

## Концептуальное моделирование текстов тувинского языка с использованием методов анализа формальных понятий

*Ч.М. Монгуш*

*Тувинский государственный университет, Кызыл*

**Аннотация:** Одним из важных составляющих этнокультурного наследия Республики Тыва является анализ текстов тувинских героических сказаний. В научной базе Тувинского института гуманитарных и прикладных социально-экономических исследований Республики Тыва хранятся магнитофонные и рукописные записи всех жанров тувинского фольклора, в том числе, в старых ветхих изданиях находятся примерно 300 тувинских героических сказаний. В настоящее время преподавателями и студентами Тувинского государственного университета создана электронная коллекция «Тувинские героические сказания» и внедрена в электронный корпус тувинского языка. В этой коллекции имеются оцифрованные тексты тувинских героических сказаний, их метаописания и сведения о сказителях. С использованием данной электронной коллекции можно решать лингвистические и филологические задачи, которые сводятся к задаче концептуального моделирования коллекции «Тувинские героические сказания». В настоящей статье рассматривается задача установления авторского стиля сказителей при описании снаряжений коня в произведениях тувинского героического эпоса. Для решения данной задачи используется алгебраический подход, который в литературе называется анализом формальных понятий. В рамках анализа формальных понятий коллекция «Тувинские героические сказания» представляется формальным контекстом. Тогда решение поставленной задачи направлено на выявление множества всех формальных понятий формального контекста и связывание их в решетку. Полученная решетка служит концептуальной моделью и основой для решения поставленной задачи.

**Ключевые слова:** электронный корпус тувинского языка, тувинский героический эпос, авторский стиль сказителя, распознавание образов, анализ формальных понятий.

### Введение

Электронная коллекция «Тувинские героические сказания» – информационная составляющая электронного корпуса тувинского языка [1]. В этой коллекции представлены паспорта тувинских героических сказаний в виде объектно-признаковой таблицы. В таблице всякий столбец отвечает некоторому признаку, а строка устанавливает признаковое описание некоторого эпоса. В паспорт произведения входят сведения о сказителе, библиографические сведения, жанровые, стилевые и другие особенности текста. Таким образом, объектно-признаковая таблица является моделью представления электронной коллекции «Тувинские героические сказания»,

---

что допускает применение математических методов при решении прикладных задач таких, как установление авторских особенностей сказителей, использование языковых стандартов и клише, применение диалектных вариантов и т.д. Эти задачи сводятся к задаче концептуального моделирования электронной коллекции «Тувинские героические сказания».

В данной статье рассматривается задача установления авторского стиля сказителей при описании доспехов коня в произведениях тувинского героического эпоса. Для решения этой задачи применяется математический аппарат – анализ формальных понятий. Он возник как прикладное направление теории решеток [2, 3]. Анализ формальных понятий на основе алгебраической теории решеток Г. Биркгофа позволяет построить смысловую структуру предметной области [4 – 6]. В рамках этого подхода электронная коллекция «Тувинские героические сказания» представляется формальным контекстом. Тогда решение рассматриваемой задачи направлено на выявление множества всех формальных понятий заданного формального контекста и связывание их в решетку формальных понятий. В свою очередь, построенная решетка является концептуальной моделью предметной области и основой для решения прикладных задач.

### **Концептуальное моделирование коллекции «Тувинские героические сказания»**

Рассмотрим основные понятия математического аппарата, применяемые в настоящей статье [7, 8].

Пусть для рассматриваемой предметной области определены непустые конечные множества:  $G$  – множество объектов,  $M$  – множество признаков. Непустое бинарное отношение инцидентности  $I$  между этими множествами трактуется следующим образом: для  $g \in G$  и  $m \in M$  пара  $(g, m)$  означает, что объект  $g$  обладает признаком  $m$  и наоборот, признак  $m$  свойственен

---

объекту  $g$ . Тогда в анализе формальных понятий тройка  $K = (G, M, I)$  называется формальным контекстом предметной области. В таблице № 1 представлен пример матрицы инцидентности  $I$  формального контекста  $K = (G, M, I)$ , где  $G = \{1, 2, 3, 4\}$ ,  $M = \{a, b, c, d\}$ . А символ «+» означает отношение инцидентности, т.е. объект  $g$  обладает признаком  $m$ .

