

Как помочь языку выжить в современном мире с помощью языковых технологий?

Трун Тростерюд
Университет Тромсё
Норвегия



Джек Рютер
Фонд Конне
Финляндия





▼ По-русски

▫ [Start page](#)

► [Фон](#)

► [Интерактивные программы](#)

► [Словари](#)

► [Тексты и списки](#)

► [Другие языки](#)

► [UnivOaһra-
Vorkshop 2012](#)

Добро пожаловать на страницы Giellatekno - саамские языковые технологии

[Davvi](#) [Åarjel](#) [Norsk](#) [English](#) [Suomeksi](#) [Русский](#)

Ресурсы, которые существуют для разные языки

- **Саамские языки:** [северосаамский](#), [луле-саамский](#), [южносаамский](#) // [пите-саамский](#), [инари-саамский](#), [колтта-саамский](#), [кильдин-саамский](#).
- **Остальные уральские языки:** [вепсский](#), [горномарийский](#), [ижорский](#), [квенский](#), [КОМИ](#), [ливский](#), [ливвиковсий](#), [лугово-восточный марийский](#), [мокшанский](#), [нганасанский](#), [удмуртский](#), [финский](#), [хантыйский](#), [эрзянский](#).
- **Остальные языки:** [бурятский](#), [гренландский](#), [инупиак](#), [корнский](#), [оджибве](#), [русский](#), [фарерский](#),

Уральские языки

F Finnisch-Ugrisch

FO Ostseefinnisch

- FO1 Finnisch
- FO2 Karelisch
- FO3 Wepsisch
- FO4 Ischorisch
- FO5 Estnisch
- FO6 Wotisch
- FO7 Liwisch

FS Samische Sprachen

- FS1 Westsamisch
- FS2 Zentralsamisch
- FS3 Ostsamisch

FU Ugrisch

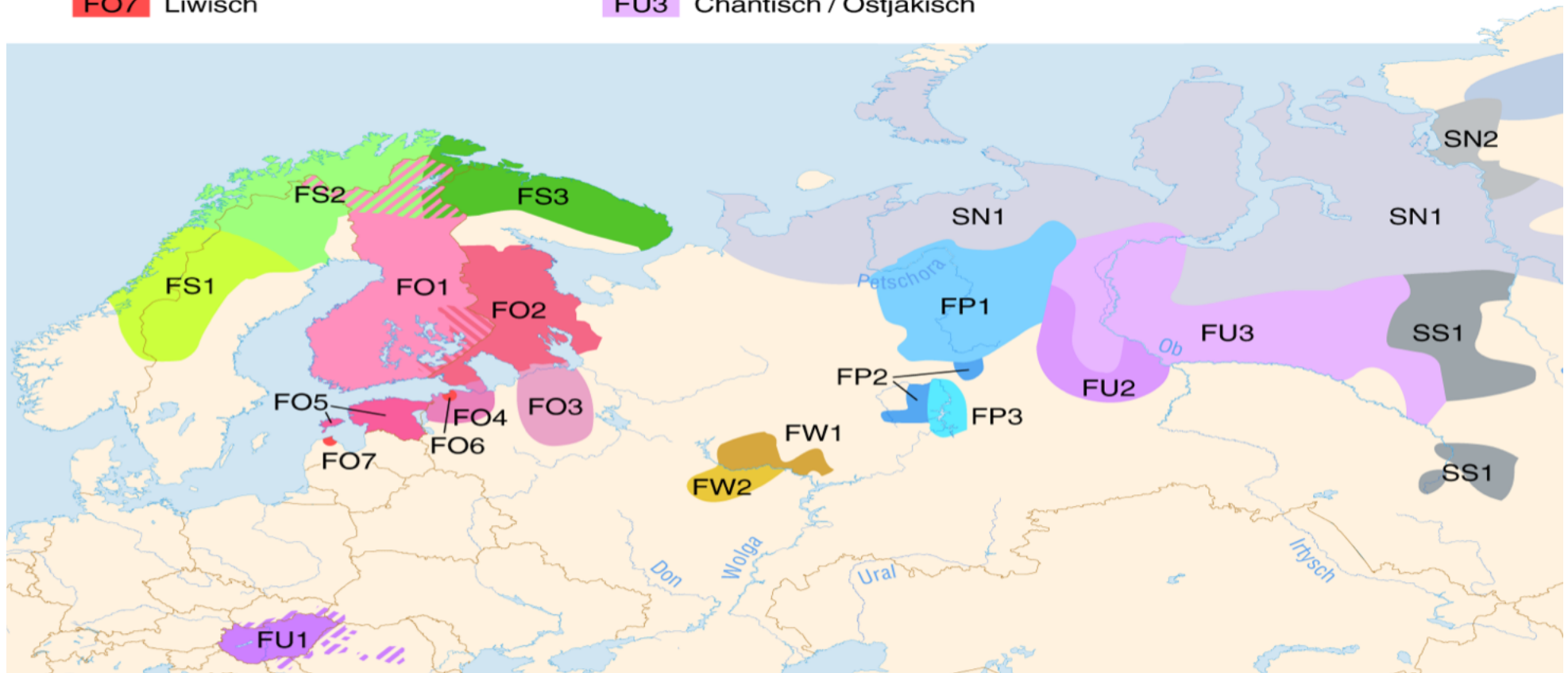
- FU1 Ungarisch
- FU2 Mansisch / Wogulisch
- FU3 Chantisch / Ostjakisch

