

T-(538|725)-MALV, Málvinnsla lexc - **L**exicon **C**ompiler

Hrafn Loftsson¹ Hannes Högni Vilhjálmsson¹

¹Tölvunarfræðideild, Háskólinn í Reykjavík

September 2007

1 Tveggja laga reglur

2 lexc

1 Tveggja laga reglur

2 lexc

Kimmo Koskenniemi (1983)

Regla	Lýsing
$a:b \Rightarrow lc _ rc$	a is transduced as b only when it has lc to the left and rc to the right
$a:b \Leftarrow lc _ rc$	a is always transduced as b when it has lc to the left and rc to the right
$a:b \Leftrightarrow lc _ rc$	a is transduced as b always and only when it has lc to the left and rc to the right
$a:b / \Leftarrow lc _ rc$	a is never transduced as b when it has lc to the left and rc to the right

Frekari skýringar:

http://www.sil.org/pckimmo/v2/doc/Rules_1.html#subsec:3.1.4

Tveggja laga reglur

Dæmi úr ensku

Examples	happy+er	party+s	marry+ed
	happi0er	parties	marri0ed
Rules	Cy+er	Cy+s	Cy+ed
	Ci0er	Cies	Ci0ed

- 1 $y:i \leftarrow C:C _ +:0 e:e r:r$
- 2 $y:i \leftarrow C:C _ +:e s:s$
- 3 $y:i \leftarrow C:C _ +:0 e:e d:d$

Reglurnar eru keyrðar **samhliða**. Sérhver regla þarf að vera uppfyllt fyrir sérhvert par af stöfum áður en næsta par er skoðað (sjá Fig. 5.10 bls. 139).

Tengsl við stöðuferjöld

- Hægt að þýða tveggja laga reglu yfir í stöðuferjald.
- Forrit sem framkvæma tveggja laga orðhlutafræðilega greiningu:
 - Notandi býr til mengi af tveggja laga reglum.
 - Forritið þýðir þær yfir í stöðuferjöld.
- Dæmi: **lexc** (Lexicon Compiler) sem er hluti af **XFST** (Xerox Finite-State Tool).
- Dæmi: **PC-Kimmo**, <http://www.sil.org/pckimmo/>

1 Tveggja laga reglur

2 `lexc`

Lexicon Compiler

- Mál og þýðandi til að skilgreina stöðuvélar (e. automata) og stöðuferjöld (e. transducers).
- Sérstaklega hentugt til að skilgreina orðasöfn.
- Byggir á tveggja laga orðhlutafræði.

Snið lexc skrár

- Multichar_Symbols declaration
- Declarations section
- Lexicon Root
- Lexicon X
- Lexicon Y
- ...
- END

Lexicon Compiler

- Mál og þýðandi til að skilgreina stöðuvélar (e. automata) og stöðuferjöld (e. transducers).
- Sérstaklega hentugt til að skilgreina orðasöfn.
- Byggir á tveggja laga orðhlutafræði.

Snið lexc skrár

- Multichar_Symbols declaration
- Declarations section
- Lexicon Root
- Lexicon X
- Lexicon Y
- ...
- END

Einfalt dæmi

```
! ex1-lex.txt (this line is a comment)
LEXICON Root
dog      # ;
cat      # ;
bird     # ;
END
```

Keyrsla

```
xfst      ; Keyrir upp xfst sem býður eftir inntaki
xfst[0]:   read lexc < ex1-lex.txt
xfst[1]:
```

Einfalt dæmi

```
! ex1-lex.txt (this line is a comment)
LEXICON Root
dog      # ;
cat      # ;
bird     # ;
END
```

Keyrsla

```
xfst      ; Keyrir upp xfst sem býður eftir inntaki
xfst[0]:  read lexc < ex1-lex.txt
xfst[1]:
```

Keyrsla

```
xfst[1]: up dog      ; frá surface til lexical
xfst[1]: down bird  ; frá lexical til surface
xfst[1]: print words
xfst[1]: clear stack
xfst[0]
```

lexc – Continuation classes

LEXICON Root

walk # ;

walks # ;

walked # ;

walking # ;

talk # ;

talks # ;

talked # ;

talking # ;

pack # ;

packs # ;

packed # ;

packing # ;

lexc – Continuation classes

```
LEXICON Root
```

```
walk    V ;
```

```
talk    V ;
```

```
pack    V ;
```

```
LEXICON V
```

```
s      # ;
```

```
ed     # ;
```

```
ing    # ;
```

```
      #;  ! <- an empty-string entry
```

Lexical transducers

Upper:Lower færslur

LEXICON Root

```
swim:swam      # ;  
fiOght:fought # ;
```

Keyrsla

```
xfst[1]: up swam      ; frá surface til lexical  
xfst[1]: down fight  ; frá lexical til surface  
xfst[1]: print words
```

Lexical transducers

Upper:Lower færslur

LEXICON Root

```
swim:swam      # ;  
fi0ght:fought # ;
```

Keyrsla

```
xfst[1]: up swam      ; frá surface til lexical  
xfst[1]: down fight  ; frá lexical til surface  
xfst[1]: print words
```


Dæmi fyrir íslensku

- Sjá skrá ice-lex.txt undir “Annað efni”

```
xfst[0]: read lexc < ice-lex.txt
```

```
xfst[1]: up hestur
```

```
xfst[1]: down hest+NO+KK+FT+N
```

```
xfst[1]: down gest+NO+KK+FT+N
```