

**ПОДГОТОВКА МНОГОСЛОВНЫХ  
КОНСТРУКЦИЙ ДЛЯ БД ПОРТАЛА  
ТЮРКСКОЙ МОРФЕМЫ С ПОМОЩЬЮ  
ПРОГРАММНОЙ ОБРАБОТКИ  
ДВУЯЗЫЧНЫХ СЛОВАРЕЙ**

**Аюпов М.М.**

*Институт прикладной семиотики  
Академии наук Республики Татарстан*

*23 сентября 2021 г.*

# При создании больших многоязычных баз данных важная роль принадлежит заполнению этих баз данных.

## Портал "Тюркская Морфема"

Вы вошли как: Читатель

[Вики](#) [Форум](#) [Обзор](#) [Вход в систему](#)

Выбранный язык базы данных:	
Татарский	▼
Общая часть	
Грамматика	▼
Тезаурус	▼
Ситуации	▼
Языковая часть	
Морфемы	▼
Морфотактика	▼
Ситуации в языке	▼
Многословные выражения	▼

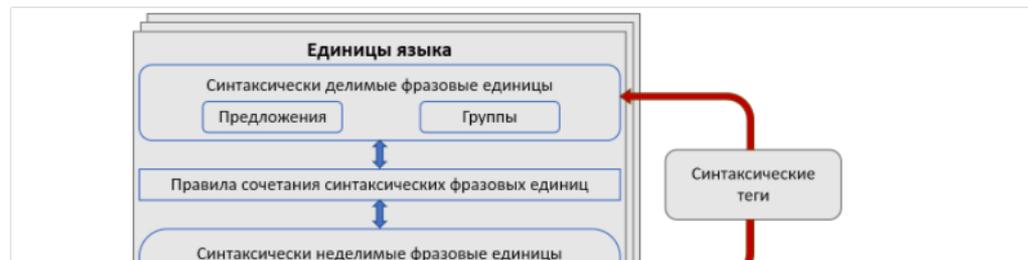
### Модель Тюркской Морфемы

Для проведения научных исследований в области тюркологии и типологии агглютинативных языков необходим программный инструментарий, который учитывает структурно-функциональные особенности рассматриваемых языков.

Портал «Тюркская морфема» – это инструментарий, который позволяет производить исследования с учетом этих особенностей тюркских языков и соответствует требованиям к научно-исследовательской деятельности в области компьютерной лингвистики, лингвистической типологии.

Портал создан на базе структурно-параметрической функциональной модели тюркской морфемы и содержит специальные лингвистические базы данных, описывающие языковые единицы тюркских языков на разных лингвистических уровнях: морфологическом, синтаксическом, семантическом.

Базы данных портала также могут быть использованы в учебном процессе, как информационно-справочная система по тюркским языкам.



Языки портала тюркской морфемы, БД для которых заполнена с помощью программной обработки двуязычных словарей:

- татарский,
- киргизский,
- узбекский,
- крымско-татарский,
- казахский,
- чувашский,
- башкирский,
- гагаузский,
- ногойский,
- азербайджанский,
- алтайский.

Russian	Semant
бегун	1192

Tatar	Semant
йөгерешче	1192

Uzbek	Semant
yuguruvchi	1192

Tatar	Uzbek
йөгерешче	yuguruvchi

Перевод между тюркскими языками:

- морфологический разбор текста по правилам соответствующего языка;
- перевод основы на нужный язык;
- обратный морфологический сбор по правилам языка, на который переводится текст.

Xalqaro ilmiy-amaliy konferensiya  
ishiga ham muvaffaqiyatlar tilayman.



