



# Морфологический анализатор кыргызского языка

*Исраилова Нелла Амантаевна,  
Бакасова Пери Султановна*

КГТУ им. И.Раззакова, Кыргызстан



# Морфологическая модель кыргызского языка

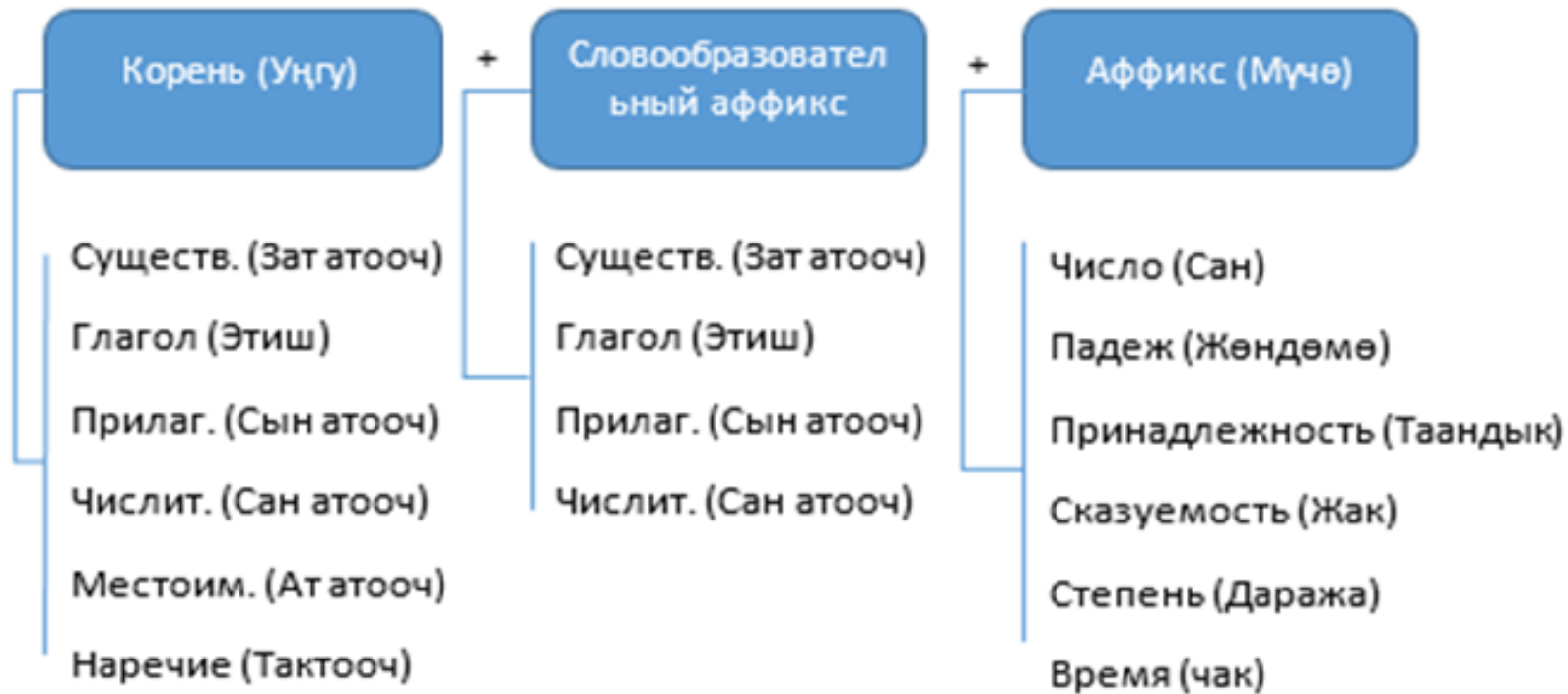
TurkLang 2017

- Кыргызский язык характеризуется строгой последовательностью в присоединении аффиксов к корню (или основе), например, для образования имени существительного к корню сначала присоединяется словообразующий суффикс, затем формообразующие аффиксы: числа, принадлежности, падежей и сказуемости.
- Пропуск одного из этих аффиксов не меняет порядка их присоединения.



