

The US-Iceland NSF Workshop, Reykjavík, June 1st, 2007

The electronic corpus of Old Icelandic and its use in syntactic research

Eiríkur Rögnvaldsson

Outline

- The Old Icelandic corpus
- Tagging the corpus
- Searching for syntactic constructions
- A shallow parser for Icelandic

The Old Icelandic corpus

- Around two million running words
 - mostly narrative texts
- Texts from the 13th and 14th centuries
 - none of them preserved in the original
 - manuscripts often considerably younger
- Almost all in modern Icelandic spelling
 - taken from recent editions

Narrative texts

- Íslendingasögur (The Icelandic Family Sagas)
 - 984,000 words
- Sturlunga saga (12th and 13th centuries)
 - 260,000 words
- Heimskringla (Sagas of the Kings of Norway)
 - 232,000 words
- Landnámabók (The Book of Settlements)
 - 37,000 words

Other available texts

- Grágás (Old Icelandic law)
 - approximately 200,000 running words
- Religious literature (Book of Homilies)
 - 85,000 running words
- Biskupasögur (Sagas of bishops)
 - size unknown
- A few texts from the 16th-18th centuries
 - the New Testament, autobiographies, homilies, etc.

Problems

- Problems with the text as syntactic sources
 - Dating
 - Authenticity
 - Errors and changes made by scribes and editors
 - Word-for-word translations
 - Poetic license
 - Differences between spoken and written language
 - etc.

Variations between manuscripts

- [...] none of these texts is found in the original; most of them are only preserved in manuscripts from the fourteenth and fifteenth (and in a few cases sixteenth) centuries. This makes it extremely difficult to assess the validity of these texts as linguistic evidence, since it is often impossible to know whether a certain feature of the preserved text stems from the original or from the scribe of the preserved copy, or perhaps from the scribe of an intermediate link between the original and the preserved manuscript. It is well known that scribes often did not retain the spelling of the original when they made copies; instead, they used the spelling that they were used to. In many cases, two or more manuscripts of the same text are preserved, and usually they differ to a greater or lesser extent.

Different age of the manuscripts

- However, it is usually assumed that the syntax of Old Icelandic did not change much in the thirteenth and fourteenth centuries. Therefore, I feel justified in lumping together various narrative texts from these centuries and treating them as if they exhibit the same stage of the language. In working with these texts, I have not noticed any significant syntactic differences between those that are assumed to be relatively old and preserved in older manuscripts, and those that are considered relatively young and are preserved in younger manuscripts. It is possible that future research will show that it is illegitimate to treat these texts as roughly contemporaneous; but in doing so, I follow the standard practice of traditional syntactic descriptions (see especially Nygaard 1905; Heusler 1967).
 - Eiríkur Rögnvaldsson: Word Order Variation in the VP in Old Icelandic

Editorial changes

- Manuscript text in square brackets
 - ok lagði Hallgrímr í gegnum hann [í hann ok í gegnum] og svá í slána ... Gunnarr tók atgeirinn ok lagði í gegnum Hallgrím ____ [ok kvað þa visu ...]. Gunnar [ok þat efndi G.] ____ [at hann] bar atgeirinn jafnan síðan [meðan hann lifði].
 - (*Brennu-Njálssaga*, p. 68; from Sigurðsson 1985:16.)

Scribal changes

- Kjartan Ottosson – seminar in Oslo, 2001:
 - General characteristics of the copies:
 - rearrangements of word order very common, especially in local context
 - other syntactic elements more constant
- Jóns saga helga A og Augustinus saga
 - ca. 1275-1300 → ca. 1350
 - hann qveðr þa hana > hann kveðr hana þaa
 - biðr þa þat eigi gera > eigi þat giora
 - þorði villumaðrinn eigi þat gera > þat eigi at gera