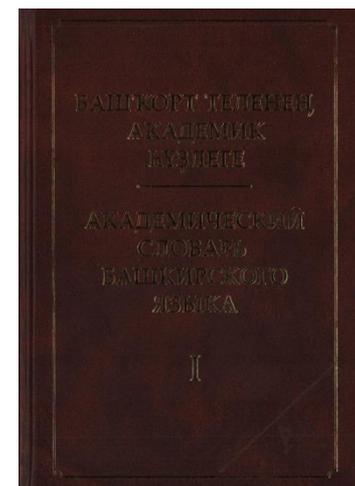
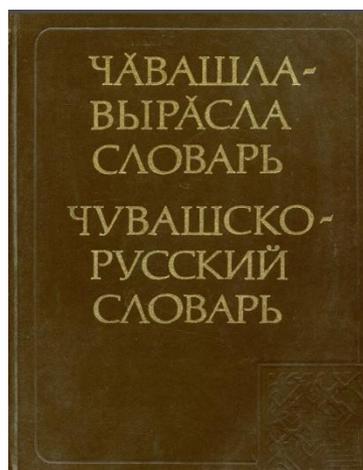
Халыкара фәнни-гамәли  
конференция эшендә уңышлар  
телим.

В БД портала тюркской морфемы необходимо хранить многословные конструкции.

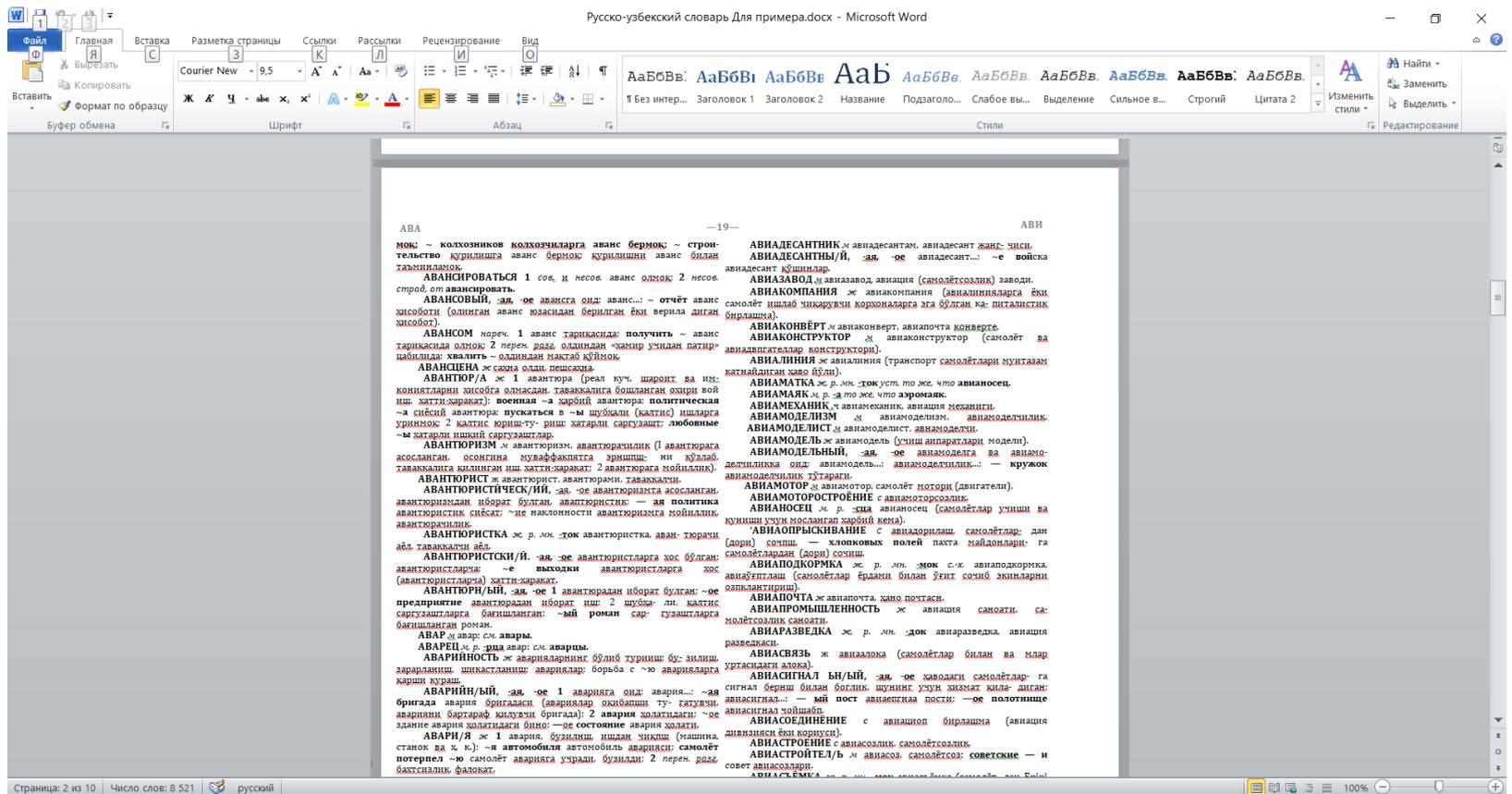
от альфы до омеги  
↓ ↓  
альфадан омегага

