

# Málvinnsla - Skiladæmi 2

Háskólinn í Reykjavík - Tölvunarfræðideild

September 2007

## 1 Dæmi I - 10%

Beitið (handvirkri) orðhlutafræðilegri greiningu (e. morphological parsing) á orðin sem fram koma hér að neðan. Þið eigið að sýna öll myndön<sup>1</sup> sem koma fyrir í hverju orði, með "+" á milli myndana. T.d. fyrir orðið *fínn* mynduð þið skila greiningunni: *fín+n*. Athugið að við lítum svo á að orð samanstandi af myndönunum forskeyti+stofn+viðskeyti.

1. *læsa*
2. *virkur*
3. *bóndann*
4. *sígrænt*
5. *nemandi*

## 2 Dæmi II - 30%

Í þessu dæmi eigið þið að búa til stöðuvél í JFlex (kallið hana NafnKennsl.flex) sem svarar því hvort tiltekin nafnorð í skránni *nafnord.txt* (sjá að neðan) eru lögleg eður ei. Notið eftirfarandi kóða sem grunn að stöðuvélinni ykkar:

```
%% Grunnur að stöðuvél sem ber kennsl á ósamsett íslensk nafnorð

%public
%class NafnKennsl
%standalone
```

---

<sup>1</sup>Myndan er minnsta afmörkuð eind í byggingu orðsins.

```

%unicode

%{
%}

WhiteSpace = [ \t\f\n\r]

KK1Et =                               /* hestur */

/* Beygingaflokkar */

KK1 = ({KK1Et}|{KK1Ft}){WhiteSpace}+    /* hestur */
KvK1 = ({KvK1Et}|{KvK1Ft}){WhiteSpace}+ /* súpa */
HK1 = ({HK1Et}|{HK1Ft}){WhiteSpace}+    /* svín */

Error = {WordChar}+{WhiteSpace}+

/* Nafnorð er samsett af stofni + beygingarendingu (sleppum forskeytum) */

Ord = bekk{KK8} |
      eining{KvK3} |
      gald{KK4} |
      gest{KK2} |
      fund{KK6} |
      hest{KK1} |
      hól({KK7}|{HK1})|
      hugs{KvK4} |
      maur{KK5} |
      rjóð{HK2} |
      skól{KK3} |
      súp{KvK1} |
      svín{HK1} |
      von{KvK2}

%%
{Ord} { System.out.println("OK: " + yytext());}
{Error} { System.out.println("Error: " + yytext());}