# Морфологическая структура слов кыргызского языка

TurkLang 2017





На основании анализа и грамматики кыргызского языка можно выделить следующие основные правила при словоизменении имен существительных:

- Форма множественного числа указывает на расчлененную (не собирательную) множественность и выражается при помощи аффикса **-лар** с фонетическими вариантами. Учитывается гласная последнего слога основы (или корня), а также, если слог закрытый - звонкость, глухость конечной согласной.
- Единственным нарушением закона сингармонизма следует считать присоединение аффикса **-лар, -дар, -тар** после основы, содержащей в последнем слоге гласную у или уу.
- В структуре слова аффиксы принадлежности находятся либо сразу после корня (основы), либо после корня (основы) и аффикса множественного числа, но предшествуют падежным аффиксам.



- Падежные аффиксы могут следовать сразу за корнем, а также за основой, содержащей кроме корня словообразующие аффиксы, аффиксы множественного числа, аффиксы принадлежности.
- При склонении имен существительных, оканчивающихся на сочетания согласных *ск, нк, фт, нг, нд, мн*, после основы добавляются гласные *-ы, и* или *у*, после чего эти слова склоняются как слова с основой, оканчивающейся на гласную.
- Имена существительные в кыргызском языке могут спрягаться
- При морфологическом способе к словообразующей основе присоединяются словообразовательные аффиксы. Аффиксы, присоединяющиеся к основе имен существительных.



- В кыргызском языке окончания имен существительных делятся на четыре вида. Описанные ниже окончания непосредственно будут использоваться в разрабатываемом алгоритме определения основы слова.
- Обозначим через  $E_i$  - следующие множества окончаний, для  $i = 1, 2, 3, 4$ .

$E_i$	Обозначения
$E1$	окончания множественного числа
$E2$	окончания принадлежности
$E3$	падежные окончания
$E4$	окончания сказуемости

TurkLang 2017



Где, С – словообразующий суффикс (-чу, -лык, -дык, -чылык, -лаш, -ма и т.д)



В таблице описаны определения морфемного состава ( $E_i$ , где  $i = 1, 2, 3, 4, 5$ ):

№	Виды окончаний	Окончания
1	Оконания множественного числа - $E1$	<i>-лар, -дар, -тар, -лер, -дер, -тер, -лор, -дор, -тор, -лөр, -төр, -дөр</i>
2	Окончания принадлежности - $E2$	<i>-м, -ым, -им, -ум, -үм, -ң, -ың, -иң, -уң, -үң, -ңыз, -ңиз, -ңуз, -ңүз, -иңиз, -уңуз, -үңүз, -ыңыз, -быз, -биз, -буз, -бүз, -ыбыз, -ибиз, -убуз, -үбүз, -ңар, -ыңар, -ңыздар, -ыңыздар, -сы, -ы</i>
3	Падежные окончания - $E3$	<i>-нын, -дын, -тын, -га, -ка, -ны, -ды, -ты, -да, -та, -дан, -тан</i>
4	Окончания сказуемости – $E4$	<i>-мын, -сың, -сыз, -быз, -сыңар, -сыздар</i>



Для удобства реализации была исследована систематизация окончаний.

Наименование	Обозначение
<i>O</i>	корень слова (основа)
A1	E1 + E2, окончание множественного числа + окончание принадлежности
A2	E1 + E3, окончание множественного числа + падежное окончание
A3	E1 + E4, окончание множественного числа + окончание сказуемости
A4	E2 + E3, окончание принадлежности + падежное окончание
A5	E2 + E4, окончание принадлежности + окончание сказуемости
A6	E3 + E4, падежное окончание + окончание сказуемости





Тогда, существительное во множественном числе - O + E1, П-р, окучуу + лар

Правило	Пример
$O + A1$	окучуу + лар + ыбыз
$O + A1 + E3$	окучуу + лар + ыбыз + дан
$O + A1 + A6$	окучуу + лар + ыбыз + дан + сыздар
$O + A2$	окучуу + лар + дан
$O + A2 + E4$	оюнчу + лар + дан + быз
$O + A3$	окучуу + лар + быз
$O + A4$	окучуу + нуз + дан
$O + A4 + E4$	окучуу + нуз + дан + мын
$O + A5$	окучуу + нуз + мун
$O + A6$	окучуу + дан + мын



# Особенности и закономерности морфологии кыргызского языка

- Структурно-типологическая характеристика кыргызского языка связана с его принадлежностью к агглютинативным языкам. Для описания языков агглютинативного типа применяется набор признаков, учитывающих не только морфологические, но и синтаксические и фонетические особенности.



## Морфологические признаки агглютации:

1. Корень слова – в именительном падеже выступает в чистом виде, таким образом является центром всей парадигмы склонения;
2. Между морфемами четко сохраняется граница;
3. Строгая последовательность присоединения аффиксов;



# Фонетические признаки агглютинации:

1. Наличие сингармонизма;
2. Фиксированное ударение, которое способствует сохранению фонетической целостности слова.