Таблица № 1

Формальный контекст  $K = (G, M, I)$

$g \backslash m$	$a$	$b$	$c$	$d$
1	+		+	
2			+	+
3	+			+
4	+	+	+	
5		+	+	

Если  $A \subseteq G$  и  $B \subseteq M$ , то пара операторов:

$$A' = \bigcap_{g \in A} g' = \{m \in M \mid \forall g \in A (g, m) \in I\}, \quad (1)$$

$$B' = \bigcap_{m \in B} m' = \{g \in G \mid \forall m \in B (g, m) \in I\}. \quad (2)$$

называется отображениями Галуа и задает соответствие между частично упорядоченными множествами  $(2^G, \subseteq)$  и  $(2^M, \subseteq)$ . В этом случае  $A'$  – множество признаков, характерных всем объектам из  $A$ , а  $B'$  – совокупность объектов, обладающих всеми признаками из  $B$ . Двойное применение отображения Галуа определяет оператор замыкания на совокупность всех подмножеств множеств объектов, признаков, и обозначается:  $A''$  и  $B''$ .

Формальным понятием называется пара множеств  $(A, B)$ , где  $A \subseteq G$  и  $B \subseteq M$ , если выполняется  $A' = B$  и  $B' = A$ . Другими словами, множества  $A, B$  образуют формальное понятие  $(A, B)$  в формальном контексте  $K = (G, M, I)$  тогда и только тогда, когда:

$$A = A'' \text{ и } B = B''. \quad (3)$$

Рассмотрим подмножества  $A = \{4, 5\}$  и  $B = \{b, c\}$  множеств  $G, M$  из примера. Для подмножеств  $A$  и  $B$  найдем  $A''$  и  $B''$ , используя формулы (1) и (2):

$$A'' = (A')' = (\{b, c\})' = \{4, 5\} = A.$$

$$B'' = (B')' = (\{4, 5\})' = \{b, c\} = B.$$

Тогда согласно формуле (3) пара  $(A, B)$  образует формальное понятие формального контекста  $K = (G, M, I)$ , т.е. объекты 4 и 5 обладают признаками  $b, c$ . В таблице № 1 формальное понятие  $(A, B)$  выделено двойными линиями. Определение формального понятия полностью соответствует традиционной трактовке термина «понятие» в гуманитарных науках. Объемом формального понятия является множество  $A$ , а содержанием – множество  $B$ . А также, если уменьшить объем формального понятия, то увеличивается его содержание и наоборот, если уменьшить содержание, то объем формального понятия станет больше.

Обозначим через  $FC$  – множество всех формальных понятий формального контекста  $K = (G, M, I)$ . Тогда введем на  $FC$  отношение частичного порядка  $\sqsubseteq$ :

$$(A_1, B_1) \sqsubseteq (A_2, B_2), \text{ если } A_1 \subseteq A_2 \text{ (или } B_2 \subseteq B_1),$$

где  $A_1, A_2 \subseteq G$  и  $B_1, B_2 \subseteq M$ .

На множество всех формальных понятий  $FC$  определим операции пересечения  $\sqcap$  и объединения  $\sqcup$  при помощи  $\cap$  и  $\cup$ :

$$(A_1, B_1) \sqcap (A_2, B_2) = (A_1 \cap A_2, (A_1 \cap A_2)'),$$

$$(A_1, B_1) \sqcup (A_2, B_2) = ((B_1 \cap B_2)', B_1 \cap B_2).$$

В этом случае, частично упорядоченное множество  $(FC, \sqsubseteq)$  образует решетку  $L = (FC, \sqcap, \sqcup)$ , которая называется решеткой формальных понятий контекста  $K = (G, M, I)$ . Решетка формальных понятий формального контекста является концептуальной моделью, которая определяет смысловую

структуру исследуемой предметной области, и служит основой для решения прикладных задач [8, 9]. В качестве примера рассмотрим задачу распознавания авторских особенностей при описании снаряжений коня в текстах тувинского героического эпоса.

### Установление авторского стиля сказителей тувинского героического эпоса

Для решения задачи распознавания авторского стиля сказителей с помощью математического аппарата анализа формальных понятий необходимо:

- 1) формировать формальный контекст тувинских героических сказаний  $K = (G, M, I)$ ;
- 2) найти множество всех формальных понятий  $FC$ ;
- 3) построить концептуальную модель рассматриваемой предметной области  $L$ ;
- 4) производить запросы на установление авторского стиля сказителей.