More scribal changes

- Magnúss saga góða (Morkinskinna)
 - ca. 1350-1400 → ca. 1475-1500
 - hafa tekit > tekit hafa
 - sva lez hann fyr hafa gert > letz so fyr giort hafa
 - malit þitt > þitt mal
- Ólafs saga Tryggvasonar hin mesta
 - ca. 1387-95 → 1498
 - þessa alla menn > alla þessa menn
 - er gætt þeirra skyllde hafua > er þeirra skyldu gætt hafa
 - lettu sinne ferd æigi > lettu eigi sinne ferd

Tagging the corpus

Lemmatized concordance

Egla	464	..Egils að hann væri flutningsmaður við konung . að Aðalsteinn konungur gerði...
Egla	375	..og sumar kom þá það Bárður sér orlofs konung að fara að vitja ráðs þess er honum...
Egla	433	..staðfest. Þá báðu sendimenn Ólaf konung að gefa sér tólm til að þeir hittu enn...
BjHít	82	..var þá þar og túlkaði vel málið við konung að hann fengi fé sitt. Konungur lét...
Egla	427	..er þeir tóku ræður sínar þá það Þórir konung að hann skyldi eigi fyrirkunna hann...
Egla	427	..vel faðir" segir Arinbjörn, "því við konung að hann telji ekki að um vist Egils....
HallÓ	1249	..Hallfreði mikill skaði að um Ólaf konung að hann undi engu og ætlaði þá að...
Egla	411	..hrið og leituðu þess dagraðs að hitta konung að hann var í góðu skapi, báru þá upp...
Egla	381	..inn í Naumudal og fóru svo í svig við konung að þeir hittu hann að öðru hverju. Tók...
Fóstb	794	..virðing að Danir tignuðu hann nær sem konung að því sem Þormóður hefir um ort...
Egla	458	..við því búist að yrkja lof um Eirik konung . ' Arinbjörn það hann freista...
HallM	1212	..þótti mikið fyrir skilnaði við konung . Auðgísl son Hallfreðar fór austur til...
Laxd	1545	..þenna því að hann hafði eigi fundið konung á þeim vetni. Þangað var og kaupstefnu...
Vatn	1853	..lyfinguna á konungsskipinu og kvaddi konung á þessa leið: "Heill, heill, herra." ...
Laxd	1598	..kristnum mönnum. En þó vil eg finna konung áður en eg ráði þetta til staðar því að...
Njála	129	..vom. En þó gakk þú að finna konung áður þú farir." Hrútur gerði svo...
Egla	464	..til Lundúna og hitti þar Aðalstein konung . Bar hann fram jartegnir og orðsending...
Egla	410	..var þá í hinum mestum kærleikum við konung . Björn og þeir Þórólfur fóru...
GíslS	858	..að við Noreg skipi sínu og hitta konung brátt og ganga fyrir hann og kveðja...

Searching raw text

- Simple search using wildcards
 - *lang**
 - *langa, langar, langi, langaði, langað*
 - engan okkar langar til að deyja
 - *langa v.*
 - hann heyrði til þeirra langar leiðir
 - *langur adj.*
- Topicalization in embedded clauses
 - *Ég veit [að þennan mann þekkir þú ekki]*

