

О ВОЗМОЖНОСТИ ЛОГИЧЕСКОГО АНАЛИЗА ТЕКСТА (на примере восточных средневековых сведений о народах Восточной Европы)

ON THE POSSIBILITY OF A LOGICAL ANALYSIS OF TEXT (on the example of some oriental medieval text about peoples of Eastern Europe)

Шпирко Сергей Валерьевич,

кандидат физико-математических наук,
старший научный сотрудник факультета
управления и прикладной математики
Московского физико-технического института
e-mail:shpirko@yahoo.com

Sergey V. Shpirko

Рассматривается задача извлечения знаний из текста на естественном языке. Для ее решения предлагается подход, основанный на логике распознавания А. Д. Закревского в пространстве многозначных признаков. Для представления знаний в виде дизъюнктов рассматривается набор синтаксических трансформаций исходного текста. Выясняются границы применимости предложенного подхода. Работа алгоритма демонстрируется на примере анализа одного исторического текста.

Ключевые слова: извлечение знаний, конъюнктивная нормальная форма, матрица дизъюнктов, логический вывод, метод резолюций, логика предикатов, сколемизация, синтаксические трансформации, импликация.

In this paper we consider a problem of knowledge discovery from natural language text. We present new approach based upon A. D. Zakrevskij logic of recognition in space of multi-valued attributes. We consider some set of syntactic transformation for needed data representation as disjunctive matrix. We investigate applicability boundaries of this approach. We illustrate the algorithm by an example of some historical text.

Keywords: knowledge discovery, conjunctive normal form, disjunctive matrix, logical reasoning, resolution method, predicate logic, skolemization, syntactic transformations, implication

Согласно современным текстологическим представлениям текст средневекового произведения в процессе своего исторического бытования подвергался многочисленным изменениям¹. Такие изменения могли производиться как осознанно, так и бессознательно. В результате каждый сохранившийся список произведения является своего рода «слоеным пирогом», содержащим в себе разновременные пласты, порой взаимно противоречивые. Текстолог в целях восстановления истории текста должен находить, вскрывать подобные пласты. Аналогичные задачи уже становились объектом приложения математических

методов. Так, для задачи генеалогической классификации списков средневекового произведения применялись метод групп² и аппарат теории нечетких множеств³. В настоящей работе автор предлагает рассматривать текст на естественном языке как некоторую стройную, непротиворечивую логическую систему (в идеале). Данный подход позволяет для определенного класса текстов проводить его логический анализ, выявляя места компилятивной сшивки и логические противоречия в структуре. Далее, в тексте может содержаться и неявная информация, которую автор произведения или копиист при переписке по тем или иным причинам

опускает (например, повествуя обо всем известных фактах). Построенный на основе предложенного подхода алгоритм строит формальным образом логические цепочки и формирует логические умозаключения, что позволяет выявлять подобного рода скрытую информацию и решать задачу добывания нового знания. Концептуально развиваемый в данной статье подход опирается на две идеи, предлагаемые автором к рассмотрению.

Первая идея состоит в применении к решаемой проблеме развитого инструментария логики распознавания А. Д. Закревского⁴; вторая — в использовании трансформаций (включая синтаксические) к исходному тексту, что перекликается с идеей синонимических преобразований Е. В. Падучевой⁵. Поясним суть предложенных идей.

Отправной точкой для логики распознавания является модель задачи, построенная на основе исходных данных и знаний об окружающем мире. В качестве исходных данных рассматриваются объекты любой природы с конечным набором признаков. В случае бинарных признаков каждый такой объект представляется в виде обычного булева вектора (0 — если признак отсутствует, 1 — если присутствует). В многозначном случае булев вектор является секционированным. Каждая секция соответствует своему признаку, и длина секции равна числу его альтернативных значений.

Знания об окружающем мире представляются в виде матрицы. Данная матрица (будем обозначать ее через D) соответствует некоторой конъюнктивной нормальной форме (КНФ): $D = d_1 \wedge d_2 \wedge \dots \wedge d_n$, где $d_i = d_{i1} \vee d_{i2} \vee \dots \vee d_{imi}$, $i = 1, \dots, n$ — элементарные дизъюнкты. Каждой строке матрицы D соответствует элементарный дизъюнкт, который описывает некоторую область локализации объекта. А конъюнкция дизъюнктов определяет множество выполнимости матрицы D , или, что эквивалентно, допустимую область локализации для объектов. Таким образом, наши знания об окружающем мире можно интерпретировать как набор предписаний (соответствующих строкам матрицы D), которым должны подчиняться объекты нашей модели мира.

Вернемся теперь к рассмотрению ранее озвученной идеи синтаксических трансформаций текста. Весь текст естественно связать с матрицей D , а каждое его предложение — со своим элементарным дизъюнктом. Для подобного представления нам будет необходимо описать и обосновать соответствующие правила синтаксических трансформаций текста. Коль скоро такая матрица дизъюнктов будет получена, то дальнейший анализ должен заключаться в формальном применении к ней инструментария логики распознавания.

Наводящей для нас будет фундаментальная идея предикативности из области лингвистики.

Так, согласно классическому представлению: «Законченность предложения обусловлена наличием в его составе (главных) компонентов, выражающих предмет и приписываемый ему признак... Предикативность — это форма приписывания признака предмету в соотнесенности с моментом коммуникации»⁶.

Введем фиктивную переменную x , которую будем воспринимать как элемент некоторого универсального множества. Станем считать, что в каждом предложении текста оба его главных члена выражают признаки элемента x . Точнее говоря, сказуемое будет уточнять, доопределять признак, выраженный подлежащим. Поясним эту мысль на простом примере.

Пример 1. Рассмотрим предложение «Человек смертен». В нем «человек» является предметом (подлежащим), а «смертен» — приписываемым ему признаком (сказуемым). С учетом введенной переменной x данное предложение можно переписать в виде: «Если x является человеком, то x является смертным».

Итак, от исходного предложения мы перешли к рассмотрению условного предложения. Часть, стоящую в придаточном предложении, называют антецедентом, а в главном — консеквентом. Далее, и антецедент, и консеквент можно рассматривать как одноместные предикаты, т. е. утверждения о свойствах элемента x ⁷. А само условное предложение — как логическую импликацию вида: $\forall x A(x) \Rightarrow B(x)$. В данной формуле стоит символ $\forall x$, который означает «для всех x ». Данную формулу можно воспринимать как некоторое предписание. А именно, если какая бы то ни была переменная x обладает признаком $A(x)$, то она должна обладать и признаком $B(x)$. Связь исходного предложения с представлением в виде дизъюнкта становится явно выраженной, если учесть известный силлогизм: $A(x) \Rightarrow B(x) \approx \neg A(x) \vee B(x)$ (или x не является человеком, или x является смертным).

Возвращаясь к представлению в форме КНФ, отметим, что в логике распознавания существуют два способа получения знания. Первый способ можно назвать восхождением от частного к общему. Его удобно неформально охарактеризовать на примере одной задачи распознавания данных⁸. В ней исходные данные представляют собой выборку небольшого числа археологических объектов с бинарными признаками. На этапе обучения у объектов выборки находят несочетающиеся комбинации признаков. На основе найденных комбинаций формируют так называемые интервалы запрета, т. е. такие максимальные интервалы пространства признаков, в которые не попадет ни один объект из выборки. Далее все подобные интервалы сводятся в вышеупомянутую матрицу

дизъюнктов (в этом случае она называется матрицей запретов).

