.
88. Мальковский, М. Г. Осин, А. И. Визуализация смысла сложных для восприятия фрагментов текста. Труды международной конференции Диалог 2001. Москва, 2001г.
89. Р. О. Якобсон. «Шифтеры, глагольные категории и русский глагол». Принципы типологического анализа языков различного строя. Москва, 1972г.
90. Apresian, Ju. D., Boguslavsky, I. M., Iomdin, L. L., Lazursky, A. V., Sannikov, V. Z, Sizov, V. G., Tsinman, L. L. «ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT». MTT 2003, First International Conference on Meaning – Text Theory. Paris, École Normale Supérieure, Paris, June 16–18, 2003.
91. Blanchon, H. «An Interactive Disambiguation Module for English Natural Language Utterances». Proceedings of NLPRS 95. Seoul, Dec 4-7, 1995.

92. Blanchon, H. & Fais, L. «Asking Users About What They Mean: Two Experiments & Results. Proc. HCI'97». San Francisco, California. August 24-29, 1997.
93. Boguslavsky, Igor M., Iomdin, Leonid L., Lazursky, Alexander V., Mityushin, Leonid G., Sizov, Victor G., Kreydlin, Leonid G., Berdichevsky, Alexander S. «Interactive Resolution of Intrinsic and Translational Ambiguity in a Machine Translation System». CICLing 2005. Lecture notes in computer science. A. Gelbukh (ed.), Springer-Verlag Berlin Heidelberg 2005.
94. Boitet, C., Blanchon, H. Multilingual «Dialogue-Based MT for monolingual authors: the LIDIA project and a first mockup». Machine Translation, 9/2 1994.
95. Hutchins W. «Machine translation: past, present, future». Ellis Horwood, Chichester 1986.
96. Maruyama H., Watanabe, H., and Ogino, S. «An interactive Japanese parser for machine translation». Karlgren, H., ed. Proceedings of the 13th International Conference on Computational Linguistics, v. 2. Helsinki. 1990.
97. Tufis D., Ion, R., Ide, N. «Fine-Grained Word Sense Disambiguation Based on Parallel Corpora, Word Alignment, Word Clustering and Aligned Wordnets». Proceedings of the 20th International Conference on Computational Linguistics, Geneva, August 23–27, 2004.
98. «Universal Networking Language (UNL) Specifications». Version 2005. <http://www.unl.org/unlsys/unl/unl2005/> // Universal Networking Digital Language Foundation.
99. Лазурский А.В., Бердичевский А.С., Крейдлин Л.Г., Митюшин Л.Г., Сизов В.Г. «Интерактивное разрешение лексической и синтаксической неоднозначности в системах автоматической обработки естественного языка». Институт проблем передачи информации РАН, Москва, 2003г.
100. В.П. Захаров «Корпусная лингвистика». Санкт-Петербургский государственный университет, Санкт-Петербург, 2005г.
101. Богуславский И.М., Григорьев Н.В., Иомдин Л.Л., Крейдлин Г.Е., Фрид Н.Е., Чардин И.С. Разработка синтаксически размеченного корпуса русского языка // Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных». СПб, изд-во Санкт-Петербургского университета, 2002, с. 40-50.

102. Иомдин Л.Л., Сизов В.Г., Цинман Л.Л. (2001) Использование эмпирических весов при синтаксическом анализе. // Обработка текста и когнитивные технологии. Труды конференции «Когнитивное моделирование». Казань, Отечество, № 6, С. 64-72
103. Чардин И.С. (2001) Использование аннотированного корпуса для снятия синтаксической неоднозначности в лингвистическом процессоре ЭТАП-3. Материалы 2-ой Всероссийской конференции "Теория и практика речевых исследований" (АРСО-2001). Москва, Издательство МГУ, с.26-27.
104. Андрющенко В. М. Концепция и архитектура машинного фонда русского языка. М., 1989.
105. Богуславский И. М. и др. Аннотированный корпус русских текстов: концепция, инструменты разметки, типы информации” // Труды Международного семинара по компьютерной лингвистике и её приложениям "Диалог-2000". Протвино, 2000.
106. Empirical Methods in the new Millenium: Linguistically interpreted Corpora. Programme and Abstracts // Электронная публикация по адресу: <http://wwwling.arts.kuleuven.ac.be/sle2001/empirical-programme.htm>
107. Д. А. Поспелов Информатика: Энциклопедический словарь для начинающих - М.: Педагогика - Пресс, 1994.
108. Кузнецов О. П., Адельсон-Вельский Г. М. Дискретная математика для инженера. - 2-е изд., перераб. и доп. - М.: Энергоатомиздат, 1988.
109. Поспелов Д. А. Фантазия или наука? - М.: Наука, 1982.
110. Искусственный интеллект: Справ. В 3 кн. - М. Радио и связь, 1990.
111. Уотермен Д. Руководство по экспертным системам. - М.: Мир.
112. Реальность и прогнозы искусственного интеллекта: Сб. ст.- М.: Мир. 1987.
113. Попов Э. В. Экспертные системы: решение неформализованных задач в диалоге с ЭВМ. - М.: Наука, 1987.
114. Кузнецов В. Е. Представление в ЭВМ неформальных процедур. - М.: Наука, 1989.
115. Минто В. Дедуктивная и индуктивная логика: Пер. с англ. С.Котляревского.. – СПб: Комета, 1995.

116. Хорошевский В.Ф. Механизмы вывода в экспертных системах. - М.: МИФИ, 1988.
117. Хорошевский В.Ф. Автоматизация программирования экспертных систем. - М.: МИФИ, 1988.
118. Гаврилова Т. А., Долныкова А. А. Профессиональный подбор и подготовка специалистов в области инженерии знаний. // Новости искусственного интеллекта. - 1992, N1, с. 81-92..
119. Голенков В.В. Гулякина Н.А. Математические основы искусственного интеллекта. УМО вузов Республики Беларусь по образованию в области информатики и радиоэлектроники – Минск, 2003.
120. გოგიანიშვილი გ., ს. სერგეევი ს., ჩხაიძე ლ. ქართული ენის სინტაქსის არსის ფორმალური სახის აგების ინტელექტუალური სისტემა (სამოქალაქო სარჩელის შინაარსის ავტომატიზებული კოდირების მაგალითზე) - "შრომები-მართვის ავტომატიზებული სისტემები" №1 , თბ., 2006. 11-14.
121. Сергеев С. Структура базы знаний системы формализации содержания формулы гражданского иска. - Журнал "Интеллекти" #3 (29), Тб., 2007. 47-49
122. Sergeev S. Linguistic model for evaluating relevancy of civil suit formulae presented in native Georgian language texts. - Journal "Inteleqti" #1 (30), Тб., 2008. 58-60
123. გოგიანიშვილი გ., სერგეევი ს., ჩხაიძე ლ. ბუნებრივი ქართული ენის ტექსტების რელევანტობის შეფასების ლინგვისტური მოდელი (სამოქალაქო სარჩელის ფორმულის მაგალითზე). - VI კონფერენცია. ბუნებრივ ენათა დამუშავება არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი. თბ., 2008. 11-13