Double object patterns

(10) *Sentences with two non-finite verbs and two objects (= (6)):*

- a. (XP) - V_{fin} - V_{aux/mod} - V_{main} - NP_{IO} - NP_{DO}
- b. (XP) - V_{fin} - V_{aux/mod} - NP_{IO} - V_{main} - NP_{DO}
- c. (XP) - V_{fin} - V_{aux/mod} - NP_{IO} - NP_{DO} - V_{main}
- d. (XP) - V_{fin} - V_{aux/mod} - NP_{DO} - V_{main} - NP_{IO}
- e. (XP) - V_{fin} - V_{main} - V_{aux/mod} - NP_{IO} - NP_{DO}
- f. (XP) - V_{fin} - NP_{IO} - V_{aux/mod} - V_{main} - NP_{DO}
- g. (XP) - V_{fin} - NP_{IO} - V_{aux/mod} - NP_{DO} - V_{main}
- h. (XP) - V_{fin} - NP_{IO} - V_{main} - V_{aux/mod} - NP_{DO}
- i. (XP) - V_{fin} - NP_{IO} - NP_{DO} - V_{aux/mod} - V_{main}
- j. (XP) - V_{fin} - NP_{IO} - NP_{DO} - V_{main} - V_{aux/mod}
- k. (XP) - V_{fin} - NP_{DO} - V_{aux/mod} - V_{main} - NP_{IO}
- l. (XP) - V_{fin} - NP_{DO} - V_{aux/mod} - NP_{IO} - V_{main}
- m. (XP) - V_{fin} - NP_{DO} - V_{main} - V_{aux/mod} - NP_{IO}
- n. * (XP) - V_{fin} - V_{main} - NP_{IO} - V_{aux/mod} - NP_{DO}
- o. * (XP) - V_{fin} - V_{main} - NP_{IO} - NP_{DO} - V_{aux/mod}
- p. * (XP) - V_{fin} - V_{main} - NP_{DO} - V_{aux/mod} - NP_{IO}
- q. * (XP) - V_{fin} - NP_{IO} - V_{main} - NP_{DO} - V_{aux/mod}
- r. * (XP) - V_{fin} - NP_{DO} - V_{main} - NP_{IO} - V_{aux/mod}

Double object sentences

Setningar með tveim sögnum í fallhætti og tveim andlögum:

- a. ... en Eyjólfur kvaðst hafa gefið Bjarna sauðina.
- b. ... en kveðst segja honum mundu fyrstum tíðindin.
- c. Gísli kveðst aldrei mundu Kolbeini eiða sverja.
- d. Ófeigur ... kveðst ekki mundu mikla tillögu veita honum.
- e. ... en hann mundi fengið hafa henni mikinn áverka.
- f. Ásbjörn lést honum mundu gefa fé til.
- g. ... hví Bolli mun sér hafa þar svo staðar leitað ...
- h. Hallbjörn ... kveðst enn hverjum manni veitt hafa búðarrúm
- i. Bárður kvaðst honum engi mundu segja.
- j. ... að þessi maður mundi honum sanna hluti sagt hafa.
- k. Þykist eg nú það hafa launað þér.
- l. Hann kveðst fyrir löngu það hafa honum spáð ...

Modern Icelandic Corpus

- Modern Icelandic Corpus
 - source files of the IFD
 - *Icelandic Frequency Dictionary (Íslensk orðtíðnibók)*
 - 100 texts – 500,000 running words
- Lemmatized and morphologically tagged
 - 662 different tags
 - compared to 45 in the *Penn Treebank Tagset*

From the IFD corpus

- f p k e n hann hann
- s f g 3 e þ o átti eiga
- n h e o afmæli afmæli
- a o í í
- n k e o dag dagur
- c og og
- n k e n g hvolpurinn hvolpur
- n k e n - m Vaskur Vaskur
- s f g 3 e þ var vera
- n v e n afmælisgjöf afmælisgjöf

Available PoS taggers

- TnT (Trigrams'n'Tags) by Thorsten Brants
 - data-driven tagger based on HMM
 - <http://www.coli.uni-saarland.de/~thorsten/tnt/>
 - trained on Icelandic texts by Sigrún Helgadóttir
- IceTagger
 - rule-based tagger using finite-state methods
 - <http://nlp.ru.is/>
 - developed by Hrafn Loftsson

Tagging Old Icelandic texts

- Recently, we have started a project which aims at POS tagging the whole Old Icelandic corpus. We use Brant's TnT tagger, which has been trained on Modern Icelandic, to tag a part of the corpus. This part is then manually corrected and the tagger retrained on the corrected tags and applied to the next part of the corpus, and so on. Even though the tagger makes a number of errors on the first pass, this procedure has proven to be workable, due to the morphosyntactic similarities between Old and Modern Icelandic.