Второй способ можно назвать восхождением от общего к частному. О нем пишет А. Д. Закревский: «Знания могут поступать от эксперта — специалиста в рассматриваемой проблемной области. В описываемой формальной системе они оформляются в виде дизъюнктов и ответственность за их достоверность берет на себя эксперт... Вводимые им дизъюнкты могут также проверяться на отсутствие противоречий с базой данных»⁹. Мы будем следовать именно этому способу получения знаний. А именно, знания от экспертов будут поступать в виде готовых формул-импликаций.

Данная статья состоит из трех частей и приложений. В первой части дается неформальное введение в логику распознавания А. Д. Закревского. Особый акцент делается на изложении фундаментальных понятий, лежащих в основе логики распознавания: принципа дедукции и метода резолюции. Вторая часть статьи посвящена построению правил трансформации (включая синтаксические) исходного текста, приводящих к необходимому представлению в виде матрицы дизъюнктов. Выясняются границы применимости подобных трансформаций. В третьей части работа предложенного алгоритма демонстрируется на примере логического вывода из текстов, приводимых Б. Н. Заходером (приложение 1¹⁰, приложение 2¹¹) и В. В. Бартольдом (приложение 3¹²). Данные тексты являются переводами на русский язык сведений средневековых арабских и персидских географов и историков о народах, населявших Восточную Европу.

1. Логика распознавания в пространстве многозначных признаков

Сделаем неформальное введение в логику распознавания, следуя работе А. Д. Закревского¹³.

Напомним, что логика распознавания работает с матрицей КНФ, состоящей из элементарных дизъюнктов.

Определение 1. Элементарным дизъюнктом называется дизъюнкция конечного числа литералов, т. е. формула вида $d_{i_1} \vee d_{i_2} \vee \dots \vee d_{i_{m_i}}$.

Определение 2. Пустой дизъюнкт — единственный невыполнимый дизъюнкт, обозначаемый как L .

Для представления данных и знаний будем использовать секционированные булевы векторы. Каждый сектор отделяется от соседних точкой и соответствует своему признаку. Если объект обладает некоторым значением признака, то в соответствующем компоненте сектора ставим 1, иначе — 0. Соответствие между значениями признаков и компонентами определяем по линейному трафарету.

Пример 2. Пусть переменная x характеризуется двумя многозначными признаками x_1 и x_2 , причем у первого признака четыре альтернативных значения, а у второго — пять: $x_1 = \{a, b, c, d\}$, $x_2 = \{e, f, g, h, k\}$. Нам известно, что если переменная x принимает значения a и d , то она должна принимать и значение e : $((x_1 = a) \wedge (x_1 = d)) \Rightarrow (x_2 = e)$. Тогда с учетом приведенного силлогизма получаем элементарный дизъюнкт:

$$\begin{aligned} ((x_1 = a) \wedge (x_1 = d)) \Rightarrow (x_2 = e) &\approx \\ \neg((x_1 = a) \wedge (x_1 = d)) \vee (x_2 = e) &\approx \\ (x_1 = b) \vee (x_1 = c) \vee (x_2 = e). & \end{aligned}$$

Упорядочим значения у признаков с помощью следующего трафарета: $\frac{x_1}{abcd} \frac{x_2}{efghk}$. Тогда полученный

дизъюнкт можно представить в виде секционированного булева вектора: 0110.10000.

Рассмотрим теперь два вышеозвученных фундаментальных понятия логики распознавания: принцип дедукции и метод резолюций. Про принцип дедукции можно сказать, что он отвечает на вопрос — *Что делать?*, а про метод резолюции — *Как делать?*

1.1. Принцип дедукции

Определение 3. Матрицу дизъюнктов назовем *выполнимой*, если истинным является каждый ее дизъюнкт. В противном случае матрицу будем называть *невыполнимой*.

Данное определение можно сформулировать также следующим эквивалентным образом.

Определение 4. Матрица будет *выполнимой* тогда и только тогда, когда множество ее корней не пусто.

Здесь под корнем матрицы понимается вектор, у которого равна одна единица в каждой секции, а его покомпонентная конъюнкция с *каждой* строкой матрицы отлична от нулевого вектора. Еще в таком случае говорят, что у матрицы существует *столбцовый минор*, покрывающий все ее строки.

Центральная задача дедуктивного вывода формулируется следующим образом. Пусть заданы матрица дизъюнктов D и некоторый дизъюнкт d . Необходимо проверить, является ли он логическим следствием матрицы дизъюнктов: $D \Rightarrow d$. Другими словами, обращается ли элементарная дизъюнкция d в единицу на множестве выполнимости матрицы D .

Обозначим через $\neg d$ вектор, полученный из вектора d путем его покомпонентного отрицания, а через $D \wedge \neg d$ — матрицу, полученную из матрицы D путем покомпонентной конъюнкции каждой ее строки с вектором $\neg d$.

Принцип дедукции: Дизъюнкт d логически следует из матрицы дизъюнктов D тогда и только тогда, когда невыполнима матрица $D \wedge \neg d$.

Доказательство. С учетом силлогизма имеем цепочку эквивалентных формул:

$$\neg(D \Rightarrow d) \approx \neg(\neg D \vee d) \approx \neg(\neg D) \wedge \neg d \approx D \wedge \neg d.$$

Пример 3. Рассмотрим предложение *Язык болгар и русов не похож на язык буртасов*.

Сначала представим его в виде секционированного вектора-дизъюнкта. Применяя синтаксические трансформации (см. часть 2), приходим к конструкции: *Если x является представителем народа болгар или x является представителем народа русов, то x имеет язык, не похожий на язык буртасов*. С учетом введенных в третьей части статьи признаков данная конструкция представима в виде импликации:

$$((x_1 = "2") \vee (x_1 = "3")) \Rightarrow (x_8 = "2").$$

Отсюда с учетом силлогизма получаем дизъюнкт

$$d = \neg((x_1 = "2") \vee (x_1 = "3")) \vee (x_8 = "2") \approx (x_1 = "1") \vee (x_1 = "4") \vee (x_8 = "2").$$

Докажем теперь, что данный дизъюнкт является логическим следствием матрицы D . Рассмотрим подматрицу D' матрицы D , состоящую из строк и секций, релевантных X -му предложению Приложения 2 (строки с 20-й по 27-ю, секции 1-я, 5-я, 6-я, 7-я, 8-я):

$$D' = \begin{pmatrix} 1011 & 00 & 00 & 10 & 00 \\ 1110 & 10 & 00 & 00 & 00 \\ 0111 & 01 & 00 & 00 & 00 \\ 1011 & 00 & 00 & 00 & 01 \\ 1101 & 00 & 00 & 01 & 00 \\ 1110 & 00 & 01 & 00 & 00 \\ 1101 & 00 & 00 & 00 & 01 \\ 0111 & 00 & 01 & 00 & 00 \end{pmatrix}$$

Из Определения 4 следует, что для невыполнимости матрицы $D \wedge \neg d$ достаточно доказать невыполнимость $D' \wedge \neg d$ (поскольку все остальные секции рассматриваемых строк являются нулевыми и поэтому незначимыми).