Формирование формального контекста  $K = (G, M, I)$  осуществляется на основе баз данных «Тувинские героические сказания», «Клише и стандарты в текстах тувинских героических сказаниях» и разработанной программы «Программа формирования контекста для электронной коллекции Тувинские героические сказания» [10]. Формальный контекст  $K = (G, M, I)$  содержит 12 тувинских героических сказаний:

$G = \{ \text{Демир-Шилги аъттыг Тевене-Мөге; Мөге Шагаан-Тоолай; Танаа-Херел; Каңгывай-Мерген; Алдын-Чаагай; Хан-Шилги аъттыг Хан-Хүлүк; Арзылаң-Кара аъттыг Чечен-Кара Мөге, Арзылаң-кара аъттыг Хунан-Кара, Сарыг-Хемниң иштин чурттаан Тавын-Хаан, Хан-Шилги аъттыг Хан-Күчү-Маадыр, Алдын-сарыг аъттыг Аңчы-Кара, Элестей ашак} ,$

а в качестве признаков выбираются языковые стандарты, описывающие снаряжения коня:

$M = \{ \text{стандарты, характеризующие седло; стандарты, характеризующие хлыст; стандарты, характеризующие потник; стандарты, характеризующие аркан, стандарты, характеризующие узду; стандарты, характеризующие лассо; стандарты, характеризующие стремяна} \}$ .

Нахождение множества всех формальных понятий  $FC$  и построение решетки формальных понятий  $L$  реализуется с помощью разработанной программы «Программа FSCorpus концептуального моделирования тувинских текстов методами анализа формальных понятий». Применение данного комплекса программ к построенному формальному контексту  $K = (G, M, I)$  дало множество  $FC$ , содержащее тридцать формальных понятий. С помощью найденных формальных понятий программа FSCorpus построила концептуальную модель, которая определяет смысловую структуру исследуемой предметной области. Данная решетка формальных понятий является основой для решения лингвистических и филологических задач, в том числе, распознавания авторских особенностей сказителей тувинских героических сказаний.

Для установления авторских особенностей необходимо выполнять запросы на выявление знаний из решетки формальных понятий. Например, задан запрос: необходимо найти все языковые стандарты, которые характеризуют авторский стиль сказителя Ооржак Ч-Х.Ч., т.е. найти все формальные понятия, включающие в содержание Ооржак Ч-Х.Ч.

При реализации данного запроса программа FSCorpus отбирает из 30 формальных понятий 8 и строит решетку, которая описывает авторские особенности сказителя Ооржак Ч-Х.Ч. (рис. 1).

На рис. 1 формальное понятие:

---

$\{(q_0, q_7, q_{11}); (m_{43}, m_{48})\} = \{(\text{Демир-Шилги аъттыг Тевене-Мөге, Арзылаң-кара аъттыг Хунан-Кара, Элестей ашак}); (\text{Ооржак Чанчы-Хөө Чапаажыкович, Барун-Хемчикский р-н})\}$

означает, что авторский стиль сказителя Ооржак Чанчы-Хөө Чапаажыковича из Барун-Хемчикского района выявляются в эпосах Демир-Шилги аъттыг Тевене-Мөге, Арзылаң-кара аъттыг Хунан-Кара, Элестей ашак.

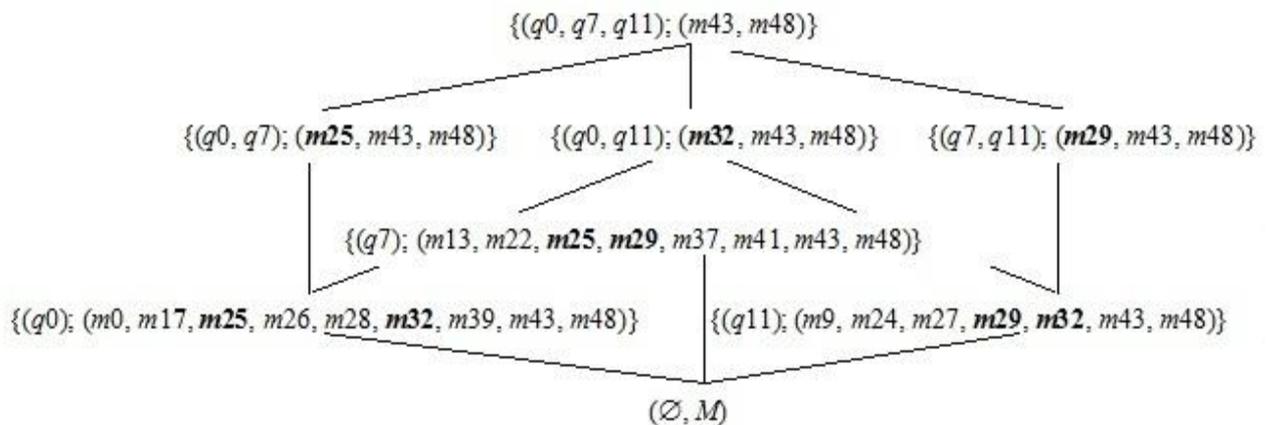


Рис. 1. – Решетка формальных понятий, характеризующая авторский стиль сказителя Ооржак Ч-Х.Ч.