An example from the tagged corpus

129	það	Pron	Pers	Neut		Sg	Nom
130	var	Verb	Ind	Act	3Pers	Sg	Past
131	sagt	Adj	Strong	Neut	Pos	Sg	Nom
132	eitt	Numeral		Neut		Sg	Acc
133	vor	Noun		Neut		Sg	Acc
134	að	Conj/Comp					
135	hann	Pron	Pers	Masc		Sg	Nom
136	keypti	Verb	Subj	Act	3Pers	Sg	Past
137	sér	Pron	Pers	Masc		Sg	Dat
138	skip	Noun		Neut		Sg	Acc
139	og	CoorConj					
140	færi	Verb	Subj	Act	3Pers	Sg	Past
141	á	Prep/Prt					
142	Strandir	Noun		Fem		Pl	Acc
143	norður	Adv					

Some corrections

129	það	fphen	fphen
130	var	sfg3eþ	sfg3eþ
131	sagt	ssg	sþghen
132	eitt	tfheo	tfheo
133	vor	nheo	nheo
134	að	c	c
135	hann	fpken	fpken
136	keypti	sfg3eþ	svg3eþ
137	sér	fpkeþ	fpkeþ
138	skip	nhfn	nheo
139	og	c	c
140	færi	svg3eþ	svg3eþ
141	á	ao	ao
142	Strandir	nkeo	nvfo
143	norður	aa	aa

- 131
 - supine >
past participle
- 136
 - indicative >
subjunctive
- 138
 - nominative >
accusative
- 142
 - masculine singular >
feminine plural

Tags in continuous text

- Tags follow each word
 - ég **fp1en** stökk **sfg1eþ** á **aa** eftir **aþ** strætó **nkeþ** og **c** veifaði **sfg1eþ** , , vagnstjórinn **nkeng** sá **sfg3eþ** mig **fp1eo** og **c** stoppaði **sfg3eþ** . . ég **fp1en** tautaði **sfg1eþ** takk **au** og **c** brosti **sfg1eþ** til **ae** hans **fpkee** um **ao** leið **nveo** og **c** ég **fp1en** lét **sfg1eþ** miðann **nkeog** detta **sng** . .

Searching for syntactic constructions

Embedded topicalization

Advanced word search [X]

Mask:

að ef þegar svo þó nema	c	*	n??o* n??þ* n??e* f???o f???þ f???e	

In middle: 1 2 3 4 5

Words between: 0 0 0 0

▲	▲	▲	▲
▼	▼	▼	▼

Must be in this order: 1-2 2-3 3-4 4-5

Max. list size: rows

Clear All Load... OK Cancel

Lemmas... Save... Help

The results

- Search with these patterns in the IFD corpus returned more than 400 results
- Most of them show oblique subjects
 - ég held [að **henni** leiðist vinnan]
 - og af hverju ekki [ef **mér** leyfist að spyrja?]
 - það var óneitanlega léttir [þegar **þessu** var lokið]
 - allt og sumt var [að **hana** hafði dreymt eitthvað um nóttina]

Subordinate topicalization

- However, a few examples show topicalization
 - Guð veit [að það geri ég ____ líka]
 - hún fullyrti [að það myndi hún gera ____]
 - ég gerði hins vegar alls ekki ráð fyrir [að þessu sæti myndi fylgja ____ seta á Alþingi]
 - þeir vissu þó vel [að þetta máttu þeir ____ ekki]
 - reynt var að leiða henni fyrir sjónir [að þetta yrðu allir að gera ____]
 - Vigdís forseti segir [að bestu ljóð um vor og sumar hafi karlar ort ____ til kvenna]

Transitive expletives

Advanced word search ✕

Mask:

fphen	*	s??3??	*	n??n* f???n l??n??

In middle: 1 2 3 4 5

Words between: 0 0 0 0

<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="button" value="▲"/>	<input type="button" value="▲"/>	<input type="button" value="▲"/>	<input type="button" value="▲"/>
<input type="button" value="▼"/>	<input type="button" value="▼"/>	<input type="button" value="▼"/>	<input type="button" value="▼"/>

Must be in this order: 1-2 2-3 3-4 4-5

Max. list size: rows

<input type="button" value="Clear All"/>	<input type="button" value="Load..."/>	<input type="button" value="OK"/>	<input type="button" value="Cancel"/>
<input type="button" value="Lemmas..."/>	<input type="button" value="Save..."/>		<input type="button" value="Help"/>