С учетом трафарета полученный дизъюнкт d представим в виде секционированного булева вектора: $d = (1001.00.00.00.01)$ (здесь также указаны лишь релевантные секции — 1-я, 5-я, 6-я, 7-я, 8-я). Тогда $\neg d = (0110.11.11.11.10)$ и

$$D' \wedge \neg d = \begin{pmatrix} 0010 & 00 & 00 & 10 & 00 \\ 0110 & 10 & 00 & 00 & 00 \\ 0110 & 01 & 00 & 00 & 00 \\ 0010 & 00 & 00 & 00 & 00 \\ 0100 & 00 & 00 & 01 & 00 \\ 0110 & 00 & 01 & 00 & 00 \\ 0100 & 00 & 00 & 00 & 00 \\ 0110 & 00 & 01 & 00 & 00 \end{pmatrix}$$

Нетрудно видеть, что у подматрицы $D' \wedge \neg d'$ любой столбцовый минор не покрывает одновременно 4-ю и 7-ю строки (из-за 1-й секции). Следовательно, данная подматрица невыполнима, а вектор d является логическим следствием исходной матрицы D .

1.2. Правила редукции матрицы

В реальных задачах матрицы дизъюнктов могут быть очень большого размера. Поэтому встает вопрос их упрощения (редукции), при котором они не теряли бы своего свойства выполнимости/невыполнимости. Ниже приводятся без доказательства шесть таких правил.

Пусть u и v — некоторые строки матрицы дизъюнктов D , а p и q — некоторые ее столбцы. Пусть a и b — некоторые векторы. Будем обозначать $a \geq b$, если данное неравенство выполняется покомпонентно.

Правила редукции:

1. Если $u \geq v$, то строка u удаляется;
2. Если строка u содержит секцию без нулей, то она удаляется;
3. Если столбец p — пустой (состоит из одних нулей), то он удаляется;
4. Если в строке содержатся единицы лишь в одной секции, то удаляются все столбцы этой секции, содержащие нули в данной строке;
5. Если $p \geq q$ и столбцы p и q принадлежат одной секции, то столбец q удаляется;
6. Обозначим через d_i^{jk} значение k -го компонента j -й секции, стоящего в i -й строке матрицы дизъюнктов, и пусть $d_i^{jk} = 1$. Далее, обозначим через d вектор, получающийся из i -й строки заменой j -й секции на такую, в которой $d_i^{jk} = 0$, а остальные компоненты имеют значение «единица».

Тогда единичное значение d_i^{jk} можно заменить на нулевое.

Вообще говоря, необходимо добиваться максимального упрощения матрицы дизъюнктов, при котором все ее строки содержат минимальное число единиц, при этом ни одну единицу нельзя заме-

нить на нуль без потери эквивалентности преобразований (см. правила редукции).

1.3. Метод резолюции

Следующим фундаментальным понятием логики распознавания является метод резолюции. Он позволяет строить цепочки логического вывода и извлекать тем самым новое знание.

Пусть даны два секционированных вектора дизъюнкта u и v .

Определение 5. Будем называть резольвентой векторов u и v по i -й переменной и обозначать через $u < i > v$ вектор, получающийся покомпонентной конъюнкцией i -х секций векторов u и v и покомпонентной дизъюнкцией остальных секций данных векторов.

Данная операция резолюции является обобщением операции обобщенного склеивания в теории булевых функций¹⁴.

Однако не все получающиеся резольвенты достойны дальнейшего рассмотрения. Они могут оказаться, например, лишь расширением одиночных дизъюнктов, присутствующих в матрице, и быть отброшены (см. правила редукции). Нас интересуют лишь те векторы, резольвента которых является действительно новым знанием.

Определение 6. Будем называть векторы дизъюнкты u и v смежными по i -й переменной, если их покомпонентная дизъюнкция i -й секции отличается от каждой из этих секций, а в любой из остальных секций существует некоторая компонента с нулем в обоих векторах.

Пример 4. Рассмотрим 16-ю и 29-ю строки матрицы дизъюнктов $D: d_{16} = (1011.00.10.10.10.10.10)$ и $d_{29} = (0000.10.00.00.00.01.00)$. Здесь приведены только значимые (ненулевые) секции обоих векторов: 1-я, 23-я, 25-я, 29-я, 30-я, 33-я, 36-я. Нетрудно видеть, что данные векторы смежны по 33-й переменной. Тогда их резольвента $d_{16} < 33 > d_{29} = (1011.10.10.10.10.00.10)$.

Заметим, что 16-я строка матрицы дизъюнктов соответствует IV и VI предложениям, перечисляющим источник богатства болгар. Одним из таких источников являются куницы (33-я переменная). 29-я строка соответствует вспомогательному утверждению о том, что понятие «куница» включается в более общее понятие «меха» (в промышленном смысле). Полученную резольвенту можно перевести на естественный язык следующим образом: «У болгар — вьючные животные (или лошади), кольчуги, оружие и меха». Таким образом, в данном примере в результате операции резольвирования был произведен переход от частного к обобщенному понятию (от мехов куниц к мехам вообще).

Возвращаясь к проблеме логического вывода, отметим, что в процессе резольвирования исход-

ная матрица дизъюнктов расширяется включением в нее новых дизъюнктов, являющихся следствием некоторых пар дизъюнктов, уже вошедших в матрицу. В связи с этим возможны три варианта¹⁵:

- получен пустой дизъюнкт \perp . В этом случае матрица дизъюнктов невыполнима (см. принцип дедукции). Следовательно, исходная модель мира неверна, ее необходимо корректировать;
- полученная матрица не содержит ни одной пары дизъюнктов, допускающих резольвирование;
- процесс новых следствий продолжается бесконечно долго или заикливается.

В заключение данной части статьи рассмотрим одну важнейшую задачу, решаемую в рамках логики распознавания.

1.4. Задача распознавания

Рассмотрим некоторый объект и матрицу дизъюнктов, отражающую закономерности того мира, которому принадлежит изучаемый объект. Пусть нам указан некоторый признак u объекта, значение которого не известно. Будем называть его целевым.

Задачу распознавания можно сформулировать следующим образом: как по известной матрице дизъюнктов и *частичной информации* об объекте вычислить его целевой признак. Или, добавим, максимально сузить область возможной локализации объекта.

Поясним, что здесь понимается под частичной информацией. Будем считать, что частичная информация об объекте задается совокупностью элементарных запретов $x_i \neq v_j$, где x_i — это признак объекта, а v_j — одно из его возможных значений. Таким образом, этот элементарный запрет указывает на то, что значение признака x_i данного объекта отлично от v_j . В крайнем случае, когда элементарные запреты запрещают все значения признака, кроме одного, то значение данного признака будет задано однозначно.

Сопоставим частичной информации об объекте вектор-конъюнкт r , в котором все элементарные запреты отображены значением нуль в соответствующих компонентах. Таким образом, вектор-конъюнкт r задает исходную интервальную локализацию нашего объекта. Зная матрицу дизъюнктов, эту первоначальную локализацию можно уточнить.

Утверждение 1. Область возможного существования объекта, задаваемого вектором-конъюнктом r , образуется множеством корней матрицы $D^* = D \wedge r$, где D — матрица дизъюнктов.