## Синтаксические признаки агглютинации:

- 1. Твердый порядок слов в предложении;
- 2. Определение находится перед определяемым словом;
- 3. Дополняющее слово находится перед дополняемым словом;
- 4. Сказуемое в конце предложения.

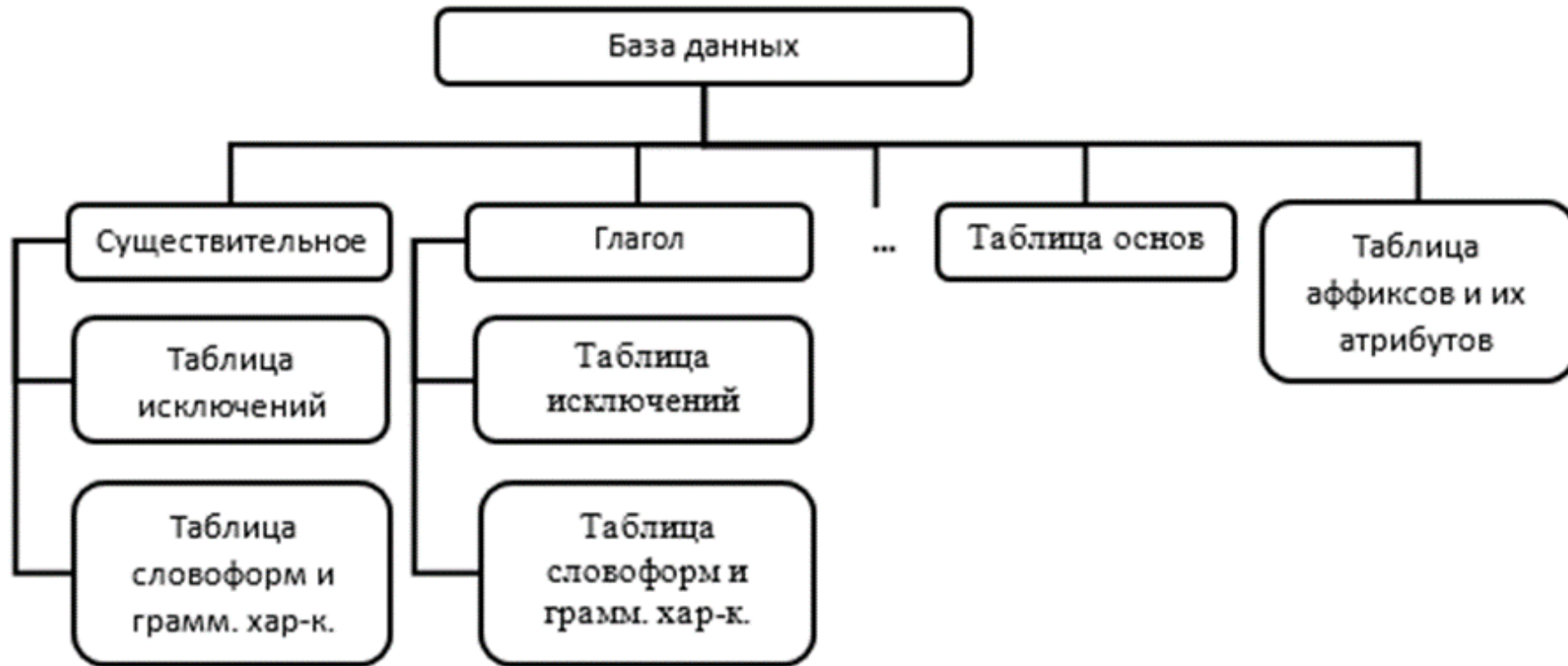


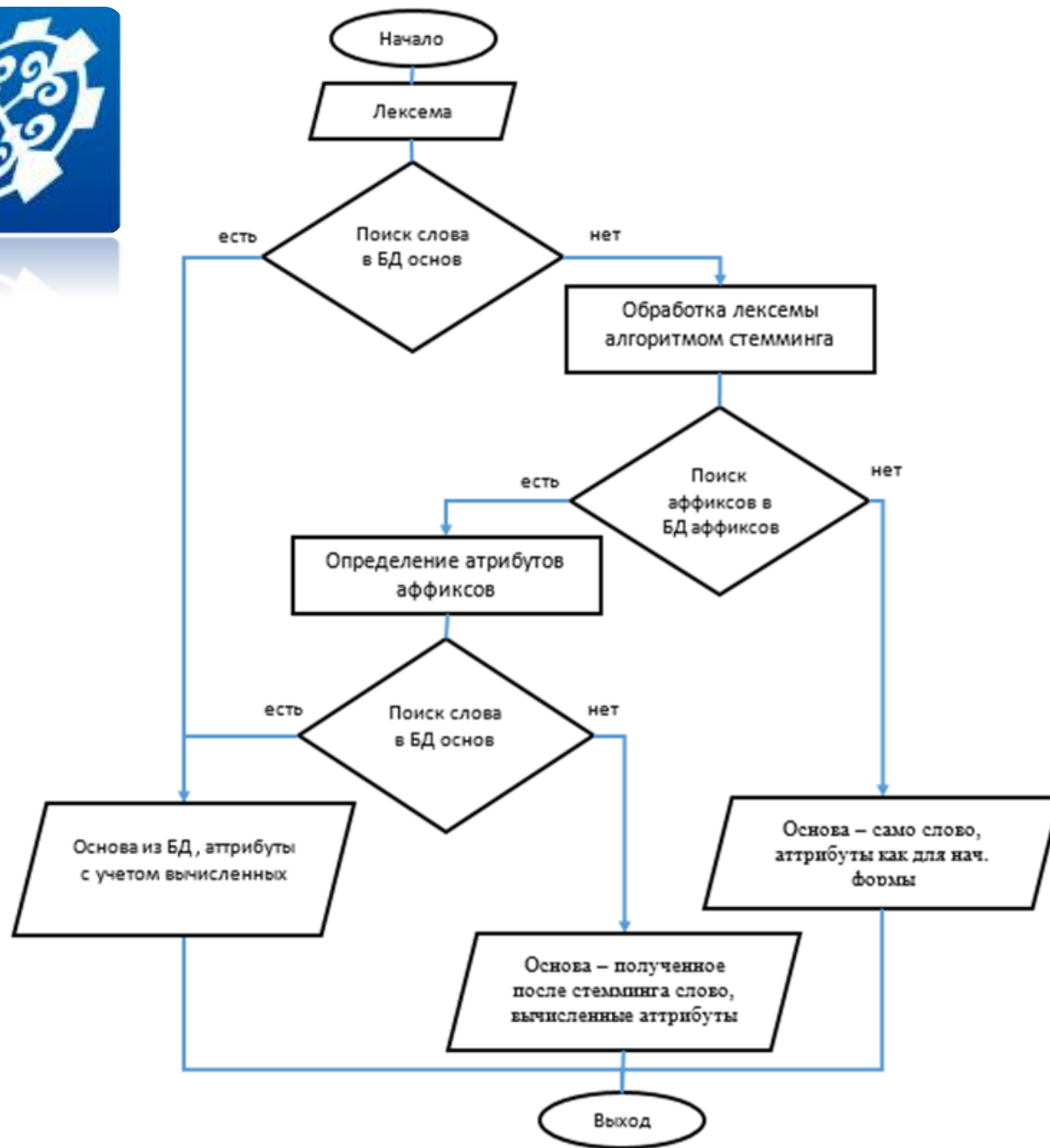
# Структура морфологической базы данных кыргызского языка

- Морфологическая база данных должна содержать всю информацию, необходимую для работы процедур морфологического анализа и синтеза.
- С каждой флексией должен быть связан набор значений ГП, приписываемый основе с данной флексией. Если в морфологической модели учитываются какие-либо типичные особенности словоизменения (например, чередование букв в основе), то информация о них также должна храниться в базе данных.
- Морфологическая БД помимо лексем с регулярным словоизменением должна содержать лексемы с отсутствующими формами, с супплетивными, неизменяемые существительные. Кроме того, БД обязательно должна содержать омонимичные лексемы (с полной и частичной омонимией).