FP Finnisch-Permisch

- FP1 Komi-Syrjänisch
- FP2 Komi-Permjakisch
- FP3 Udmurtisch / Wodjakisch

FW Finnisch-Wolgaisch

- FW1 Tscheremissisch
- FW2 Mordwinisch



Языковая политика 1955 года

- Северная Норвегия
 - Детям запрещалось разговаривать на языке Саами в интернатах

Языковая политика 1955 года

- Северная Норвегия
 - Детям запрещалось разговаривать на языке Саами в интернатах
- Южная Африка
 - Дети получали образование на их родном языке
 - ... но у них не было доступа к английскому языку

Языковая политика 1955 года

- Северная Норвегия
 - Детям запрещалось разговаривать на языке Саами в интернатах
- Южная Африка
 - Дети получали образование на их родном языке
 - ... но у них не было доступа к английскому языку
- Марий Эл
 - Дети получали основное образование на их родном языке
 - ... И так же они изучали русский язык

Норвегия с 1980 года

- Новая языковая политика для:
 - северосаамского языка (20000)
 - луле-саамского языка (500)
 - южносаамского языка (300)
- Право использовать родной язык и получать официальную информацию на саамских языках
- Право получать образование на саамских языках



Radiologisk avd. venterom
Radiologijja osd. vuordinlatnja

BRANNDØR
holdes lukket

← Røntgen

ráđđehus.no

Stoltenberga nubbi ráđđehus

Departemeanttat

Fáddá A-Å

Neahttakarta

Hjelp

Oktavuotta

Bokmål | Nynorsk | Sámeigiella | English



RÁÐÐEHUS.NO

Dieđut ráđđehusas ja departemeanttain



Oza olles ráđđehusa.no:as

Oza

Čálánhápmi a a a

Norgga ekonomii ja ovdána bures

- Norgga ekonomii ja lea ceavzán bures ruhtadanroasu váttisvuodaid maŋŋá. Árvoráhkadeapmi lea dál lassánan ovcci jahkenjealjádasa maŋŋálagaid. Barggolašvuotta lea buoret go 2008 buoremus áiggi, ja bargguhisvuotta bissu vuollegas dásis. Ráđđehusa reviderejuvvon bušehta árvalusain mii láchit dilálašvuodaid dasa, ahte norgga ekonomii ja ovdána bures ain viidáseappotge, ruhtadanministtar Sigbjørn Johnsen dadjala. **(Ruhtadandepartemeanta, 15.5.2012)**

Vearro- ja divatrilevdadusat >

Elektrovnnalaš vearrokoarttat >

EDAG álkidahtta ealáhusaid barggu >

Norga áigu dahkat loatnasoahpamuša IMF:in >



2012 dárkkistuvvon nátionála bušehta



Galledii Sámedikki

Stáhtačáli Eli Blakstad galledii otna Sámediggeráđi Vibeke Larsena. Čoahkkima fáddán lei plánejuvvon elfápmollinnjá gaskal Ofuohta, Báhcavuona ja Hármafuohta. Čoahkkima sámediggeráđi departemeanta