Examples from the ÍSTAL-corpus

- In a quick search I made in all 6000+ examples of *það* a few years ago, I couldn't find any such sentences in the ÍSTAL corpus
- The above-mentioned method returned a few examples
 - *það átti enginn skap saman*
 - *það þekkja allir Rósu*
 - *það vita allir hver Rósa er*
 - *það heldur enginn að þú sért hommi*

The “new construction”

Advanced word search [X]

Mask:

s??3?? ss? sn?	*	spghen	*	n??o* n??p* n??e* f???o f???p f???e

In middle: 1 2 3 4 5

Words between: 0 0 0 0

<input type="button" value="▲"/>	<input type="button" value="▲"/>	<input type="button" value="▲"/>	<input type="button" value="▲"/>
<input type="button" value="▼"/>	<input type="button" value="▼"/>	<input type="button" value="▼"/>	<input type="button" value="▼"/>

Must be in this order: 1-2 2-3 3-4 4-5

Max. list size: rows

<input type="button" value="Clear All"/>	<input type="button" value="Load..."/>	<input type="button" value="OK"/>	<input type="button" value="Cancel"/>
<input type="button" value="Lemmas..."/>	<input type="button" value="Save..."/>		<input type="button" value="Help"/>

Old Icelandic

- Eftir það var lokið þinginu
 - Og er þetta var sagt Snorra goða
 - Í því var sleppt blámanninum
 - og var skipt landinu í helminga
 - Eftir það var lýst áverkunum
 - þótti varnað sér sæmdar
 - Var skipt verkum með húskörlum
 - Í því var lokið stofuhurðinni
 - Svo var háttað húsinu
 - Þá var beitt útgöngu með konum
 - Síðan var kvatt féránsdóms
 - og var komið griðum á
- The search returns several examples of post-verbal objects in passive sentences, **but**
 - they all show full NPs
 - no examples with pronominal objects have been found

Object shift

Advanced word search [X]

Mask:

In middle: 1 2 3 4 5

Words between: 3 0 0 0

Must be in this order: 1-2 2-3 3-4 4-5

Max. list size: rows

Object shift in Old Icelandic

- Gunnar hallmælti **Hallgerði ekki** um fyrir sendimanninum
- Nú leita þeir um skóginn og finna **Gísla eigi**
- En þó viljum vér **þenna kost eigi**
- fann hann **Snorra ekki** í þessi ferð
- Létta þeir **ferðinni eigi** fyrr en þeir koma suður í Þrándheim
- En þó viljum vér **þenna kost eigi**
- og fundu **Þórð eigi** sem von var að
- er hann dræpi **Þórð eigi** og förunauta hans
- fyrir því drap eg **Þórð eigi** og alla skipshöfn hans
- Nú bjó Skúta **þetta mál ekki** til

A shallow parser for Icelandic

Annotation scheme

- [Phrase]
 - NP(s)
 - AP(s)
 - AdvP
 - PP
 - CP
 - SCP – subordinate conj.
 - VP(i/b/s/p/g)
 - InjP – interjection phrase
 - MWEx – multiword expr.
- { *Function }
 - *SUBJ
 - *OBJ
 - *OBJAP – obj. with adjective
 - *OBJNOM – nominative obj.
 - *IOBJ
 - *COMP – predicative compl.
 - *QUAL – genitive qualifier
 - *TIMEX – temporal expr.
 - *X > - belongs to following V
 - *X < - belongs to preceding V

Original text

- pabbi nken hennar fpvee var sfg3eþ vinur nken minn feken og c við fp1fn tefldum sfg1fþ oft aa saman aa á ao kvöldin nhfog og c hann fpken studdi sfg3eþ mig fp1eo og c ég fp1en hélt sfg1eþ meira lheovm að cn segja sng áfram aa að cn koma sng til ae hans fpkee eftir ao að c allt fohen var sfg3eþ orðið spghen skrytið lhensf og c hún fpven hætt spgven að cn hringja sng í ao mig fp1eo . .