# Структура морфологической базы данных





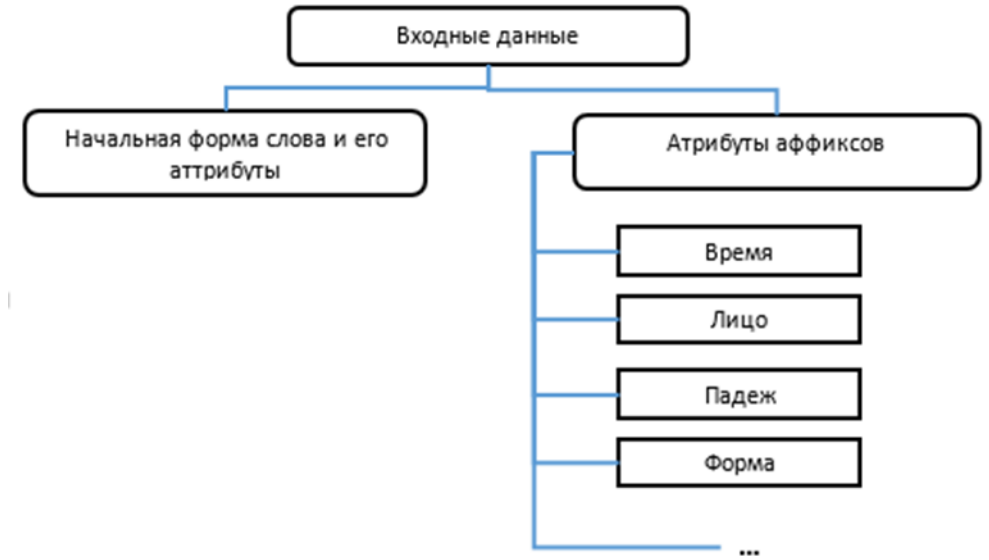
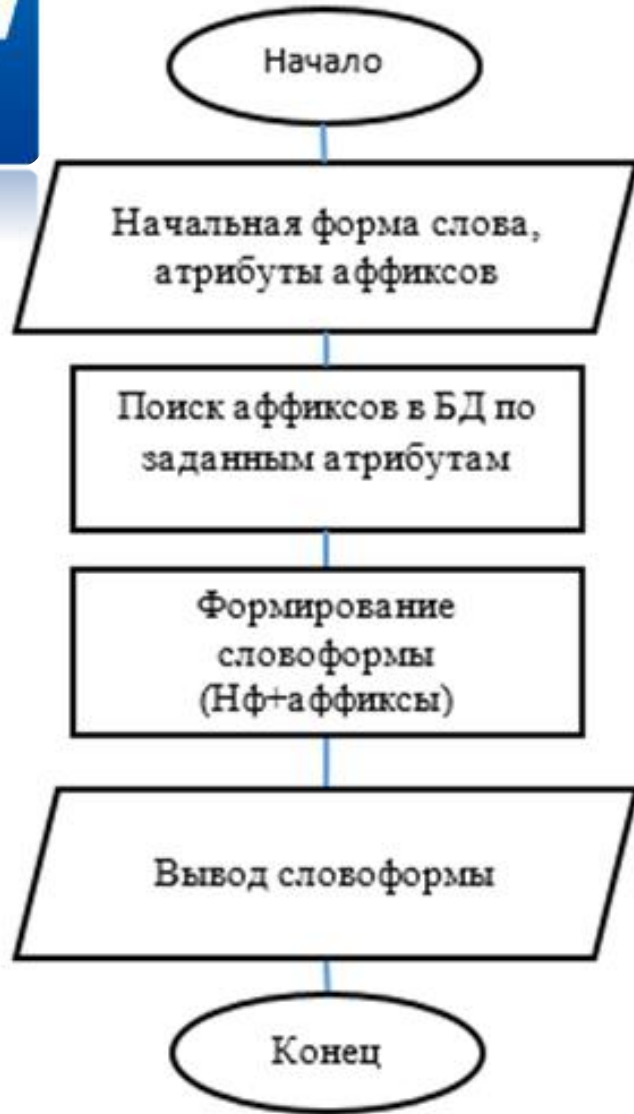
Алгоритм  
морфологического  
анализа кыргызского  
языка





# Алгоритм морфологического синтеза кыргызского языка

- Модель морфологического синтеза разработана на основе грамматики кыргызского языка с учетом семантических правил.
- Входными данными морфологического синтеза являются выходные данные морфологического анализатора (корень слова, полученный в результате морфологического анализа и его атрибуты) и заданные атрибуты аффиксов.



