

Íslenskur trjábanki  
Verkfæri fyrir meginlegar rannsóknir í setningafræði

Anton Karl Ingason, Einar Freyr Sigurðsson og Joel C. Wallenberg

1. apríl 2011  
Málvísindakaffi  
Háskóli Íslands

## Efnisyfirlit

<b>1</b>	<b>Staða verkefnisins</b>	<b>2</b>
<b>2</b>	<b>Greiningaraðferð</b>	<b>2</b>
2.1	Liðir, mörk og orð . . . . .	2
2.2	Innri gerð liða . . . . .	4
2.3	Beygingarliðurinn . . . . .	6
<b>3</b>	<b>Leitað á vefnum</b>	<b>7</b>
3.1	Corpald leitarviðmótið . . . . .	7
3.2	Ýmis dæmi um leitarfyrirspurnir . . . . .	8

# 1 Staða verkefnisins

IcePaHC 0.4: Útgáfudagur 8. apríl 2011  
Núverandi staða: 419.073 orð

## 2 Greiningaraðferð

### 2.1 Liðir, mörk og orð

- (1) **VB: Sögn í nafnhætti**

VB  
|  
stela

- (2) **VBPI og VBPI: Nútíð (P) og þátíð (D) framsöguháttar (I)**

VBPI  
|  
stelur

- (3) **VBPS og VBPS: Viðtengingarháttur (S)**

VBDS  
|  
stæli

- (4) **VAN: Þolmyndarháttur**

VAN  
|  
drepinn

- (5) **VBN: Lýsingarháttur þátíðar**

VBN  
|  
drepíð

- (6) **BE: vera**

BEPI  
|  
er

- (7) **HV: hafa**

HVDI  
|  
hafði

- (8) **MD: máttarsögn**

MD  
|  
skulu

- (9) **Einfaldur atviksliður**

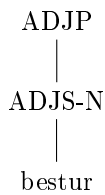
- ADVP  
|  
ADV  
|  
vel
- (10) **Atviksliðir með hlutverk: tíðaratviksorð**  
ADVP-TMP  
|  
ADV  
|  
núna
- (11) **Staðaratviksorð**  
ADVP-LOC  
|  
ADV  
|  
hérna
- (12) **Stefnuatviksorð**  
ADVP-DIR  
|  
ADV  
|  
þangað
- (13) **Nafnliður í nefnifalli: N-N (noun-nominative)**  
NP  
|  
N-N  
|  
hestur
- (14) **Nafnliður í þágufalli: PRO-D (pronoun-dative)**  
NP  
|  
PRO-D  
|  
mér
- (15) **Lýsingarorðsliður í þolfalli: ADJ-A (adjective-accusative)**  
ADJP  
|  
ADJ-A  
|  
góðan
- (16) **Lýsingarorðsliður í eignarfalli: ADJ-G (adjective-genitive)**



(17) **Lýsingarorð í miðstigi ADJR**

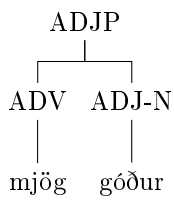


(18) **Lýsingarorð í efsta stigi ADJS**

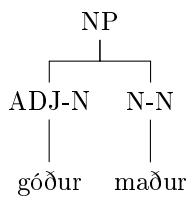


## 2.2 Innri gerð liða

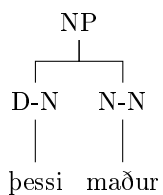
(19) **Lýsingaorðsliður með atviksorði og lýsingarorði**



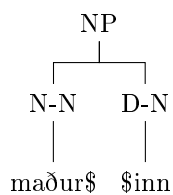
(20) **Nafnliður með lýsingarorði og nafnorði**



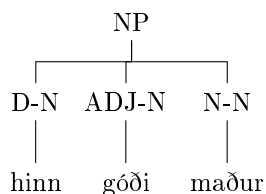
(21) **Ákveðinn nafnliður (ábendingarfornafn)**



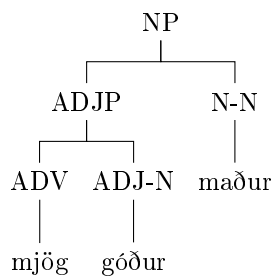
(22) **Ákveðinn nafnliður (viðskeyttur greinir)**



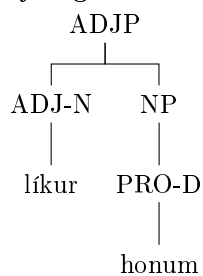
(23) **Ákveðinn nafnliður (laus greinir)**



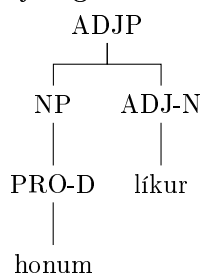
(24) **Nafnliður með lýsingarlið (stærri en bara lo.) og nafnorði**



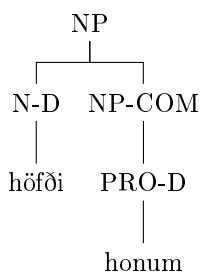
(25) **Lýsingarorðsliður yfirskipaður nafnlið**