Процесс уточнения локализации объекта можно продолжить. Для этого данный процесс удобно

расчленив на поиск отдельных элементарных запретов $x_i \neq v_j$, логически следующих из матрицы D^* .

Поставим в соответствие элементарному запрету $x_i \neq v_j$ секционированный булев вектор $s(i, v_j)$, у которого все компоненты i -й секции, кроме соответствующей значению v_j , равны единице, а остальные секции состоят из одних нулей.

Если учитывать принцип дедукции, то справедливо следующее:

Утверждение 2. Запрет $x_i \neq v_j$ следует из матрицы D^* тогда и только тогда, когда невыполнима матрица дизъюнктов $D^* \wedge \neg s(i, v_j)$.

Пример 5. В качестве иллюстрации проверим, можно ли на основе матрицы дизъюнктов уточнить религиозную принадлежность народа буртасов.

Итак, частичная информация задается элементарными запретами по признаку

$$x_1 : x_1 \neq "2", x_1 \neq "3", x_1 \neq "4".$$

Соответствующий вектор-конъюнкт r определяет однозначно значение данного признака: $r^1 = 1000$, все остальные секции $r^j, j = 2, \dots, 34$ состоят из одних единиц. Образует подматрицу D' , состоящую из 7-й, 28-й строк матрицы

$$D : D' = \begin{pmatrix} 0111 & 00 & 10 \\ 0000 & 01 & 01 \end{pmatrix}.$$

Здесь приведены только 1-я, 2-я и 3-я секции исходной матрицы. Остальные секции нулевые и поэтому незначимые.

Отсюда, с учетом Утверждения 1, следует, что область локализации объекта, задаваемого вектором r , задается множеством корней матрицы

$$D^* = D' \wedge r = \begin{pmatrix} 0000 & 00 & 10 \\ 0000 & 01 & 01 \end{pmatrix}.$$

В качестве целевого возьмем признак x_2 и рассмотрим элементарный запрет $x_2 \neq "1"$, соответствующий утверждению о том, что буртасы не являются мусульманами. Данному элементарному запрету соответствует секционированный булев вектор $s(2,1) = (0000.01.00)$. Тогда

$$D^* \wedge \neg s(2,1) = \begin{pmatrix} 0000 & 00 & 10 \\ 0000 & 00 & 01 \end{pmatrix}.$$

Нетрудно видеть, что полученная матрица не имеет корней (из-за 3-й секции). Следовательно, и исходная матрица также является невыполнимой. Таким образом, наше исходное предположение оказалось верным, а буртасы действительно не являются мусульманами.

В связи с рассмотренным примером сделаем небольшое историографическое отступление. Вопрос о мусульманстве буртасов не является праздным, а дискуссия по нему насчитывает не одно

столетие (причины ее возникновения подробно излагаются в работе Д. А. Хвольсона¹⁶). Например, в известном средневековом сочинении «Худуд ал-'алам» приводится противоположное сообщение о религиозной принадлежности буртасов¹⁷. Современному состоянию данной дискуссии посвящена работа Т. М. Калининой¹⁸. Подчеркнем, что мы ни в коем случае не пытаемся вступать в дискуссию со специалистами по данному вопросу. Цель настоящей статьи состоит в том, чтобы показать, как на основе имеющейся исходной информации можно строить модель мира в виде матрицы дизъюнктов и делать на ее основе правдоподобные выводы.

II. Правила трансформации исходного текста

Целью данной части статьи является описание и обоснование правил трансформации исходных предложений, приводящих к формуле $\forall x A(x) \Rightarrow B(x)$, где $A(x), B(x)$ — одноместные предикаты, а знак \forall — квантор общности. Формулы такого вида называют сколемовскими⁶. Как было показано ранее, данные формулы представимы в виде элементарных дизъюнктов.

Тип предложений, допускающих подобное представление, можно условно охарактеризовать как описательно-определятельный. В таких предложениях подлежащее, описывающее один признак универсальной переменной x , доопределяется другим ее признаком (выраженным сказуемым). Как уже отмечалось, наиболее четко подобное представление проявляется в условных предложениях. Необходимо отметить, что подобное представление должно носить «всеобщий» характер. Поэтому возможность сколемизации является критично важным требованием.

Также подчеркнем, что рассматриваться будут исключительно синтаксические трансформации. Под ними будем понимать правила толкования, позволяющие из исходного текста выделять синтаксическую конструкцию и заменять ее на синонимическую, но с более явной связью между формой и смыслом²⁰. Явлениями, подлежащими толкованию, будут, в частности, следующие:

1. Тип синтаксической связи (сочинительная, определятельная);
2. Грамматические категории, смысл которых связан с синтаксическим контекстом (например, отрицание, залог, падеж).

Параллельно с толкованием будет происходить и формирование списка признаков. Необходимым требованием для каждого признака является наличие альтернативных, т. е. взаимоисключающих и в сумме описывающих весь признак, значений.

II.1. Сочинительный тип связи (без отрицания)

Сочинительным типом связи соединяются две равноправные (однородные) части предложения. Этими частями могут быть как главные члены предложения (сказуемое и подлежащее), так и второстепенные (дополнение, определение и др.).

Сочинительная связь будет использована для формирования полного списка признаков модели, а также для разбиения конструкций текста.

Пример 6. Рассмотрим 23-е предложение из приложения 1: «У них деревянные дома, в них они живут зимою, и шатры, в них они обитают летом». Это сложная конструкция содержит два однородных члена в роли сказуемых: «дома», «шатры». В то же время они не являются альтернативными. Человек может жить исключительно в домах или шатрах, а может жить в обоих этих типах жилищ (посезонно, как в данном примере). Поэтому исходный признак необходимо разбивать на два: «иметь жилищем дом» и «иметь жилищем шатер». Причем оба признака являются бинарными, характеризующиеся двумя значениями: 1 — если соответствующий признак есть, 0 — в противном случае.

Есть и противоположные примеры, когда исходный признак не имеет смысла разбивать.

Пример 7. Рассмотрим 34-е предложение Приложения 1: «Буртасы делятся на два разряда: одни предают мертвых сожжению, другие хоронят их». В данной конструкции присутствуют два сказуемых, определяющих признак «способ погребения»: «сожжение» и «захоронение». Понятно, что над человеком можно лишь единожды произвести ритуал погребения. И этот ритуал будет характеризоваться альтернативными признаками: или сожжение, или захоронение, или иной способ погребения. Таким образом, исходный признак дробить не нужно.

В большинстве случаев сочинительная связь — это перечень возможных признаков, и разбиения исходной конструкции на несколько не требуется. В то же время, если какие-либо признаки должны обязательно сочетаться, то необходимо разбиение исходной конструкции на их соответствующее количество.

Пример 8. Возвращаясь к Примеру 6, заметим, что исходя из контекста человек должен проживать как в деревянных домах, так и в шатрах. Поэтому исходное предложение необходимо разбить на две конструкции: «У них деревянные дома, в них они живут зимою», «У них шатры, в них они обитают летом».

II.2. Сочинительный тип связи (с отрицанием)

Будем считать отрицательными предложения, содержащие лексему *нет* или частицу *не*. Если так-

же присутствует сочинительная связь, то исходное предложение должно, как и в Примере 8, разбиваться на конструкции, число которых равно количеству однородных членов.