В решетке формальные понятия означают:

- $\{(q_0, q_7); (m_{25}, m_{43}, m_{48})\} = \{(\text{Демир-Шилги аъттыг Тевене-Мөге, Арзылаң-кара аъттыг Хунан-Кара}); (\text{Алдын шалба, Ооржак Чанчы-Хөө Чапаажыкович, Барун-Хемчикский р-н})\},$
- $\{(q_0, q_{11}); (m_{32}, m_{43}, m_{48})\} = \{(\text{Демир-Шилги аъттыг Тевене-Мөге, Элестей ашак}); (\text{Алдан кулаш сыдым, Ооржак Чанчы-Хөө Чапаажыкович, Барун-Хемчикский р-н})\},$
- $\{(q_7, q_{11}); (m_{29}, m_{43}, m_{48})\} = \{(\text{Арзылаң-кара аъттыг Хунан-Кара, Элестей ашак}); (\text{Хүмүш чүген, Ооржак Чанчы-Хөө Чапаажыкович, Барун-Хемчикский р-н})\}.$

Таким образом, при описании аркана сказитель Ооржак Чанчы-Хөө Чапаажыкович применяет клише «Алдын шалба», а лассо характеризует с помощью языкового стандарта «Алдан кулаш сыдым», сказитель изображает узду, применяя стандарт «Хүмүш чүген».

В решетке последние формальные понятия описывают отдельные произведения сказителя Ооржак Чанчы-Хөө Чапаажыкович со своими характерными языковыми стандартами.

Аналогичным образом выполняется запрос: необходимо найти все языковые стандарты, которые характеризуют авторский стиль сказителя Ондар Тевек-Кежеге. В этом случае программа FCSOcrpus выдала решетку формальных понятий, которая приведена на рис. 2.

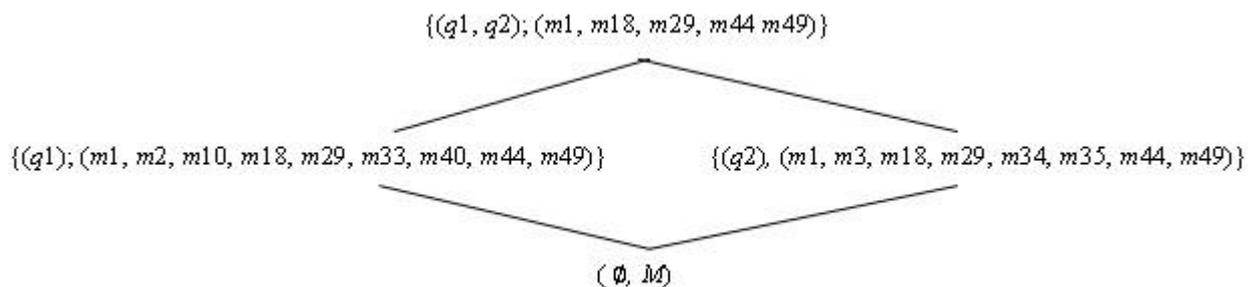


Рис. 2. – Решетка формальных понятий, характеризующая авторский стиль сказителя Ондар Тевек-Кежеге

В решетке согласно формальному понятию  $\{(q1, q2); (m1, m18, m29, m44 m49)\} = \{( \text{Мөге Шагаан-Тоолай, Танаа-Херел}); ( \text{Арт болган алчайган-калчайган эзер, Хөл болган хөлбең кара чонок, Хүмүш чүген, Ондар Тевек-Кежеге, Сут-Хольский р-н})\}$  сказителю Ондар Тевек-Кежеге из Сут-Хольского района присущи языковые стандарты «Арт болган алчайган-калчайган эзер», «Хөл болган хөлбең кара чонок», «Хүмүш чүген» при описании седла, потника и узды.

Следует заметить, что сказители Ооржак Чанчы-Хөө Чапаажыкович и Ондар Тевек-Кежеге используют общий языковой стандарт ««Хүмүш чүгөн» при описании узды.