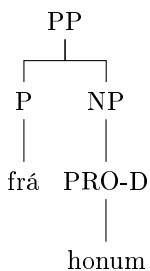
(26) **Lýsingarorðsliður yfirskipaður nafnlið með nafnliðinn á undan lýsingarorðinu**



(27) **NP-COM: Nafnliður yfirskipaður nafnlið í þágufalli**

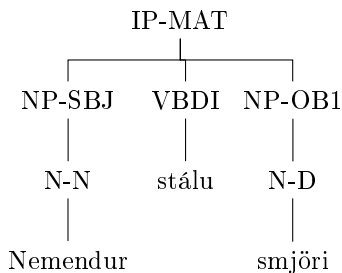


(28) **PP: Forsetningarliður**

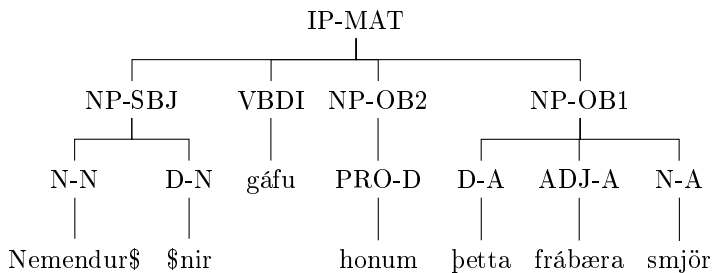


### 2.3 Beygingarliðurinn

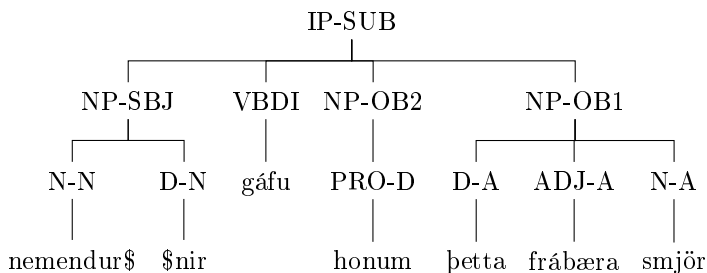
(29) **IP-MAT: aðalsetning – frumlag (SBJ), sögn, beint andlag (OB1)**



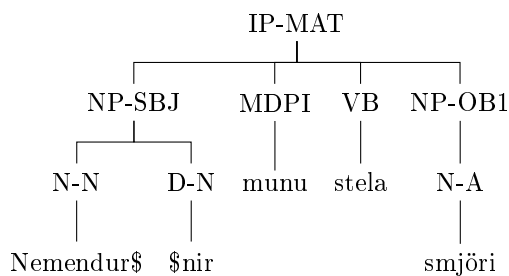
(30) **IP-MAT: aðalsetning – frumlag (SBJ), sögn, óbeint (OB2) og beint andlag (OB1)**



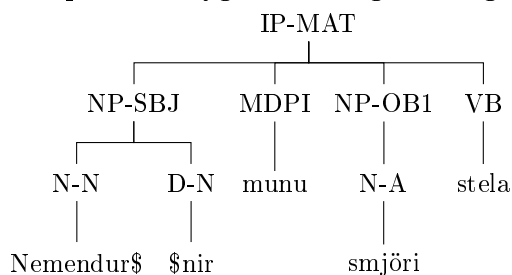
(31) **IP-SUB: aukasetning: Ég veit að nemendurnir ...**



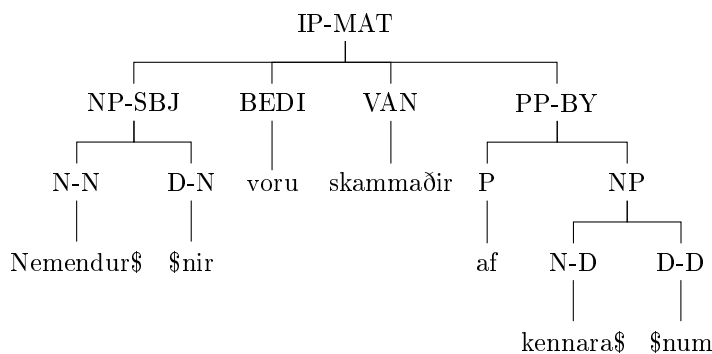
(32) **VO: persónubeygd háttarsögn-aðalsögn-andlag**



(33) **OV: persónubeygd háttarsögn-andlag-aðalsögn**



(34) **Polmynd með af-lið**



### 3 Leitað á vefnum

#### 3.1 Corpald leitarviðmótið

Fyrirvari: Því fer fjarri að leitarviðmótið sé tilbúið. Það er þó nú þegar hægt að leita að ákveðnum tegundum mynstra.