Пример 9. Рассмотрим предложение VII Приложения 2: «У болгар нет золотой и серебряной монеты». Данное предложение необходимо разбить на две конструкции: «У болгар нет золотой монеты» и «У болгар нет серебряной монеты».

II.3. Определительный тип связи

К определительным будем относить конструкции с прилагательными, обособленными определениями (в частности, причастными оборотами) и определительными придаточными.

Обособленные определения и определительные придаточные будем исключать из дальнейшего рассмотрения.

Пример 10. Рассмотрим 33-е предложение Приложения 1: «Наибольшая часть достояния буртасов — куницы, из шкур которых делают шубы». Здесь присутствует придаточное предложение «из шкур которых делают шубы».

Конструкции с прилагательным также удаляются (если только это прилагательное не разграничивает отдельные значения признака).

Пример 11. В рассматриваемом 33-м предложении присутствует словосочетание «наибольшая часть достояния буртасов» с прилагательным «наибольшая». В результате приходим к конструкции «Часть достояния буртасов — куницы».

В то же время встречаются конструкции, в которых прилагательные нельзя удалять из дальнейшего анализа.

Пример 12. Рассмотрим предложение X Приложения 2: «Язык болгар походит на язык хазар, у буртасов другой язык, язык русов не походит на язык хазар и буртасов». В нем роль одного из сказуемых играет прилагательное «другой».

Пример 13. Возвращаясь к Примеру 9, обратим внимание, что оба прилагательных «золотая» и «серебряная» в словосочетаниях «золотая монета» и «серебряная монета» определяют два бинарных признака (см. II.1). Поэтому они не могут быть удалены.

II.4. Исходная залоговая конструкция

Будем считать, что залоговая конструкция является исходной, если глагольная лексема имеет исходное синтаксическое употребление. Сказуемое выражено глаголом, подлежащее выражает субъект действия, а прямое дополнение — объект действия.

Пример 14. Вернемся к Примеру 12. Нетрудно видеть, что одна из его конструкций «Язык болгар походит на язык хазар» имеет залогово-исходную форму. Сказуемое выражено глаголом «походит»,

а подлежащее — лексемой «язык». Для конструкции «У буртасов другой язык» это не так. В ней роль сказуемого играет указательное местоимение *другой*.

В Примере 10 конструкция «Наибольшая часть достояния буртасов — куницы» также не имеет залогово-исходной формы. В ней роль сказуемого играет существительное в именительном падеже («куницы»), а глагол-связка отсутствует.

Все подобные конструкции должны быть преобразованы в залогово-исходную форму. При этом подлежащее также должно находиться в немаркированной форме (в именительном падеже). В результате получаем: «Буртасы имеют другой язык», «Наибольшая часть достояния буртасов есть куницы».

II.5. Контекстная подстановка

Небольшая, но важная операция. Она состоит в подстановке имен в те места конструкции, где они неявно подразумеваются.

Пример 15. В конструкциях, рассматриваемых в Примере 8, необходимо заменить местоимение «они» на подразумеваемое имя «буртасы». В результате получаем: «У буртасов деревянные дома, в них буртасы живут зимою», «У буртасов шатры, в них буртасы обитают летом».

II.6. Эквивалентность конструкций

В конструкции могут встретиться глагольные лексемы, которые задают некое отношение эквивалентности между двумя ее частями. К таким лексемам относятся, например: «быть равным», «быть похожим», «быть подобным», «не быть равным», «не быть похожим», «не быть подобным», «быть другим» и др. В этом случае исходную конструкцию необходимо разбивать на две с «зеркальной» заменой ее частей.

Пример 16. Вернемся к Примеру 12. Приведенная в нем конструкция «Язык болгар походит на язык хазар, у буртасов другой язык, язык русов не походит на язык хазар и буртасов» содержит три подобные лексемы. Причем в рассматриваемой конструкции присутствует еще и сочинительная связь.

В результате разбиения получаем восемь отдельных конструкций: «Язык болгар походит на язык хазар», «Язык хазар походит на язык болгар», «У буртасов язык не походит на язык болгар», «У болгар язык не походит на язык буртасов», «Язык русов не походит на язык хазар», «Язык русов не походит на язык буртасов», «Язык хазар не походит на язык русов», «Язык буртасов не походит на язык русов».

II.7. Представление в форме импликации

Переходим к заключительной стадии синтаксических трансформаций. В результате предыдущих стадий мы имеем конструкцию вида NV. Здесь

через символ N обозначено подлежащее, а через V — сказуемое (или группа сказуемого в случае зависимых слов или однородных членов). Сама конструкция находится в залогово-исходной форме. Подлежащее и сказуемое связаны друг с другом предикативной связью, т. е. подлежащее выражает предмет, а сказуемое — приписываемый ему признак.

Введем фиктивную переменную x — элемент универсального множества. Как уже отмечалось, в этом случае оба главных члена предложения можно рассматривать как признаки (свойства) переменной x . При этом подлежащее является признаком, а сказуемое — его уточнением, признаком признака. Свяжем с ними два соответствующих предиката: $N \leftrightarrow A(x)$ и $V \leftrightarrow B(x)$. Если исходная конструкция NV носит «всеобщий» характер, то можно перейти к сколемовской формуле в виде импликации: $\forall x A(x) \Rightarrow B(x)$.

Пример 17. Рассмотрим одну из конструкций Примера 16: «Язык болгар походит на язык хазар». Данная конструкция будет эквивалентна следующей: «Если x является представителем народа болгар, то x имеет язык, похожий на язык хазар». С учетом введенных в третьей части статьи признаков последняя конструкция запишется в виде: $(x_1 = "2") \Rightarrow (x_7 = "1")$ (ср. 20-ю строку матрицы дизъюнктов).

Важно подчеркнуть, что далеко не всегда конструкция NV носит «всеобщий» характер. В этом случае сколемизация неприменима. Следовательно, невозможно построить и элементарный дизъюнкт.

Пример 18. Рассмотрим предложение VIII Приложения 2: «...в обоих городах деревянные строения». Смысл данного предложения заключается в том, что отдельные представители народа болгар проживают в деревянных строениях. На языке логики предикатов данной конструкции соответствует формула с квантором существования: $\exists u: A(u) \Rightarrow B(u)$, где предикаты $A(u)$ и $B(u)$ соответствуют признакам x_1 и x_{12} соответственно. К сожалению, преобразовать данную формулу в сколемовский дизъюнкт не представляется возможным. И данное предложение находится вне построенной в третьей части матрицы дизъюнктов.

С учетом приведенных в данной части статьи правил трансформации понятно, насколько важно четко представлять себе синтаксическую структуру предложения. Заключительную часть автор хотел бы посвятить неформальному введению в один весьма полезный язык представления синтаксических структур — грамматику зависимостей²¹.

В основе данной грамматики лежит представление о синтаксической структуре предложения как иерархии связей. Типы связей могут быть сочинительные, подчинительные, предикативные.

Часть из них мы уже разбирали: сочинительная, определительная (как один из вариантов подчинительной) связи. Во главе этой иерархии стоит сказуемое, связанное с подлежащим предикативной связью.

