

Джек М. Рютер (Хельсинки)

Хельсинкиса университетын кыв туялысь

Ижкарын перымса кывъяс симпозиум вылын лыддьомтор

Финляндияса академия отсөгөн Хельсинкиса университетын вöчöны ичöт войтырлөн кывъяса гижöд чукöрлысь проект, кодi лöб Войвыв Муясса öтувъя проектлөн юкöнөн. Öти боксянь, Хельсинкиса университетын эмöсь перымса кывъяс туялысь йöз. Мöд боксянь, танi эмöсь и йöз, кодъяс уджалöны тöдан машиналысь кыв туялöмын. Сидзи, лингвистика кафедра-са профессор Киммо Коскенниемилöн UNIX вылын лöсьöдöм TWOL уджтасъяс отсөгөн гижöд чукöр проектн лои позянлун артмöдны и кывчукöр юкөн.

Öни аслам проектн вöчöма нин вель уна – кывчукöрын дас сюрс сайö кыв да кыв артмöдан 95 ордым, кодъяс серти петкöдчö кывсикасьяслөн артмöм да вежласьöм. Воис кад, кор позьö лыддьöдлыны компьютерса коми кывкудлысь сьöкыдлуньяс.

Тöдан машинаö медводз колö пыртны газет-небöг гижöдъяс. Кöть öнiя кадö газет-небöгтö лэдзöны тöдан машинаөн, UNIX-ын уджалiгөн ковмис и материалсö мöд ногөн пуктавны, медым кужны лыддьынысö. Та могысь ковмис кириллича шыпасъяс пыдди лöсьöдны торъя анбур. Эска, мый матысса кадö позьö кириллича шыпасъяс лыддьыны веськыда UNIX вывсьыс.

Өнi верман лыддыны сөмын куим лыдакывья пасьяс, шуам: а = \340, б = \341, в = \342 да с. в.

Отарсянь кӧ, омӧль, мӧдарсянь кӧ, бур. Кириллича пыдди артмӧдӧм системасӧ позьӧ лыддыны мушар пасьята быдлаын: а = а, б = в, в = в, г = г да с. в. Эмӧсь и сэтшӧм пасьяс, кодьясӧс ковмис мӧд ногӧн вӧчны: ю = u1, ӕ = o1, я = a1, i = i2, э = e2, ы = i3, ӧ = e3, ж = z4, ч = c4, ш = s4, щ = s5, ь = w, ь = y.

Водзӧ сета материалсӧ кирилличаӧн, тӧдса гижанногсӧ кок-ниджык лыддынысӧ.

Кӧть тӧдан машинаса материалсӧ босьтан веськыда небӧг-газет лӧдзанинысь, колӧ тӧдны, кызди гижны кывьясӧс.

1-кӧ, эм Молодцовлӧн гижанног, код подув вылын кывьяс-сӧ позьӧ гижны шуанног серти. Оз ков гижны *область* роч ног, позьӧ и коми ногӧн, *область* артмас, мый Молодцов серти роч кирилличаӧн, *морковь* водзтӧ моз лоас *морков*. Литератураын эм и *гӧрд вуж*. Кывчукӧрад колӧ пыртны и выль формаяс.

2-кӧ, коми газетьясын, рекламаясын эм уна материал, кодӧс комиӧдӧма роч да мукӧд кыв вылысь, мый вӧсна паныдась-лӧ интӧм, абу стӧч коми кыв: *экономикаса министерство* (колӧ: *экономика министерство*), *челядь сӧянiн* (колӧ: *челядьлӧн сӧ-янiн*).

3-кӧ, коми кывйын эмӧсь шыяс, кодьяс вежласьӧны либӧ ты-далӧны сӧмын сӧк, кор на бӧрын мунысь вежлӧг либӧ мукӧд си-кас кывйив заводитчӧ гора шыӧн: *вӧв + ыс : вӧлыс; кыв + ын : кывйын; син + ас : синмас* да с. в.

Тадзи кывчукӧр гижигӧн колӧ кутны тӧд вылын, кутшӧм веж-ласьӧмьяс паныдасьлӧны. Сьӧкылдуньяс лӧбны *к* да *л* шыяскӧд.