Bráðabirgðavefslóð: <http://www.hi.is/~antoni/corpald>

Leitarforritið virkar best í Google Chrome vafranum en ætti þó líka að virka ágætlega í Firefox og líklega Opera og Safari líka. Corpald hefur ekki verið prófað í Internet Explorer og er líklega mun hægvirkara þar, ef það virkar yfir höfuð. Þegar slóðin er slegin inn hleður Corpald inn öllum trjábakanum og þetta getur tekið tíma, jafnvel mínútu eða lengur. Hraði nettengingar og það hversu öflug tölva notandans er hefur áhrif á hraðann. Þegar búið er að hlaða inn öllum gögnunum er hægt að slá inn hverja fyrirspurnina á eftir annarri án biðtíma á milli. Fyrirspurnir eru mishraðvirkar og ef niðurstöður eru mjög margar er eðlilegt að það taki nokkurn tíma að keyra þær.

Í þessu sýnishorni af leitarviðmóti er aðeins hægt að tilgreina eina keðju af venslum í gegnum hrísluna. Fjórar tegundir vensla eru í boði:

**doms:** dominates = yfirskipun

**idoms:** immediately dominates = bein yfirskipun

**sprecedes:** sister-wise precedence = kemur á undan (til vinstri við) systur í trénu

**siprecedes:** sister-wise immediate precedence = kemur næst á undan systur í trénu

Til að leita að tilteknum venslum þarf að tilgreina merkimiða á hnútum í hríslunni, nöfn á liðum, mörk eða orð, og hvernig þessir hnútar tengjast. Hægt er að leita að merkimiðum eftir nákvæmum streng, upphafi strengs eða enda strengs:

**NP-SBJ:** leitar að nákvæma strengnum NP-SBJ

**NP-\***: leitar að streng sem byrjar á NP-, þ.e. öllum nafnliðum, t.d. bæði NP-SBJ og NP-OB1

**\*-D:** leitar að streng sem endar á -D, þ.e. öllum þágufallsörkum, t.d. N-D, PRO-D, ADJ-D

### 3.2 Ýmis dæmi um leitarfyrirspurnir

- (35) **Þágufallsfrumlög:**  
NP-SBJ idoms \*-D  
Mér líkar þetta.
- (36) **Eignarfallsandlög:**  
NP-OB1 idoms \*-G  
Allir sakna hans.
- (37) **Tækisfall (NP adjunct):**  
IP-\* idoms NP-ADT  
Hún greip sverðið tveimur höndum.
- (38) **Nafnliður yfirskipaður nafnlið í þágufalli:**  
NP-\* idoms NP-COM  
Hann hafði hatt á höfði sér.
- (39) **Nafnliður innan forsetningarliðar yfirskipaður nafnlið í þágufalli:**  
PP idoms NP-\* idoms NP-COM  
Hann hafði hatt á höfði sér.
- (40) **Nafnliður undir beygingarlið (frumlag, andlag, etc.) yfirskipaður nafnlið í þágufalli:**  
IP-\* idoms NP-\* idoms NP-COM  
(hér kemur reyndar nokkur mökkur villna í IcePaHC 0.3, kannski eru engin svona dæmi í raun)
- (41) **Lýsingarliður yfirskipaður nafnlið:**  
ADJP idoms NP  
Hún er lík Lady Gögu.
- (42) **Lýsingarliður yfirskipaður nafnlið sem kemur á eftir lýsingarorði:**  
ADJP idoms ADJ\* sprecedes NP  
Hún er engum lík.
- (43) **Lýsingarliður yfirskipaður nafnlið sem kemur á undan lýsingarorði:**  
ADJP idoms NP sprecedes ADJ\*  
Hún er lík Lady Gögu.
- (44) **VO-háttarsögn: háttarsögn–nafnháttur–andlag:**  
IP-\* idoms MD\* sprecedes VBN sprecedes NP-OB\*  
Þau munu stela smjörinu.

Ath. að IP-\* nær yfir alla beygingarliði. Hægt er að takmarka sig við aðalsetningar með IP-MAT, aukasetningar með IP-SUB og nafnháttarsetningar með IP-INF. Ráðlegt er að enda á stjörnu ef maður er að gera rannsókn því að stundum eru fleiri merkimiðar á eftir MAT og SUB. Hægt er að nota IP-MAT\* og IP-SUB\* til að vera nokkuð viss um að ná öllum dæmum sem skipta máli.

- (45) **OV-háttarsögn: háttarsögn–andlag–nafnháttur:**  
IP-\* idoms MD\* sprecedes NP-OB\* sprecedes VBN  
Þau munu smjörinu stela.
- (46) **VO-hafa: hafa–lýsingarháttur–andlag í aðalsetningu:**  
IP-\* idoms HV\* sprecedes VBN sprecedes NP-OB\*

Nemandinn hafði skrifað gula miðann.

(47) **OV-hafa: hafa–andlag–lýsingarháttur í aðalsetningu:**

IP-\* idoms HV\* sprecedes NP-OB\* sprecedes VBN

Nemandinn hafði gula miðann skrifað.

(48) **Af-liðir í IcePaHC 0.3 (ónákvæmt):**

IP-\* idoms VAN sprecedes PP idoms P idoms af

Nemandinn var skammaður af kennaranum.

Nemandinn var skammaður af því að hann skrifaði gula miðann.

(49) **Af-liðir í IcePaHC 0.4:**

IP-\* idoms VAN sprecedes PP-BY

Nemandinn var skammaður af kennaranum.