Грамматику зависимостей удобно изображать графически в виде дерева зависимостей. Далее автор в своем изложении сведений по деревьям зависимостей следует согласно работе И. П. Севбо²². Дерево зависимостей будем представлять в виде неориентированного графа. Узлы соответствуют словам предложения, а дуги — связям между ними. Изолированных узлов не бывает. Каждое слово входит в словосочетание, в котором оно является или главным или зависимым. В случае, когда связь в предложении никак не маркирована (например, сочинительная), вводим для нее дополнительный узел.

Будем обозначать узлы графа следующими метками: ● — для полнозначных слов, ○ — для предлогов, ⊕ — для сочинительных союзов. Дуги будем обозначать следующими метками: — — подчинения слов, - - - - — однородность слов, = = = = — подчинение предложений. В сложноподчиненных предложениях с расчлененными составными союзами подчинительная связь устанавливается между частями этого союза. При этом часть союза, стоящую в главном предложении, будем обозначать ⊕, а в придаточном — ⊙.

Пример 19. Отобразим с помощью деревьев зависимостей процесс синтаксических преобразований одной из рассмотренных ранее конструкций. Отображение синтаксических преобразований остальных конструкций проводится аналогичным образом (рис. 1).

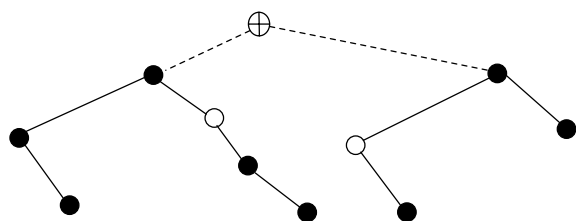


Рис. 1. Язык болгар происходит на язык хазар, у буртасов другой язык

Вершиной графа является дополнительный узел, отражающий сочинительную связь; однородные члены (сказуемые *походит* и *другой*) связаны штриховой линией. После преобразования в исходно-залоговую форму имеем граф, изображенный на рисунке 2.

Далее разбиваем граф на два графа (рис. 3–4).

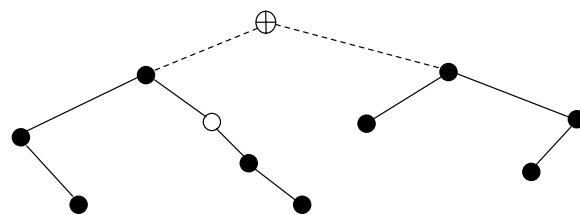


Рис. 2. Язык болгар походит на язык хазар, буртасы имеют другой язык

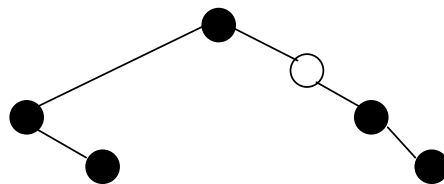


Рис. 3. Язык болгар походит на язык хазар

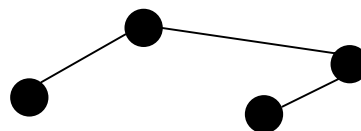


Рис. 4. Буртасы имеют другой язык

Для представления данных конструкций в виде импликации достаточно преобразовать их в сложноподчиненные. Заметим, что при этом преобразовании «левая» ветка графа (соответствующая подлежащему) перейдет в antecedent, а «правая» (соответствующая группе сказуемого) — в consequent.

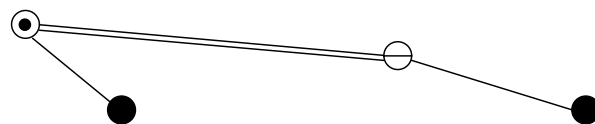


Рис. 5. Если x является представителем народа болгар, то x имеет язык, похожий на язык хазар

В терминах импликации $A(x) \Rightarrow B(x)$ antecedentу соответствует предикат $A(x)$ (x является представителем народа болгар), а предикату $B(x)$ — consequent (x имеет язык, похожий на язык хазар). И antecedent и consequent изображены на рисунке 5 отдельными узлами в виде метки ●. Аналогично преобразуется и вторая конструкция (рис. 6).



Рис. 6. Если x является представителем народа буртасов, то x имеет другой язык

II. Пример построения модели мира

Работу предложенного подхода продемонстрируем на примере анализа текстов, приведенных в приложении 1 и 2. Конечной целью является построение на их основе матрицы дизъюнктов согласно алгоритму, описанному в первой и второй частях статьи.

Анализировать будем не все тексты целиком, а только части, выделенные жирным шрифтом. Данные части представляют собой цельный по смыслу комплекс сведений об определенных сторонах жизни рассматриваемых народов (их религиозная, языковая принадлежность, форма прав-

ления, вид жилища и проч.). Для удобства выделенные части пронумерованы. В приложении 1 сохраняется оригинальная нумерация арабскими цифрами, а в приложении 2 автор вводит нумерацию римскими цифрами. Чтобы сослаться на какую-либо выделенную часть текста, автор будет указывать ее номер (арабскими для приложения 1 и римскими цифрами для приложения 2).

III.1. Список признаков

С учетом изложенных во второй части рассуждений, автором предлагается следующий список признаков (см. табл. 1).

Таблица 1

Список признаков

N	Признак	Значения
x_1	являться представителем народа	1 — бургасов; 2 — болгар; 3 — русов, 4 — хазар
x_2	иметь религию_1	1 — мусульманскую; 2 — немусульманскую
x_3	иметь религию_2	1 — как у гузов; 2 — не как у гузов
x_4	иметь язык, похожий на тюркский	1 — да; 2 — нет
x_5	иметь язык, похожий на язык болгар	1 — да; 2 — нет
x_6	иметь язык, похожий на язык русов	1 — да; 2 — нет
x_7	иметь язык, похожий на язык хазар	1 — да; 2 — нет
x_8	иметь язык, похожий на язык бургасов	1 — да; 2 — нет
x_9	иметь правителем	1 — одно лицо; 2 — двух лиц; 3 — многих лиц; 4 — ни одного лица
x_{10}	иметь жилищем шатер	1 — да; 2 — нет
x_{11}	иметь войлочное жилище ²³	1 — да; 2 — нет
x_{12}	иметь жилищем (деревянный) дом	1 — да; 2 — нет
x_{13}	народ разделен на группы	1 — нет групп; 2 — две; 3 — три; 4 — много групп
x_{14}	вести войны	1 — да; 2 — нет
x_{15}	вести междоусобные войны	1 — да; 2 — нет
x_{16}	вести войны с неверными	1 — да; 2 — нет
x_{17}	вести войны с булгарами	1 — да; 2 — нет
x_{18}	вести войны с печенегами	1 — да; 2 — нет
x_{19}	вести войны с бургасами	1 — да; 2 — нет
x_{20}	способ погребения_1	1 — сжигают; 2 — хоронят в землю; 3 — иначе
x_{21}	мусульманский способ погребения	1 — да; 2 — нет
x_{22}	подчиняться чужеземной власти	1 — да; 2 — нет
x_{23}	иметь меха источником богатства	1 — да; 2 — нет
x_{24}	иметь мед источником богатства	1 — да; 2 — нет
x_{25}	иметь вьючные животные источником богатства	1 — да; 2 — нет
x_{26}	иметь крупный рогатый скот источником богатства	1 — да; 2 — нет
x_{27}	иметь овец источником богатства	1 — да; 2 — нет
x_{28}	иметь свиней источником богатства	1 — да; 2 — нет
x_{29}	иметь кольчуги источником богатства	1 — да; 2 — нет
x_{30}	иметь оружие источником богатства	1 — да; 2 — нет
x_{31}	иметь золотую монету источником богатства	1 — да; 2 — нет
x_{32}	иметь серебряную монету источником богатства	1 — да; 2 — нет
x_{33}	иметь куниц источником богатства	1 — да; 2 — нет
x_{34}	иметь верблюдов источником богатства	1 — да; 2 — нет
x_{35}	иметь коров источником богатства	1 — да; 2 — нет
x_{36}	иметь лошадей источником богатства	1 — да; 2 — нет