Multiword expressions

- pabbi nken hennar fpvee var sfg3eþ vinur nken minn feken og c við fp1fn tefldum sfg1fþ oft aa saman aa á ao kvöldin nhfog og c hann fpken studdi sfg3eþ mig fp1eo og c ég fp1en hélt sfg1eþ [MWE_AdvP meira lhevom að cn segja sng MWE_AdvP] áfram aa að cn koma sng til ae hans fpkee [MWE_SCP eftir ao að c MWE_SCP] allt fohen var sfg3eþ orðið spghen skrýtið lhensf og c hún fpven hætt spgven að cn hringja sng í ao mig fp1eo . .

Noun phrases

- *[NP pabbi nken NP] [NP hennar fpvee NP] var sfg3eþ [NP vinur nken minn feken NP] [CP og c CP] [NP við fp1fn NP] tefldum sfg1fþ [AdvP oft aa AdvP] [AdvP saman aa AdvP] á ao [NP kvöldin nhfog NP] [CP og c CP] [NP hann fpken NP] studdi sfg3eþ [NP mig fp1eo NP] [CP og c CP] [NP ég fp1en NP] hélt sfg1eþ [MWE_AdvP meira lheovm að cn segja sng MWE_AdvP] [AdvP áfram aa AdvP] að cn koma sng til ae [NP hans fpkee NP] [MWE_SCP eftir ao að c MWE_SCP] [NP allt fohen NP] var sfg3eþ orðið spghen [AP skrítið lhensf] [CP og c CP] [NP hún fpven NP] hætt spgven að cn hringja sng í ao [NP mig fp1eo NP] ..*

Prepositional phrases

- [NPn pabbi nken] [NPg hennar fpvee] [VPb var sfg3eþ VPb] [NPn vinur nken minn feken] [CP og c CP] [NPn við fp1fn] [VP tefldum sfg1fþ VP] [AdvP oft aa AdvP] [AdvP saman aa AdvP] [PP á ao [NPa kvöldin nhfog] PP] [CP og c CP] [NPn hann fpken] [VP studdi sfg3eþ VP] [NPa mig fp1eo] [CP og c CP] [NPn ég fp1en] [VP hélt sfg1eþ VP] [MWE_AdvP meira lheovm að cn segja sng MWE_AdvP] [AdvP áfram aa AdvP] [VPi að cn koma sng VPi] [PP til ae [NPg hans fpkee] PP] [MWE_SCP eftir ao að c MWE_SCP] [NPn allt fohen] [VPb var sfg3eþ VPb] [VPp orðið spghen VPp] [APn skrýtið lhensf] [CP og c CP] [NPn hún fpven] [VPp hætt spgven VPp] [VPi að cn hringja sng VPi] [PP í ao [NPa mig fp1eo] PP] ...

Subjects

- *{*SUBJ> [NPn pabbi nken] {*QUAL [NPg hennar fpvee] *QUAL} *SUBJ>}* [VPb var sfg3eþ VPb] [NPn vinur nken minn feken] [CP og c CP] *{*SUBJ> [NPn við fp1fn] *SUBJ>}* [VP tefldum sfg1fþ VP] [AdvP oft aa AdvP] [AdvP saman aa AdvP] [PP á ao [NPa kvöldin nhfog] PP] [CP og c CP] *{*SUBJ> [NPn hann fpken] *SUBJ>}* [VP studdi sfg3eþ VP] [NPa mig fp1eo] [CP og c CP] *{*SUBJ> [NPn ég fp1en] *SUBJ>}* [VP hélt sfg1eþ VP] [MWE_AdvP meira lheovm að cn segja sng MWE_AdvP] [AdvP áfram aa AdvP] [VPi að cn koma sng VPi] [PP til ae {*QUAL [NPg hans fpkee] *QUAL} PP] [MWE_SCP eftir ao að c MWE_SCP] *{*SUBJ> [NPn allt fohen] *SUBJ>}* [VPb var sfg3eþ VPb] [VPp orðið spghen VPp] [APn skrýtið lhensf] [CP og c CP] *{*SUBJ [NPn hún fpven] *SUBJ}* [VPp hætt spgven VPp] [VPi að cn hringja sng VPi] [PP í ao [NPa mig fp1eo] PP] ..