Выль кывьясас найӧ озджык вежласьны, шуам *с* бӧрын: *воск : воскыс = вос : воскыс = вос : восыс*. Мукӧддырйи кывпомса *л* либӧ *в* шыяс кольӧны вежласьтӧг, шуам: *зал : залыс* либӧ *ов : овыс*; видз. *ов : оvas = ов : олас*. Та вӧсна и колӧ содтыны кык пас: *б* лыдпасӧн кӧ, *кб* да *лб*.

4-кӧ, шы вежласьӧмьясысь кындзи эмӧсь и мукӧд гижан сьӧ-кылдуньяс, шуам кык да куим юкӧна кывьяс, кодьясӧс ӧтувтан-визьӧн гижӧны: *мам-бать : мам-батьлы = мамлы-батьлы*.

Корсюрӧ кык юкӧна кывйын кывйив содӧ бӧрья юкӧн по-мас. Тайӧ тшӧкылджыка паныдасьлӧ коми кывйын. Корсюрӧ кык юкӧна кывйын содтӧма кывйивьяс воддза и бӧрья юкӧн помас, костаныс пуктӧма ӧтувтан визь.

Видзѡдлам коми гижысь Иван Тороповдысь кыв, кодi уна нога да аслыспѡлѡс:

1 *ыши-пошиасьны* (Торопов 1995, 10 лб.)

2 *козья-пипуа-пожѡма вѡрын* (9 лб.)

3 *венѡ-кошьѡ* (10 лб.)

4 *вердысь-удысьясѡн* (17 лб.)

1 колѡ пыртны кывчукѡрѡ. *Ыши-* оз овлы торья кывйѡн, став татшѡм кык юкѡна кывсѡ пуктам аслас местаѡ.

2 оз ков пыртны кывчукѡрѡ. Пыртам кѡ, позьѡ эськѡ пуктыны нѡльѡд юкѡн, витѡдѡс тшѡтш – дженьыдджыка +а- = 'и': вѡр, кѡнi эм и коз, и пипу, и пожѡм и с. в.

3-ыс сьѡкыдлуна. Кыкнан кывйыс овлѡны торйѡн. Кыкнан юкѡныс вежласьѡны, мый тѡдан машинаыдлы вывтi сьѡкыд гѡ-гѡрвоны. Сѡмын контекст серти тѡдан, эм-ѡ налѡн торья вежѡртас.

4 пыр кокниджык, кор кык юкѡна кадакывсьыд артмѡ эмакыв, сы вѡсна мый танi сѡмын веськыд юкѡныс артмѡдѡма. Колѡ-ѡ 3 да 4 пыртны кывчукѡрѡ, тѡдѡ сѡмын велѡдчѡм коми кывья морт. Танi колѡ кутны тѡд вылын, мый тѡдан машинаыд сѡмын компьютер – 'артасян'.

Водзѡ коми кыв правилѡяс йывсьыс, кодьясѡс лѡсьѡдѡма тѡдан машиналы:

шыпасчукѡр: а б в г д е ё ж з и i й к кб:к л лб:л м н о ѡ пр ст у ф х ц ч ш щ ъ ы ь э ю я #:0 %- %_:0 ;

диакритическѡй пас: %+ ;

торья чукѡрьяс

! став согласнѡй

C = б в г д ж з й клмнпрстфхцчшщъь;

! ь-(небыд пас) кындзи став согласнѡй

C1 = б в г д ж з й клмнпрстфхцчшщъь;

! став согласнѡй, кодьяс бѡрын вермас лоны сѡмын и

C2 = б в г ж й кмпрфхчшщ;

! гора шыпасьяс

V1 = а е ё и i о ѡ у ы э ю я;

! шыпасьяс

Segm = б в г д ж з й клмнпрстфхцчшщъ

ь

а е ё и = о ѡ у ы э ю я;

! кыв су (ѡтвѡтѡм чукѡрьясын, торйѡн гижѡм кывьясын)

WB = # %- %_;

Гижан правилёяс

"л:в" л:в <=> V1 _ [WB: | %+: C1:];

"м:0" м:0 <=> н _ [WB: | %+: C1:];

"к:0" к:0 <=> [с | ш | с ь | ч] _ [WB: | %+: C1:];

"т:0" т:0 <=> п _ [WB: | %+: C1:];

"ы:0" ы:0 <=> V1: + C:* _ %+: V1;