Входные данные морфологического синтеза



# Результаты тестирования

TurkLang 2017

## Морфологиялык анализдөө

Морфологиялык анализдөө үчүн сөздү жазыңыз

 ң ө ү Анализдөө  
Зат атооч Этиш Сын атооч Сан атооч Ат атооч Тактооч

китептерим - зат атооч, Атооч жөндөмө, көптүк сан, таандык мүчөсү менен  
КИТЕП - зат атооч, Атооч жөндөмө, жекелик сан

### СӨЗДҮН ЖӨНДӨЛҮШҮ

Жөндөмө	Жекелик сан	Көптүк сан
Атооч	китеп	китептер
Илик	китептин	китептердин
Барыш	китепке	китептерге
Табыш	китепти	китептерди
Жатыш	китепте	китептерде
Чыгыш	китептен	китептерден

### ТААНДЫК МҮЧӨСҮ МЕНЕН

Жак	Жекелик сан	Көптүк сан
Биринчи	китебим	китептерим
Экинчи	китебиң	китептериң
Экинчи (сыйлуу)	китебиңиз	китептериңиз
Үчүнчү	китеби	китептери

Для проверки правильности определения основы слов были протестированы существительные и глаголы с разными вариациями суффиксов и окончаний.



# Примеры результатов тестирования слов.

Название словоформы	Аффиксы	Количество аффиксов	Часть речи
Тегирмен	<i>чи + лер + ди + ки + лер + ден + сиз + дер + би</i>	9	Сущ.
Табышмак	<i>тар + ыңыз + дар + дан + бы</i>	5	Сущ.
Кам	<i>сыз + дан + дыр + ыл + ба + ган + дык + тан</i>	8	Сущ.
Жаз	<i>дыр + а + сыңар + бы</i>	4	Глаг.
Оку	<i>ган + сыңар + бы</i>	3	Глаг.
Бекит	<i>кен + биз</i>	2	Глаг.



## Морфологиялык анализдөө

Морфологиялык анализдөө үчүн сөздү жазыңыз

Ң

Ө

Ү

Анализдөө

Зат атооч

Этиш

Сын атооч

Сан атооч

Ат атооч

Тактооч

тегирменчилерденсиздерби - зат атооч, Чыгыш жөндөмө, көптүк сан, суроо түрү, баяндоочтук категориясында  
ТЕГИРМЕНЧИ - зат атооч, Атооч жөндөмө, жекелик сан

### СӨЗДҮН ЖӨНДӨЛҮШҮ

Жөндөмө	Жекелик сан	Көптүк сан
Атооч	тегирменчи	тегирменчилер
Илик	тегирменчинин	тегирменчилердин
Барыш	тегирменчиге	тегирменчиерге
Табыш	тегирменчини	тегирменчилерди
Жатыш	тегирменчиде	тегирменчилерде
Чыгыш	тегирменчиден	тегирменчилерден

### ТААНДЫК МҮЧӨСҮ МЕНЕН

Жак	Жекелик сан	Көптүк сан
Биринчи	тегирменчим	тегирменчибиз
Экинчи	тегирменчиң	тегирменчиңер
Экинчи (сыйлуу)	тегирменчиңиз	тегирменчиңиздер
Үчүнчү	тегирменчиси	тегирменчиси

### ТААНДЫК МҮЧӨСҮ МЕНЕН (Көптүк сан)

Жак	Жекелик сан	Көптүк сан
Биринчи	тегирменчилерм	тегирменчилерибиз
Экинчи	тегирменчилерң	тегирменчилериңер
Экинчи (сыйлуу)	тегирменчилериңиз	тегирменчилериңиздер
Үчүнчү	тегирменчилери	тегирменчилери

## Морфологиялык анализдөө

Морфологиялык анализдөө үчүн сөздү жазыңыз

Ң

Ө

Ү

Анализдөө

Зат атооч

Этиш

Сын атооч

Сан атооч

Ат атооч

Тактооч

окугансыңарбы - этиш, Белгисиз өткөн чак, көптүк сан, суроо түрү  
ОКУ - этиш

### АЙКЫН КЕЛЕР ЧАК

Жак	Жекелик сан	Көптүк сан
Биринчи	окуймун	окуйбуз
Экинчи	окуйсун	окуйсунар
Экинчи (сыйлуу)	окуйсуз	окуйсuzдар
Үчүнчү	окуйт	окушат

### АРСАР КЕЛЕР ЧАК

Жак	Жекелик сан	Көптүк сан
Биринчи	окурмун	окурбуз
Экинчи	окурсун	окурсунар
Экинчи (сыйлуу)	окурсuz	окурсuzдар
Үчүнчү	окур	окушат



# Заключение

TurkLang 2017

- В процессе реализации морфологического анализатора кыргызского языка разработаны модули морфологического анализа и синтеза.
- Проведено тестирование, в результате которого морфологический анализатор показал устойчивую работу на множестве слов с различными морфохарактеристиками.



**TurkLang** 2017

Спасибо за внимание!