III.2. Формирование дизъюнктов

Результаты работы предложенного алгоритма по построению матрицы дизъюнктов сведены в таблицу 2. В первом столбце данной таблицы указан порядковый номер строки матрицы (номер соответствующего дизъюнкта). Во втором столбце приведена конструкция, на основе которой построен дизъюнкт. Если конструкция получена из текстов приложений 1 и 2, то в третьем

столбце указана ссылка на номер соответствующего предложения. Если конструкция вспомогательная, то она также выделяется в третьем столбце. В четвертом столбце приводится непосредственно дизъюнкт в виде секционированного булева вектора. Причем для удобства указываются только значимые (ненулевые) секции. Порядковые номера соответствующих секций приводятся в пятом столбце таблицы.

Таблица 2

Алгоритм построения матрицы дизъюнктов

NN	Конструкция	NN предл.	Дизъюнкт	NN секций
1	У буртасов деревянные дома	23	0111.10	1, 12
2	У буртасов шатры	23	0111.10	1, 10
3	Буртасы находятся в подчинении у царя хазар	24	0111.10	1, 22
4	У буртасов нет главы или начальника	25	0111.0001	1, 9
5	Буртасы постоянно воюют с булгарами и печенегами	28	0111.10	1, 14
6	Буртасы постоянно воюют с булгарами и печенегами	28	0111.10.10	1, 17, 18
7	Вера буртасов походит на веру гузов	30	0111.10	1, 3
8	Наибольшая часть достояния буртасов — куницы. Имеется у буртасов мед, а также верблюды, крупный рогатый скот (или коровы), овцы (или свиньи)	33	0111.10.10.10.10.10.10	1, 24, 26, 27, 28, 33, 34, 35
9	Буртасы делятся на два разряда	34	0111.0100	1, 13
10	Одни буртасы предают мертвых сожжению, другие буртасы мертвых хоронят	34	0111.110	1, 20
11	Царь болгар по имени Ал.м.с исповедует ислам	I	1011.1000	1, 9
12	Царь болгар по имени Ал.м.с исповедует ислам	I	1011.01	1, 22
13	Булгары делятся на три разряда: каждый из этих трех разрядов носит свое название	II	1011.0010	1, 13
14	Булгары постоянно нападают на буртасов и полонят их	III	1011.10	1, 14
15	Булгары постоянно нападают на буртасов и полонят их	III	1011.10	1, 19
16	У болгар вьючные животные (или лошади), кольчуги и оружие. Основное имущество у болгар — меха куницы	IV, VI	1011.10.10.10.10.10	1, 25, 29, 30, 33, 36
17	Равным образом болгарские кладбища подобны мусульманским	V	1011.10	1, 21
18	У болгар нет золотой монеты	VII	1011.01	1, 31
19	У болгар нет серебряной монеты	VII	1011.01	1, 32
20	Язык болгар походит на язык хазар	X	1011.10	1, 7
21	Язык хазар походит на язык болгар	X	1110.10	1, 5
22	У буртасов отличный от болгар язык	X	0111.01	1, 5
23	У болгар отличный от буртасов язык	X	1011.01	1, 8
24	Язык русов не походит на язык хазар	X	1101.01	1, 7
25	Язык хазар не походит на язык русов	X	1110.01	1, 6
26	Язык русов не походит на язык буртасов	X	1101.01	1, 8
27	Язык буртасов не походит на язык русов	X	0111.01	1, 6
28	Вера гузов не является мусульманской	вспом.	01.01	2, 3
29	Куницы являются мехами	вспом.	10.01	23, 33
30	Верблюды являются вьючным скотом	вспом.	10.01	25, 34
31	Лошади являются вьючным скотом	вспом.	10.01	25, 36
32	Коровы являются крупным рогатым скотом	вспом.	10.01	26, 35
33	Мусульмане хоронят мертвых в землю	вспом.	01.010	2, 20
34	С неверными сражаются мусульмане	вспом.	10.01	2, 16

Заметим, что в таблицу 2 не вошли конструкции: «В обоих городах болгар деревянные строения» и «Жители этих городов сражаются с неверными» (VII, IX). Как уже указывалось, данные конструкции сводятся к формулам с квантором существования: $\exists u: (u_1 = "2") \Rightarrow (u_{12} = "1")$ и $\exists u: (u_1 = "2") \Rightarrow (u_{16} = "1")$ соответственно. Вопрос преобразования данных формул в дизъюнкты остается открытым.

Автор признателен профессору Л. И. Бородкину за полезные консультации и советы.

Приложение 1

Фрагмент из публикации Б. Н. Заходера
«Каспийский свод о Восточной Европе.

Горган и Поволжье в IX–X вв.» (М., 1962. С. 27–28)

Буртасы

22. Буртасы находятся между булгарами и хазарами; между булгарами и хазарами расстояние в 15 дней в пути. 23. Соседствуя с хазарами, буртасы живут на реке, название которой также Буртас (или Итиль); **у них деревянные дома, в них они живут зимою, и шатры, в них они обитают летом**; Буртас — название области, так же как Рус, Хазар; Сарир — название царства, а не города и не народа. 24. **Буртасы находятся в подчинении у царя хазар, они поставляют десять тысяч всадников.** 25. **У буртасов нет главы или начальника, который управлял бы ими; в каждой местности у них имеется один или два старца, которые вершат дела (или суд) между ними.** 26. Земля буртасов — обширная; они живут в лесах (или: земля буртасов — ровная; большая часть их деревьев — халандж; это дерево они вывозят в Хорасан). 27. Земля буртасов в длину и ширину — 15 дней (или 17 дней, или полтора месяца) пути; здесь много торговых мест. 28. **Буртасы постоянно воюют с булгарами и печенегами.** 29. Буртасы — стойки и неустрашимы; они — совершенны лицом и общим видом. 30. **Вера буртасов походит на веру гузов.** 31. Когда поступит один из буртасов вероломно по отношению к другому, начнет его притеснять, причинит ему ранение или побои, тот, другой, не примирится, пока не добьется возмездия. 32. Когда буртасская девушка достигает зрелости, она перестает повиноваться отцу, выбирает сама того, кто ей понравился из мужчин, тот мужчина сватается сам (или через свата) и женится на девушке. 33. **Наибольшая часть достояния буртасов — куницы, из шкур которых делают шубы. Имеется у буртасов мед, а также верблюды, крупный рогатый скот (или коровы), овцы (или свиньи).** 34. Буртасы делятся на два разряда: **одни предают мертвых сожжению, другие хоронят их.** 35. У буртасов (обширные) па-

хотные поля. 36. От области буртасов до хазар все степь, дорога — благоустроенная, с источниками, деревьями, текучими водами; некоторые же, отправляясь от буртасов к хазарам, плывут по реке Итиль на судах, а некоторые (путешествуют) сухим путем. 37. Оружием буртасов служат два кинжала, секира, лук, у них нет панциря и кольчуги; лошадью у них владеет не каждый, а только тот, кто обладает многим достатком; их одежда — архалук и халат; у них головной убор, а чалму завязывают вокруг него. 38. У буртасов нет фруктов, и вино их — из меда. 39. К востоку от области буртасов река Итиль, к югу — хазары, к западу — в. н. нд.р., к северу — тюркские печенеги.