Ráđđehuskvartála boahhteáigi

Ráđđehus lea mearridan čohkket departemeantaid lahka ja birra dálá ráđđehuskvartála. Mo dálina áššiin galgá bargat viidáseappot, čielgá maŋŋil go ráđđehus lea árdáhan viiddit, ebas preassasas go lea gávdan



Ráđđehusa bargu

Ráđđehusa sátni geavahuvvo beaivválaččat guovtti áddejumis; olles ráđđehus čoahkis Gonagasa jodiheami vuolde stáhtaráđđeoahkkimis, ja olles ráđđehus čoahkis stáhtaráđđeoahkkimis, ja olles ráđđehus čoahkis

Россия

Языковое планирование

- Новые литературные языки
 - С новыми буквами!
- Образование на родном языке
 - Книги
 - Национальные газеты
- Новые административные границы



составитель карты Никола
Карта не учитывает администра
мые Россией на территориях до
ровой войны, поскольку террито
границы между ними должны
договорами.



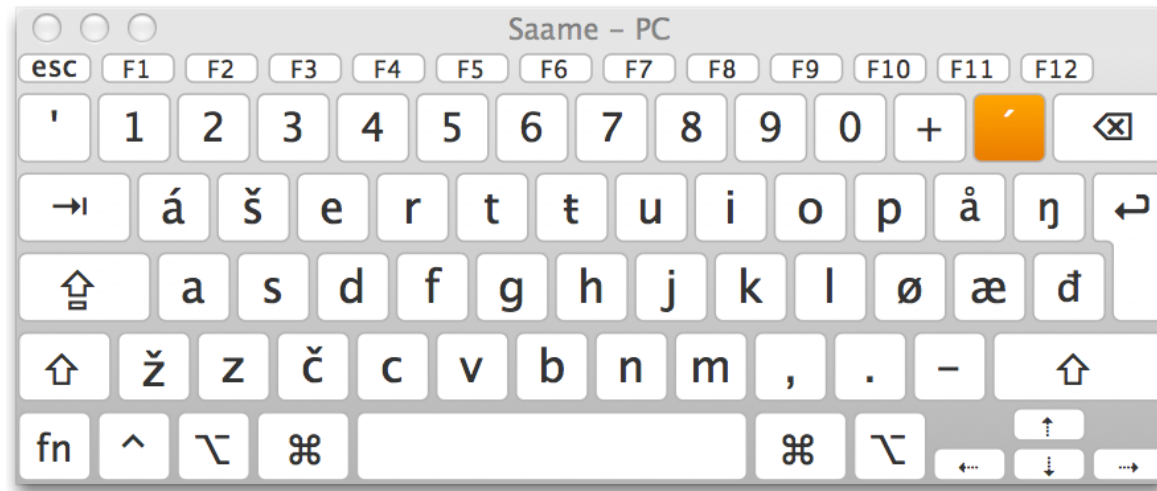
Сегодня: новые трудности

- Мы движемся в век цифровой обработки текстов
 - И мы хотим, чтобы наш язык двигался вместе с нами