от альфы до омеги  
↓  
башыннан ахырына  
кадәр

Наиболее полные двуязычные словари в основном были созданы и изданы в прошлом веке. В связи с этим у них отсутствует электронная версия.



# 1. Найти отсканированную версию нужного словаря или, если нет версии с хорошим качеством, отсканировать.



2. Двужычные словари распознаются с помощью специальных программ распознавания.

### Русско-алтайский:

БРАТЬСЯ несов. 1. (за руки) тудунар (тудун-); брать за поручни тутканан тудунар; 2. (приниматься) баштаар (башта-); алынар (алын-); браться за работу иштеп баштаар; браться за ум сагыш алынар; 3. (появляться) табылар (табыл-); Откуда берётся эта мысль? Бу шүүлте кайдан табылып jat?¶

### Русско-казахский

БРАТЬСЯ несов. см. взяться алу; ~за умақылға келу, ақылға түсу; не берусь судить шешіп айта алмаймын; не берись за это дело бұл жұмысқа кіріспе; откуда берётся столько силы? мұншалықкүш қайдан шығады?; ~за оружиеқолға қару алу¶

### Русско-ногайский

**браться** несов. см. **взяться**; браться за ум ақылды йыйып алу, берлиликти (ақылсызлықты) конов, ақыллыланув.

3. С помощью программной обработки находится нужная информация.

Русско-алтайский:

братъся за ум сагыш алынар

Русско-казахский

братъся за ум ақылға келу, ақылға түсу

Русско-ногайский

братъся за ум акылды йыйып алу, берликти (акылсызлыкты) коюв, акыллыланув

4. Обработанный словарь готовится для загрузки в базу данных.

<b>Русский</b>	<b>Алтайский</b>	<b>Казахский</b>	<b>Ногайский</b>
браться за ум	сагыш алынар	ақылға келу	ақылды йыйып алу
браться за ум		ақылға түсу	берлиликти кою
браться за ум			ақыллылану

Возникают орфографические ошибки, если встречаются неуверенно распознанные слова и слова отсутствующие в словаре программы распознавания.

Отсутствие пробелов – распространенный случай при работе с распознанным текстом.

БРА`ТЬСЯ·несов.·см.взяться·алу; ~за·умақылға·келу,·ақылға·түсу;·не·берусь·судить·шешіп·айта·алмаймын;·не·берись·за·это·дело·бұл·жұмысқа·кіріспе;·откуда·берется·столько·силы?·мұншалықкүш·қайдан·шығады?; ~за·оружиеқолға·қару·алу¶

Во время обработки словарей знак переноса строки обычно означает начало новой статьи. Но так же встречаются лишние знаки переноса строки.

самостоятельный прил. 1. {независимый) ·ОБЗ·алдылык, бойсыныксыз, бойсынувсыз; самостоятельная жизнь ·бойсыныксыз яшав; 2. {совершаемый собственными силами) ·ОБЗ·алдына, ·ОБЗ·куьши·мен; самостоятельная работа ·ОБЗ·куьши·мен·этилинетаган·куллык. ¶

самотёк м. 1. {о·движении жидкости силой·собственной тяжести) ·ОБЗИ·обзиннен·агув; 2. перен. самотёк; пустить на самотёк ·самотёкка йиберув; ·ОБЗ¶ болтаны·бол·сын·деп·коюв. ¶

## Выводы:

- получилось сэкономить время и трудозатраты при подготовке многословных конструкций для загрузки в базу данных портала тюркской морфемы;
- БД многословных конструкций должна улучшить перевод между тюркскими языками.

**Спасибо за внимание!**

**Игътибарыгыз өчен рәхмәт!**