```

Þið þurfið sem sagt að fylla inn í eyðurnar í þessari skrá og bæta við því sem þarf. Keyrið stöðuvélina á *nafnord.txt* með *java NafnKennsl nafnord.txt*. Skilið bæði .flex skránni og skrá sem sýnir úttakið ykkar. Skráin *nafnord.txt* lítur svona út:

```
hestur hest hesti hests
hesturinn hestinn hestinum hestsins
hestar hesta hestum hesta
hestarnir hestana hestunum hestanna
hestir hestina
gestur gest gesti gests
gesturinn gestinn gestinum gestsins
gestir gesti gestum gesta
gestirnir gestina gestunum gestanna
gestar gestana
súpa súpu súpu súpu
súpan súpuna súpunni súpunnar
súpur súpur súpum súpna
súpunar súpunar súpunum súpanna súpnanna
súpar súpirnar
von von von vonar
vonin vonina voninni vonarinnar
vonir vonir vonum vona
vonirnar vonirnar vonunum vonanna
vonur vonuna
skóli skóla skóla skóla
skólinn skólann skólanum skólans
skólar skóla skólum skóla
skólarnir skólana skólunum skólanna
skólur skólirnir
svín svín svíni svíns
svínið svínið svíninu svínsins
svín svín svínum svína
svínin svínin svínunum svínanna
svínur svínar svíninn
galdur galdur galdri galdurs
galdurinn galdurinn galdrinum galdursins
galdrar galdra göldrum galdra
galdarnir galdrana göldrunum galdrana
galdurinn galdrum
maur maur maur maurs
maurinn maurinn maurnum maursins
maurar maura maurum maura
maurarnir maurana maurunum mauranna
maurur maurirnir
fundur fund fund fundar
fundurinn fundinn fundinum fundarins
fundir fundi fundum funda
fundirnir fundina fundunum fundanna
funds fundarnir
höll hól hól höls
höllinn höllinn hölnum hölsins
hólar hóla hólum hóla
hólarnir hólana hólunum hólanna
hól hól hól höls
hólið hólið hólínu hölsins
hólin hólin hólunum hólanna
hólur hólirnir
bekkur bekk bekk bekkjar
bekkurinn bekkinn bekkinum bekkjarins
bekkir bekki bekkjum bekkja
bekkirnir bekkina bekkjunum bekkjanna
bekkar bekkunum
eining einingu einingar
einingin eininguna einingunni einingarinnar
einingar einingum eininga
einingarnar einingunum eininganna
einingur einingirnir
hugsun hugsun hugsun hugsunar
hugsunin hugsunina hugsuninni hugsunarinnar
hugsanir hugsanir hugsunum hugsana
hugsanirnar hugsanirnar hugsununum hugsananna
hugsur hugsirnir
rjóður rjóður rjóðri rjóðurs
```

rjóðrið rjóðrið rjóðrinu rjóðursins  
rjóður rjóður rjóðrum rjóðra  
rjóðrin rjóðrin rjóðrunum rjóðrana  
rjóðið rjóðanna

### 3 Dæmi III - 30%

Markið eftirfarandi setningar í höndunum með því að nota mörk úr íslenska markamengið (lýsing á því er aðgengileg undir “Annað efni”).

1. ég stappaði fótunum á svarta gúmmimottu.
2. hann vissi mætavel að henni veitti ekki af hverri mínútu.
3. hún reyndi að leyna kvíða sínum.
4. hún hreyfði hvorki legg né lið.
5. hvergi á byggðu bóli voru umsvif heildsala jafn víðtæk.

Notið síðan *IceTagger* markarann (á [http://nlp.ru.is/icenlp\\_isl.htm](http://nlp.ru.is/icenlp_isl.htm)) til að marka sömu setningar (ekki svindla með því að nota hann strax í upphafi :-)).

1. Bendið á þær villur sem *IceTagger* gerir (hann gerir a.m.k. eina villu í hverri af þessum setningum).
2. Getið til um hver sé líkleg ástæða viðkomandi villu.
3. Hver er hittni (nákvæmni) *IceTagger* fyrir þessar prófunarsetningar?

### 4 Dæmi IV - 30%

Í þessu dæmi eigið þið að þjálfra tölfræðimarkarann *TnT* og prófa hann á setningunum úr dæmi III.

1. Sækið <http://www.ru.is/kennarar/hrafn/Software/tnt-malv.zip>
2. Afþjappið skránni, t.d. á C:\tnt
3. Rennið yfir fyrstu 10 blaðsíðurnar í skránni *manual.pdf* til að kynnst því hvernig nota skal markarann.
4. Notið skrána *01TM.txt* (búið er að breyta “p” í “Z” í öllum mörkum í skránni því *TnT* ræður ekki við “p”-ið) til að þjálfra markarann. Sýnið skipunina sem framkvæmir þjálfunina.

5. Látið  $TnT$  marka skrána *test.txt* sem inniheldur setningarnar úr dæmi III. Sýnið skipunina sem framkvæmir mörkunina.
6. Hver er hittni (nákvæmni)  $TnT$  fyrir þessar prófunarsetningar?

## 5 Skilafrestur

Verkefninu skal skila í síðasta lagi þriðjudaginn 25. september, kl. 23:59.