"л:0" л:0 <=> л _ ь [WB: | %+: C1:];

"д:0" д:0 <=> д _ [ь | з] [WB: | %+: C1:];

"й:0" й:0 <=> V1 _ [б | в | г | д | ж | з | к | м | н | п | р | т] _ [WB: | %+: C1:];

"ьяс:ьяс" ь:ь <=> [д: з: | :ч :0*] %+: _ я с;

"ьяс:яс" ь:0 <=> [V1 (й:) | ь] %+: _ я с;

"ь+а:00я, й+а:ъ0я да с. в (1)"

X:Y <=> [[:з | :ч] | \:д :з] [ь: | й:] V1:0* %+: _ ;

X (а о у)

Y (я ё ю);

"ь+а:00я да с. в. (2)"

ь:0 <=> _ V1:0* %+: _ [а: | о: | у: | и:];

"й+а:ъ0я да с. в. (2)"

й:ъ <=> C1 _ V1:0* %+: [а: | о: | у:];

"й+а:00я да с. в. (2)"

й:0 <=> V1 _ V1:0* %+: [а: | о: | у:];

"i:i" i:i <=> [V1 | C2 | ь: | д з] :0* %+: _ ;

Prof. Kimmo Koskenniemi-лөн уджтасьяс йылысь:

AUTOMATA komitwol.aut

43 Characters in the Alphabet (102)

15 Automata Read In: 979 Transitions (9000)

113 Rows (10000) 124 Columns (5000)

59 Aligned Pairs (155)

DICTIONARY komi.lex

30098 Branches (240000)

96 Sublexicons (1000)

9389 Entries (65000)

85177 string characters (120000)

Кык ащёр кыв, налөн вежёртас, аргмёдём, английскöй вы-
лö вуджödём:

вöв

"<в0в>"

"в0л#" V ENG [be] V CONNEG SG3

"в0л#" V ENG [be] V CONNEG SG2

"в0л#" V ENG [be] V CONNEG SG1

"в0л#" N ENG [horse] NOM SG

в0вт0мтор

"<в0вт0мтор>"

"в0л+т0м+торй#" V ENG [be] V CAR1 DA-торй N
[thing] NOM SG

"в0л+т0м+торй#" V ENG [be] V CAR1 DN-торй N
[a little] NOM SG

"в0л+т0м+торй#" N ENG [horse] V CAR DA-торй N
[thing] NOM SG

"в0л+т0м+торй#" N ENG [horse] V CAR DN-торй N
[a little] NOM SG

Мый могысь позь0 в0дитчыны татш0м кывчук0р0н?

0т-к0, сй0с позь0 используйтны кызди коми-английск0й-фин кывкуд.

М0д-к0, т0дан машинаса материал подув вылын позь0 л0сь0дны коми кывлысь ст0ч гижанног петк0длысь уджтас, мый сетас отс0г газет-журналын уджалысьяслы, гижысьяслы да уна мук0д й0злы, кодьясл0н эм та выл0 коланлун. Оз позь вун0дны, мый татш0м уджтас л0 л0сь0д0ма абу озырм0дч0м выл0, а с0мын коми кыв с0вм0д0м могысь. Мед к0ть татш0м уджтас дасьт0м выл0 позяс судз0дны отс0г 0тувья удж нубдан к0сй0мьяс серти Рытыввыв Европыась.

Койм0д-к0, кор лексика0 чук0рмас 60.000 кыв да сй0 материалыс шымыртас озыр коми кывья классикьяс0с, л0 позянлун л0сь0дны сй0з шусяна торй0длана кыввел0д0м (Constraint Grammar). Кута лача, мый татш0мсяма 0тувья удж в0чны к0сй0м ловзяс Комиын.

Р0скодьяс0 вермас чинтыны неб0г-газет лэдзанін да кыв туялан учрежденн0 костын кутш0мк0 с0рнитч0м кырымал0м. Шуам, т0дан машина0 пырт0м газет гиж0дьяс да художества уджьяс, туялан текстьяс позь0 видзны дискетаяс вылын. Дискетаясл0н в0дитч0м тш0тш л0 т0дчым0н донт0мджык гиж0да материал выльысь пыртал0м дорысь. Позь0 м0впавны, мый став коми кывья художества гиж0дыс т0рас некымын CD ROM выл0.