Приложение 2

Отрывки из публикации Б. Н. Заходера
«Каспийский свод о Восточной Европе».

Том II. Булгары, мадьяры, народы Севера, печенеги, русы, славяне» (М., 1967. С. 23–46)

Булгары

Страна болгар соседит со страной буртасов. Между буртасами и булгарами три дня пути. Булгары обитают на берегу (или на берегах) реки, которая втекает в Хазарское море и называется Итиль; эта река между хазарами (или буртасами) и славянами (гузами). (I) **Царь болгар по имени Ал.м.с. исповедует ислам; у него сородичи числом пятьсот (или пятьсот тысяч человек).** Земля болгар в зарослях и дремучих чащобах, среди которых булгары и обитают. (II) **Булгары делятся на три разряда; каждый из этих разрядов носит свое название.** Хазары торгуют с булгарами, также русы привозят свои товары, также и другие народы; привозят к ним меха соболя, горноста, белки и другие меха. У болгар имеется земледелие; они выращивают разные зерновые. (III) **Булгары постоянно нападают на буртасов и полонят их.** (IV) **У болгар — вьючные животные (или лошади), кольчуги и оружие.** Они дают царю дань вьючными животными; когда кто-либо из болгар женится, царю также приносят в дар вьючное животное; с приходивших в Булгар судов царь берет десятину. Одежда болгар походит на одежду мусульман, (V) **равным образом болгарские кладбища подобны мусульманским.** (VI) **Основное имущество у болгар — меха куницы;** (VII) **у болгар нет золотой и серебряной монеты,** а расплачиваются они кунным мехом, один мех равен двум с половиной дирхемам; дирхемы же белые, круглые привозят к ним из мусульманских стран. Земля болгар в северной стороне по направлению к полюсу. Между булгарами и Хорезмом путь равен трем месяцам. У болгар два города, имя одному Сувар, другому Булгар; оба города побли-

зости (или в двух днях пути) друг от друга; (VIII) **в обоих городах деревянные строения**, соборная мечеть, живут там мусульмане, по 10 тысяч человек в каждом городе; (IX) **они сражаются с неверными**. У болгар деревья по большей части хаданк и орешник; в болгарских лесах водятся белки, соболь и другие звери с (ценным) мехом. У болгар летом ночь так коротка, что не успевает вскипеть котелок (или не успевает человек пройти более одного фарсаха); зимой же день становится таким же коротким, как ночь летом. (X) **Язык болгар походит на язык хазар, у буртасов другой язык, язык русов не походит на язык хазар и буртасов**. На восток и юг от болгар — гузы, к западу — река Итиль, на север — область печенегов. О посылке в 415/1024–1025 г. болгарским государем денег и подарков в Хорасан на построение соборных мечетей.

Приложение 3
Фрагменты из публикации В. В. Бартольда
«Введение к изданию Худуд ал-‘алам» //
Сочинения (М., 1973. Т. 8. С. 545)

Слово об области буртасов

Люди там мусульмане; у них особый язык; царя называют М-с. Они владеют шатрами и палатками. Их три группы: бахдула, ишкиль и болгар; все находятся в войне друг с другом; когда же появляется враг, они становятся друг другу друзьями.

Слово об области берадасов

Они — люди, придерживающиеся веры гузов; владеют палатками; мертвых сжигают; повинуются хазарам; доход от мехов горностаев. У них два царя, не общающиеся друг с другом.

Булгар

Булгар — город с небольшой областью, расположенный на берегу Итиля. В нем все [жители] мусульмане; из него выходит до 20 000 всадников. Со всяким войском кафиров, сколько бы его не было, они сражаются и побеждают. Это — место крепкое, богатое. Сувар — город вблизи Болгара; в нем борцы за веру, так же как и в Булгаре.

ПРИМЕЧАНИЯ

- ¹ Лихачев Д. С. Текстология. Л., 1983. 639 с.
- ² Бородкин Л. И., Милов Л. В. О некоторых аспектах автоматизации текстологического исследования (Закон Судный людям) // Математические методы в историко-экономических и историко-культурных исследованиях : сборник статей. М., 1977. С. 235–280.
- ³ Шпирко С. В. Применение теории нечетких множеств к задаче генеалогической классификации в текстологическом исследовании // Историческая информатика. 2013. № 3. С. 39–52.
- ⁴ Закревский А. Д. Представление знаний и логический вывод в пространстве многозначных признаков // Логика и компьютер. Логические языки, содержательные рассуждения и методы поиска доказательств. Вып. 2. М., 1995. С. 3–16.
- ⁵ Падучева Е. В. О семантике синтаксиса. Материалы к трансформационной грамматике русского языка. М., 2009. 296 с.
- ⁶ Левицкий Ю. А. Основы теории синтаксиса. М., 2014. 368 с.
- ⁷ Гуц А. К. Математическая логика и теория алгоритмов. М., 2009. 120 с.
- ⁸ Шпирко С. В. Логика распознавания А. Д. Закревского на примере задачи восстановления пропусков в историко-археологических исследованиях // Историческая информатика. 2014. № 1 (7). С. 62–69.
- ⁹ Закревский А. Д. Указ соч.
- ¹⁰ Заходер Б. Н. Каспийский свод сведений о Восточной Европе. Горган и Поволжье в IX–X веках. М., 1962. 280 с.
- ¹¹ Заходер Б. Н. Каспийский свод сведений о Восточной Европе. Т. II: Булгары, мадьяры, народы Севера, печенеги, русы, славяне. М., 1967. 213 с.
- ¹² Бартольд В. В. Введение к изданию Худуд ал-‘алам // Сочинения. М., 1973. Т. 8. 725 с.
- ¹³ Закревский А. Д. Указ соч.
- ¹⁴ Там же.
- ¹⁵ Гуц А. К. Указ соч.

-
- ¹⁶ Хвольсон Д. А. Известия о хазарах, бургасах, болгарах, мадьярах, славянах и руссах Абу-Али Ахмеда бен Омар ибн-Дада. Минск, 2010. 200 с.
- ¹⁷ Бартольд В. В. Указ. соч.
- ¹⁸ Калинина Т. М. Бургасы — мусульмане? // Восточная Европа в древности и средневековье : материалы конференции. М., 2014. С. 108–113.
- ¹⁹ Гуц А. К. Указ. соч.
- ²⁰ Падучева Е. В. Указ. соч.
- ²¹ Левицкий Ю. А. Указ. соч.
- ²² Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. Киев, 1981. 192 с.
- ²³ Непонятно разграничение у Б. Н. Заходера между войлочным жилищем и шатром.
-