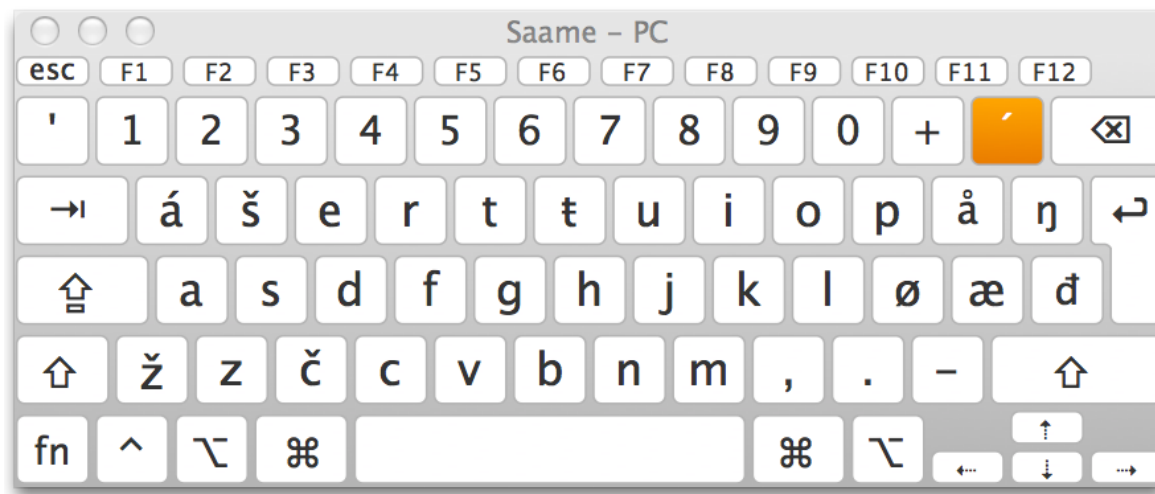
Необходимые программы

- *Клавиатуры с национальными буквами*
- Программы по проверке *правописания*
- *Словари*, распознающие измененные формы слова
- *Машинный перевод*
- *Информационные поисковики*
- *Звук*: распознавание и воспроизведение речи

Клавиатуры

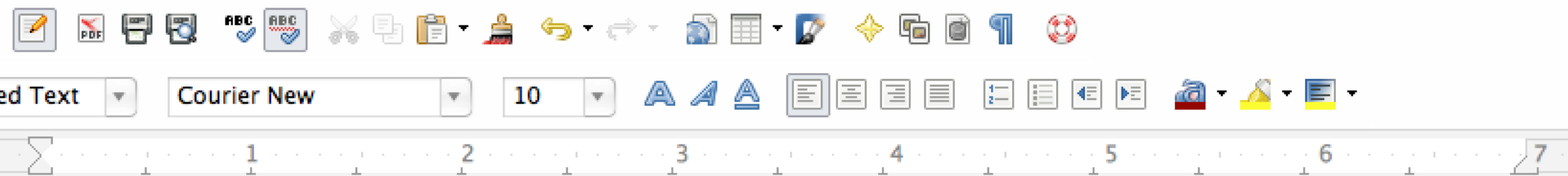


Клавиатуры



- В «Википедии» на коми языке при наборе букв *i*, *ö*, 66% набирались Латинницей, и только 33% - Кириллицей
- (2/3 текстов написаны без использование раскладки клавиатуры на коми)

Программы по проверке правописания



өшйыліс Латкинлөн сизимөд номера вугыръяс йылө. А нидзувсө яндысьтөма джыбөдісны.

Но мөдлаө Латкинлы эз мунсы. Шыр кыйөдысь кань моз куткырасис дебаркадер бөжын да кыйкнитавліс кыр йывлань. И быдөн пөсьмунлі, кор аддзис кык нылөс, горув посниа да тэрыба тувчалысьөс: өтиыс вөлі почта новлөдлысь, а мөдыс – Тамара.

Печора карысь кывтысь теплоходыс мыйлакө сөрмөдчис, эз на кыв ни тыдав, да нывъяслы, кылө, дышөдіс дебаркадер вылын өти- даың варгынысө, найө кутісны ветлөдлыны гөгөрыс. Водзджыксө кытшовтісны нырөдыс, а сэсся тотшсьөдісны и бөжланьыс, Латкин дінө.

- Абу тай дыш, – эльтана шыасис Тамара.
- А мый нө таянлы, горт олөм оз төдны-а, – вочавидзис почта новлөдлысь.
- Вола пестө поткөдлыны, – бергөдчис да шуис Ааткин, коді таөдз мышка на вөлі.
- Миян важөн нин чипасын.

Словари, понимающие измененные формы слова

Báhčaveaijohka | Industriija | Okta johka – golbma stáhta | Geavvnis | Čálalaš meattábeassanlohk
gorži | Maŋemus mátkki | Dálá elfápmu



BÁHČAVEAIJOHKA

Báhčaveaijohka lea odne rádjain Norgga ja Ruošša gaskkas. Johka vuolgá mearihis unnit jogážat mat leat vuodđun etnui mii golga Báhčaveaivutnii Girkonjárgga bol **jogas** ja Norggas leat guokte.

— Vuosttaš Digisánit —

subst. → **johka**

bekk, elv

Analyse: sg. lok.



Машинный перевод 1

- Что написали обо мне в «Коми Му»?