### **Заключение**

Экспертами было установлено, что полученные результаты в статье являются лингвистически правильными, то есть соответствуют действительности. Для полного анализа авторских особенностей сказителей тувинского эпоса с помощью анализа формальных понятий необходимо расширить формальный контекст. Также из полученных результатов следует, что созданные средства могут быть использованы для решения аналогичных задач анализа текстов в рамках электронного корпуса тувинского языка.

*Исследование выполнено при финансовой поддержке гранта Президента Российской Федерации для молодых ученых – кандидатов наук МК-2437.2022.1.1 «Разработка математического метода и средств для исследования текстов на тувинском языке».*

### **Литература**

1. Салчак А.Я., Ондар В.С. Создание русско-тувинского параллельного подкорпуса электронного корпуса тувинского языка: первые итоги // Новые исследования Тувы, 2020, № 1 URL: [nit.tuva.asia/nit/article/view/912](http://nit.tuva.asia/nit/article/view/912)
2. Ganter B., Wille R. Formal Concept Analyses: Mathematical Foundations. Springer Science and Business Media, 2012. 314 p.
3. Кузнецов С.О. Автоматическое обучение на основе анализа формальных понятий // Автоматика и телемеханика. 2001. № 10. С. 3–27.
4. Быкова В.В., Монгуш Ч.М. Декомпозиционный подход к исследованию формальных контекстов // Прикладная дискретная математика. 2019. № 44. С. 113–126.

5. Белим С.В., Богаченко Н.Ф. Использование решетки формальных понятий для построения ролевой политики разграничения доступа // Информатика и системы управления. 2019. № 1(55). С. 16–28.

6. Kuznetsov S.O., Ganter B., Eklund P.W., Sertkaya B. Machine Learning and Formal Concept Analysis // 2th International Conference on Formal Concept Analysis: proceedings. Berlin Heidelberg : Springer, 2004. Vol. 2961. P. 287–312.

7. Mongush Ch.M. Bykova V.V. On decomposition of a binary context without losing formal concepts // Journal of Siberian Federal University. Mathematics and Physics. 2019. Vol. 12. № 3. P. 323–330.

8. Маслова М.А., Дмитриев А.С., Холкин Д.О. Методы распознавания именованных сущностей в русском языке // Инженерный вестник Дона, 2021, №7. URL: [ivdon.ru/ru/magazine/archive/n7y2021/7066](http://ivdon.ru/ru/magazine/archive/n7y2021/7066)

9. Карелова О.В. К вопросу изучения индивидуального стиля автора // Известия Российского государственного педагогического университета им. А.И. Герцена. 2006. Т. 20. № 3. С. 24–29.

10. Монгуш Ч.М. Программа формирования контекста для электронной коллекции «Тувинские героические сказания» // Инженерный вестник Дона, 2018, №2. URL: [ivdon.ru/ru/magazine/archve/N2y2018/5039](http://ivdon.ru/ru/magazine/archve/N2y2018/5039).

### References

1. Salchak A.Ja., Ondar V.S. The New Research of Tuva. 2020. №1. URL: [nit.tuva.asia/nit/article/view/912](http://nit.tuva.asia/nit/article/view/912).

2. Ganter B., Wille R. Formal Concept Analyses: Mathematical Foundations. Springer Science and Business Media, 2012. 314 p.

3. Kuznetsov S.O. Avtomatika i telemekhanika. 2001. №10. pp. 3–27.

4. Bykova V.V., Mongush Ch.M. Prikladnaia diskretnaia matematika. 2019. №44. pp. 113–126.

5. Belim S.V., Bogachenko N.F. Informatika i sistemy upravlenija. 2019. №1(55). pp. 16–28.



6. Kuznetsov S.O., Ganter B., Eklund P.W., Sertkaya B. 2th International Conference on Formal Concept Analysis: proceedings. Berlin Heidelberg, Springer, 2004, vol. 2961. pp. 287–312.
7. Mongush Ch. M., Bykova V.V. Journal of Siberian Federal University. Mathematics and Physics. 2019. vol. 12. №3. pp. 323–330.
8. Maslova M.A., Dmitriev A.S., Holkin D.O. Inzhenernyj vestnik Dona, 2021, №7. URL: [ivdon.ru/ru/magazine/archive/n7y2021/7066](http://ivdon.ru/ru/magazine/archive/n7y2021/7066)
9. Karelova O.V. Izvestiia Rossiiskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena. 2006. vol. 20. №3. pp. 24–29.
10. Mongush Ch.M. Inzhenernyj vestnik Dona, 2018, №2. URL: [ivdon.ru/ru/magazine/archve/N2y2018/5039](http://ivdon.ru/ru/magazine/archve/N2y2018/5039).