

Kristín Bjarnadóttir

Beygingarlýsing íslensks nútímamáls

Kristín Bjarnadóttir

Beygingarlýsing íslensks nútímamáls

Útgáfa 1.0

Orðabók Háskólans og Edda hf. sóttu um styrk til verkefnisstjórnar menntamálaráðuneytisins í tungutækni árið 2002 til að vinna að beygingarlýsingu íslensks nútímamáls. Gengið var frá samningi um verkið 23. ágúst 2002 og því lauk 13. febrúar 2004 þegar menntamálaráðuneytinu var afhentur geisladiskur með útgáfu 1.0 af beygingarlýsingunni, alls 173.389 beygingardæmi á formi xml-skráa.

Markmiðið með verkefninu var að koma upp beygingarlýsingu á tölvutæku formi til nota í ýmiss konar tungutækniverkefni en ítarleg beygingarlýsing er grundvöllur að vélrænni greiningu á íslenskum textum, nauðsynlegur undanfari orðflokkgreiningar og setningagreiningar. Beygingarlýsingin nýtist t.d. við mörkun texta, við gerð leitarvéla, leiðréttingar- og þýðingarforrita, auk þess að vera forsenda skilvirkar orðabókargerðar og heimildasöfnunar um tungumálið.

Beygingarlýsingin er þegar notuð í ýmsum verkefnum Orðabókar Háskólans, t.d. IS-LEX, sem er samnorrænt verkefni um ís-

lensk-skandinavískar orðabækur, Markaðri íslenskrí málheild og við leit í Textasafni Orðabókarinnar. Þá er hún notuð í verkefninu Icelandic Online, sem er samstarfsverkefni Háskóla Íslands og Háskólans í Wisconsin.

Orðabók Háskólans sá að öllu leyti um vinnuna við beygingarlýsinguna og lagði til húsnæði, alla aðstöðu og aðgang að gögnum. Edda hf. lagði til beygingarlýsingu þá sem unnin var fyrir tölvuútgáfu *Íslenskrar orðabókar* (2002). Starfsmenn voru Kristín Bjarnadóttir (verkefnisstjóri), Þórdís Úlfarsdóttir, Auður Þórunn Rögnvaldsdóttir, Ragnhildur Hrönn Sigurðardóttir og Aðalsteinn Eyþórsson.

Orðaforðinn í beygingarlýsingunni er fenginn úr 3. útgáfu *Íslenskrar orðabókar* og úr söfnum Orðabókarinnar. Helstu heimildir við rannsóknir á einstökum orðum og beygingarflokkum eru Ritmálsskrá og Textasafn Orðabókarinnar, auk handbóka, greina og ritgerða um íslenska málfræði.

Útgáfa 2.0

Beygingarlýsing íslensks nútímamáls, útgáfa 2.0, hefur nú verið birt á vefsíðu Orðabókarinnar og er hún unnin upp úr útgáfu 1.0 með talsverðum endurbótum, leiðréttingum og nokkrum viðbótum.

Heildarorðaforðinn í útgáfu 2.0 skiptist svo:

Skipting beygingardæma eftir orðflokkum:			
	Orð	Orðmyndir	Meðalfjöldi orðmynda
Nafnorð	137.300	1.960.700	14,3
Lýsingarorð	26.300	2.334.900	88,8
Sagnir	7.600	538.200	70,8
Mannanöfn	4.760	19.445	4,1
Annað	70	849	12,1
Alls	176.030	4.854.094	27,6



Beygingardæmin

Takmarkið við birtingu beygingardæmanna er að einskorða efnið við raunverulegar myndir hvers orðs, þ.e. að sýna afbrigði þar sem það á við (t.d. í þágufalli eintölu af nafnorðinu *hnifur*: *hníf/hnífi*) og eyður þar sem beygingarmyndir koma aldrei fyrir (t.d. er ekki til miðmynd af sögninni *auðvelda*). Án afbrigða eru beygingarmyndir nafnorðs 16, þ.e. fjögur föll í eintölu og fleirtölu, án greinis og með greini. Beygingarmyndir sagnar í persónuhætti eru 48, auk boðháttar og lýsingarháttar, en alls geta beygingarmyndir hvernar sagnar orðið 106. Beygingarmyndir lýsingarorðs sem tekur stigbreytingu eru allt að 120, án afbrigða.

Hér er beygingardæmi nafnorðsins hnifur, eins og það birtist á vefsíðu Orðabókarinnar.

Þýðingarlýsing Orðabók Háskólans - Þýðingarmál - Þýðing - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Lýsingarorð

Þögull

Framtíðar stökkerfið er stökkerfið
líftíðar
Eftirtíðar stökkerfið er stökkerfið

Frasatíðar							
Stærk þýðing							
Stöð	Kaflar	Endar	Þýðing	Stöð	Kaflar	Endar	Þýðing
St.	þögull	þögul	þögul	St.	þögla / þögla	þögla / þögla	þögla
St.	þögla / þögla	þögla / þögla	þögla / þögla	St.	þögla / þögla	þögla / þögla	þögla
þg.	þögla / þögla	þögla	þögla / þögla	þg.	þögla / þögla	þögla / þögla	þögla / þögla
St.	þögla	þögla	þögla	St.	þögla	þögla	þögla

Vok þýðing							
Stöð	Kaflar	Endar	Þýðing	Stöð	Kaflar	Endar	Þýðing
St.	þögla / þögla	þögla / þögla	þögla / þögla	St.	þögla / þögla	þögla / þögla	þögla / þögla
St.	þögla / þögla	þögla / þögla	þögla / þögla	St.	þögla / þögla	þögla / þögla	þögla / þögla
þg.	þögla / þögla	þögla / þögla	þögla / þögla	þg.	þögla / þögla	þögla / þögla	þögla / þögla
St.	þögla / þögla	þögla / þögla	þögla / þögla	St.	þögla / þögla	þögla / þögla	þögla / þögla

Done

Þýðingarmyndum getur fjölgað talsvert þegar afbrigði eru sýnd, eins og sjá má af hluta þýðingardæmis lýsingarorðsins þögull.

frá notendum vefsíðunnar og er þeim tekið með þökkum. Eðli málsins samkvæmt lýkur verkefni af þessu tagi aldrei þar sem fengist er við lífandi mál.

Framtíðarverkefnið

Í tengslum við vinnu við markaða íslenska málheild er fyrirhugað að bæta talsverðu efni við þýðingarlýsinguna, t.d. örnefnum og fyrirtækja- og stofnanahæitum, auk viðbóta við almennan orðaforða. Þá berast ábendingar um viðbætur og leiðréttingar

Vegna tíðra endurbóta er mikilvægt að þeir sem nota þýðingarlýsinguna hafi samband við Orðabók Háskólans til að nálgast nýjustu útgáfu.