Машинный перевод 2

Коми Республика официальной портал

Русская версия | English version

Каналан Сöвет


Веськöдлан котыр

Олöмö пöртысь власьт органьяс

Выльторьяс

17.10.2012 17:39



«Коми Республика да Индия костын эмöсь ыджыд позянлуныяс экономика да культура юкöньясын сöвмöдны öтвылысь удж öта-мöдлы пöльза вайöмөн» [ фото]

Юöр сетан технологиясын да промышленностьын öтувья балаяс збыльмöдан позянлуныяс йылысь, а сіджö культура да велöдан да наука йитöдьяс йылысь сёрнитисны Сыктывкарын удж серти паныдасьлігөн Коми Республикаса Юралысьöс вежысь Александр Буров да Санкт-Петербурлын Индияса генеральной консул господин Вишвас Сапкал.

 Поиск

Гражданалөн шыöдчöмьяс

"Индöд-тшöктöмьяс..." журнал

Фотосерпаскуд

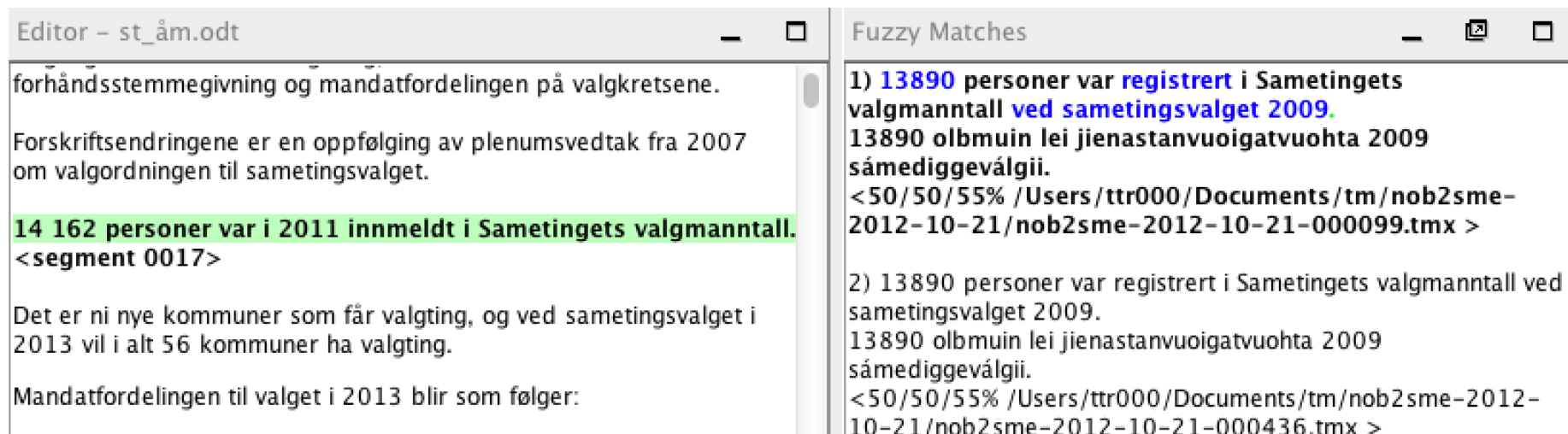
Воча йитöд

Сайт тэчас



ПРИЕМНАЯ ПРЕЗИДЕНТА

Память переводов



- Программа имеет доступ к уже переведённым ранее материалам
- Программа предлагает уже переведенные предложения как альтернативу переводимому тексту

Научное сотрудничество

- Партнеры

- Межрегиональная лаборатория информационной поддержки функционирования финно-угорских языков в Сыктывкаре
- «Giellatekno» в Университете Тромсё
- Языковая программ Фонда Конне 2012-2016

Научное сотрудничество

- Партнеры

- Межрегиональная лаборатория информационной поддержки функционирования финно-угорских языков в Сыктывкаре
- «Giellatekno» в Университете Тромсё
- Языковая программ Фонда Конне 2012-2016

- 2 главные цели:

- Создание программ для эффективной обработки текстов на коми-зырянском языке
- Создать научную основу для исследований коми-зырянского языка и родственных ему языков

Межрегиональная лаборатория информационной поддержки функционирования финно-угорских языков в Сыктывкаре

- Работают с коми-зырянским и коми-пермяцким
- Система проверки правописания
- Словари

Языковая программа Фонда Конне

- Документация
 - Сбор текстов, составление словарей
- Создание языковых анализаторов
 - вепсский, горномарийский, ижорский, коми, ливский, ливвиковский, лугово-восточный марийский, мокшанский, ненецкий, нганасанский, удмуртский, хантыйский, эрзянский.