Final version - 1

- { *SUBJ> [NP pabbi nken NP] { *QUAL [NP hennar fpvee NP] *QUAL } *SUBJ> }
- [VPb var sfg3eþ VPb]
- { *COMP< [NPn vinur nken minn feken] *COMP< }
- [CP og c CP]
- { *SUBJ> [NPn við fp1fn] *SUBJ> }
- [VP tefldum sfg1fþ VP]
- [AdvP oft aa saman aa AdvP]
- [PP á ao [NP kvöldin nhfog NP] PP]
- [CP og c CP]
- { *SUBJ> [NP hann fpken NP] *SUBJ> }
- [VP studdi sfg3eþ VP]
- { *OBJ< [NP mig fp1eo NP] *OBJ< }
- [CP og c CP]
- { *SUBJ> [NP ég fp1en NP] *SUBJ> }
- [VP hélt sfg1eþ VP]

Final version - 2

- [MWE_AdvP meira lhevom að en segja sng MWE_AdvP]
- [AdvP áfram aa AdvP]
- [VPi að en koma sng VPi]
- [PP til ae [NP hans fpkee NP] PP]
- [MWE_SCP eftir að c MWE_SCP]
- { *SUBJ> [NP allt fohen NP] *SUBJ> }
- [VPb var sfg3eþ VPb]
- { *COMP< [VPp orðið sþghen VPp] *COMP< }
- { *COMP< [AP skrýtið lhensf AP] *COMP< }
- [CP og c CP]
- { *SUBJ [NP hún fpven NP] *SUBJ }
- { *COMP [VPp hætt sþgven VPp] *COMP }
- [VPi að en hringja sng VPi]
- [PP í ao [NP mig fp1eo NP] PP]
- ...

Syntactic constituents

- Accuracy in the marking of syntactic constituents
 - using either correct tags from the IFD or tags assigned by *IceTagger*

Phrase type	F-measure correct tags	F-measure <i>IceTagger</i>	Freq. in test data
AdvP	91.8%	85.1%	8.2%
AP	95.1%	86.3%	8.1%
APs	87.0%	68.6%	0.5%
NP	96.8%	93.0%	37.6%
NPs	80.4%	74.3%	1.5%
PP	96.7%	91.3%	13.0%
VPx	99.2%	93.8%	19.3%
CP	100.0%	99.6%	5.7%
SCP	99.6%	97.6%	3.4%
InjP	100.0%	96.3%	0.2%
MWE	96.9%	92.6%	2.5%
All	96.7%	91.9%	100.0%

Syntactic functions

- Accuracy in the marking of syntactic functions
 - using either correct tags from the IFD or tags assigned by *IceTagger*

Function type	F-measure correct tags	F-measure <i>IceTagger</i>	Freq. in test data
SUBJ	68.2%	47.6%	4.7%
SUBJ>	92.7%	89.4%	30.3%
SUBJ<	83.7%	75.1%	12.3%
OBJ	0.0%	0.0%	0.2%
OBJ>	43.5%	20.0%	0.8%
OBJ<	90.2%	78.2%	19.7%
OBJAP>	71.4%	57.2%	0.2%
OBJAP<	75.0%	46.2%	0.4%
OBJNOM<	30.8%	16.7%	0.6%
IOBJ<	73.3%	51.9%	0.9%
COMP	56.9%	40.0%	2.8%
COMP>	91.3%	91.3%	1.3%
COMP<	75.1%	70.0%	12.7%
QUAL	87.7%	77.9%	10.4%
TIMEX	74.7%	55.9%	2.7%
All	84.3%	75.3%	100.0%

Results

- The results are quite good
 - 96.7% accuracy in marking syntactic constituents
 - 84.3% accuracy in marking syntactic functions
- The results could be improved by
 - making more use of the morphological tags
 - build and apply various word lists
 - improve the transducers and add some new ones

eirikur@hi.is