«Giellatekno»

- Сфокусированы над саамскими языками
 - Так же работают над другими северными языками
- Заинтересованы в:
 - Грамматическом анализе
 - Интерактивных образовательных программах

Наши общие фонды

- Грамматические модели коми и других языков
- Без этих моделей невозможно создать практические программы

```
echo << defining the rules >>
```

```
! The famous L/V
```

```
! -----
```

```
define LtoV [ л -> в || _ Flags* [ .#. | %> Cns ] ] ;
```

```
! These words are underlyingly л, and changed to в.
```

```
!€ ньл>ыс:ньлыс
```

```
!€ ньл:ньв
```

```
!€ ньл>сö:ньвсö
```

```
! Paragogic consonants
```

```
! -----
```

```
! These are consonants that are visible in front of vowel-initial suffixes.
```

```
define mDeletion м -> 0 || н _ [ .#. | %> Cns ] ;
```

```
!€ синм>ыс:синмыс
```

```
!€ синм:син
```

```
!€ синм>сö:синсö
```

LEXICON CASEPOSSLEX

DERIVEDADJ ; ! Deriving prenominal adjectives, республикаса
CASEPOSSLEX-afterDerivation ;

LEXICON CASEPOSSLEX-afterDerivation

+Nom: K ;
!+ZeroAcc: K ;
+Gen:%>лөн K ;
+Abl:%>лысь K ;
+Dat:%>лы K ;
+Acc:%>öс K ;
+Instr:%>өн K ;
+Com:%>көд K ;
+Car:%>төг K ;
+Cns:%>ла K ;
+Comp:%>ся K ;
+Ine:%>ын SPAT-COMPARATIVE ;
+Ela:%>ысь SPAT-COMPARATIVE ;
+Ill:%>ö IllCompar ;
+Apr:%>лань SPAT-COMPARATIVE ;
+Egr:%>сянь SPAT-COMPARATIVE ;
+Prl:%>өд K ;
+Tra:%>ті K ;
+Ter:%>өдз K ;
+Apr+Ine:%>ланьын SPAT-COMPARATIVE ; ! New cases in last grammar
+Apr+Ela:%>ланьысь SPAT-COMPARATIVE ; !
+Apr+Ill:%>ланьö IllCompar ; !
+Apr+Egr:%>ланьсянь SPAT-COMPARATIVE ; !
+Apr+Prl:%>ланьөд K ; !
+Apr+Tra:%>ланьті K ; !
+Apr+Ter:%>ланьөдз K ; ! Tähän asti

Коми Республика да Индия костын эмось ыджыд позянлуньяс экономика да культура юкөнъясын сөвмөдны өтвылысь удж өта-мөдлы польза вайомон.

Дать все формы слов

Устранить неоднозначность

Отправить форму

Сбросить форму

utf-8 latin

"<Коми>"

"КОМИ" **N** Sg Nom

"<Республика>"

"республика" **N** Sg Nom

"<да>"

"да" **CC**

"<Индия>"

"Индия" **N** Prop Plc Src/F Sg Nom

"<КОСТЫН>"

"КОСТЫН" **PO** Sg Ine

"КОСТ" **N** Sg Ine

"<ЭМӨСЬ>"

"ЭМ" **A** Pred Pl

"<ЫДЖЫД>"

"ЫДЖЫД" **СmpTest**

"<ПОЗЯНЛУНЬЯС>"

"ПОЗЯНЛУН" **N** Pl Nom

Выводы

- Без представленных здесь инструментов нельзя использовать язык в цифровых форматах

Выводы

- Без представленных здесь инструментов нельзя использовать язык в цифровых форматах
- Административные структуры России основаны на этнических составляющих, где язык играет ведущую роль

Выводы

- Без представленных здесь инструментов нельзя использовать язык в цифровых форматах
- Административные структуры России основаны на этнических составляющих, где язык играет ведущую роль
- Научная и прикладная деятельность должны производиться совместно

Выводы

- Без представленных здесь инструментов нельзя использовать язык в цифровых форматах
- Административные структуры России основаны на этнических составляющих, где язык играет ведущую роль
- Научная и прикладная деятельность должны производиться совместно
- Сотрудничество и обмен ресурсами являются ключевыми факторами

**Спасибо за ваше
внимание!**