

## Um upphaf *Beygingarlýsingar íslensks nútímamáls*

Það er sérlega ánægjulegt að fá að segja nokkur orð um forsögu og upphaf *Beygingarlýsingar íslensks nútímamáls* sem hér er verið að opna í nýrri og endurbættri gerð. Ég stend nefnilega í þeirri meiningu að ég beri einhvern hluta ábyrgðarinnar á upphafi þessa verks. Kristín Bjarnadóttir var nemandi minn í fyrsta námskeiðinu sem ég kenndi í íslensku, veturinn 1982-1983. Á vormisseri var viðfangsefnið einkum beygingar- og orðmyndunarfræði – ég var þá að skrifa kennslubókina *Íslenska orðhlutafræði* sem ég tindi í nemendum jafnóðum. Ég er ekki frá því að áhuga Kristínar á orðum, gerð þeirra og hegðun, megi að einhverju leyti rekja til þessa námskeiðs. Beygingarlýsingin er náttúrulega barn Kristínar og sé þessi skilningur minn réttur má ég því kannski með vissum hætti kallast afi Beygingarlýsingarinnar.

Þetta var vorið 1983 og nú verð ég að hlaupa yfir 18 ár þótt ýmislegt hafi gerst á þeim árum sem segja mætti frá ef tími væri til. En í lok ágúst 2001 kom tilkynning á póstlista sem ég var á um að sænski máltækniháskólinn, *Swedish Graduate School of Language Technology*, sem tók til starfa þetta haust, byði tvö námskeið sem unnt væri að taka í fjarnámi með staðlotum. Nemendur utan Svíþjóðar gátu sótt um að sækja þessi námskeið, og það sem skipti öllu máli – það var hugsanlegt að fá styrk til ferða og upphalds úr norrænu máltækniáætluninni sem þá var í gangi. Við Kristín höfðum oft rætt sameiginlegan áhuga okkar á máltækni – sem var nefnd tungutækni í þá daga – og ég hafði því samband við hana til að segja henni frá þessu.

Um sama leyti fékk ég póst frá Sigrúnu Helgadóttur, sem þá var nýhætt störfum á Hagstofu Íslands, og vildi spyrjast fyrir um máltæknikennslu sem til stóð að hæfist við Háskóla Íslands. Ég sagði henni frá sænska skólanum og þrem vikum síðar vorum við fjórir Íslendingar – Kristín, Sigrún, Auður Þórunn Rögnvaldsdóttir og ég – komin til Gautaborgar á staðlotu í námskeiði sem hét „Natural Language Processing“. Við sátum svo sveitt við verkefnavinnu allt haustið en námskeiðinu lauk með ráðstefnu í Växjö í janúar 2002. Þetta námskeið varð afdrifaríkt og markaði í raun upphaf reglulegs akademísks máltæknistarfs á Íslandi því að við höldum öll áfram að vinna á þessu sviði sem við hefðum varla gert án námskeiðsins, a.m.k. ekki Sigrún sem var einn helsti burðarás íslensks máltæknistarfs næstu 15 ár. Auk þess komumst við þar í kynni við ýmislegt fólk sem hjálpaði okkur af stað og við höfum síðan verið í meira og minna sambandi við.

Þetta sama haust, 2001, var svo auglýst í fyrsta skipti eftir umsóknum um styrki úr tungutækniverkefni menntamálaráðuneytisins. Við fjögur, sem kölluðum okkur Málgreiningarhópin eftir heimkomuna frá Gautaborg og verkefnavinnu í námskeiðinu, vorum svo uppveðruð og áhugasöm að við áttum frumkvæði að þremur umsóknum sem sendar voru inn um miðjan desember. Við fengum Orðabók Háskólans í lið með okkur sem lá beint við – gögnin sem við ætluðum að byggja á voru þar, Kristín vann þar, og ég var stjórnarformaður Orðabókarinnar um þær mundir. Vorið 2002 ákvað svo menntamálaráðherra að styrkja tvær þessara umsókna.

Önnur þeirra umsókna sem fengu styrk var um gerð þjálfunarlíkans fyrir íslenskan markara og byggðist á því að í iðrum tölvukerfis Orðabókarinnar leyndust grunnskrár sem gerðar höfðu verið fyrir *Íslenska orðtíðnibók* áratug áður. Það var hrein heppni að þessar skrár höfðu varðveist því að enginn sá gagnsemi þeirra fyrir, en þær urðu í raun undirstaða allrar vélrænnar málfræðigreiningar á íslensku næstu 15 árin – og að nokkru leyti enn. Þessi umsókn var send í nafni Orðabókar Háskólans og Málgreiningarhópsins. Umsóknin sem ekki fékk brautargengi í þetta skipti var um gerð íslenskrar málheildar – það verkefni fékk ekki styrk fyrr en þremur árum síðar og varð á endanum að *Markaðri íslenskri málheild*. Þess má geta að orðin *mark*, *markari*, *mörkun* og *málheild* urðu öll til við samningu þessara umsókna um miðjan desember 2001.

Hin umsóknin sem fékk styrk var send í nafni Orðabókar Háskólans og Eddu útgáfa, en rík áhersla var lögð á að fyrirtæki tækju þátt í verkefnunum til að tryggja að þau nýttust við gerð máltækniþúnaðar. Sú umsókn var hugarsmíð Kristínar sem hafði lengi haft áhuga á að vinna slíkt verk en þarna komu loks forsendur til þess. Kristín var vitanlega búin að vinna að beygingarlegri greiningu og lýsingu árum saman, einkum í tengslum við hið svokallaða norræna verkefni hjá Orðabókinni sem var undanfari ISLEX og svo við endurskoðun *Íslenskrar orðabókar* á vegum Máls og menningar. Hún samdi því þessa umsókn að mestu, í einhverju samráði við Mörd Árnason sem þá var ritstjóri *Íslenskrar orðabókar*. Kristín hafði þá þegar lagt drög að beygingarlýsingunni, komið upp miklu hráefni í hana og hafði nokkuð mótaðar hugmyndir um hvað hún vildi gera. Í umsókninni frá því í desember 2001 koma megininkenni verksins fram en þar segir í byrjun:

„Sótt er um styrk til gerðar beygingarlýsingar íslensks nútímamáls. Markmið er að koma upp beygingarlýsingu á tölvutæku formi sem birt verður í tölvuútgáfu *Íslenskrar orðabókar* [...] frá Eddu hf. og á vefsíðu Orðabókar Háskólans [...]. Aðgangur að beygingarupplýsingunum í formi beygingardæma yrði öllum frjáls án endurgjalds á heimasíðu Orðabókarinnar. Jafnframt yrði til tölvutæk beygingarlýsing sem yrði fól til nota í verkefni af ýmsu tagi.“

Þarna var sem sé þegar í upphafi lögð áhersla á hið tvíþætta gildi Beygingarlýsingarinnar – annars vegar sem uppflettigagn fyrir almenning og hins vegar sem gagnasafn til nota í máltækni-verkefnum.

Þetta var haustið 2001 og nú eru aftur liðin 18 ár. Á þeim tíma hefur margt gerst sem ég ætla ekki að rekja en læt nægja að nefna að Kristín hefur verið óþreytandi í að byggja Beygingarlýsinguna upp, auka og bæta. Það hefur skilað sér í því að Beygingarlýsingin hefur fullkomlega staðið undir þeim vonum sem við gerðum okkur í upphafi og þeim fyrirheitum sem gefin voru, enda er hún annars vegar eitt mikilvægasta og mest notaða stuðningsgagn íslenskra málnotenda, og hins vegar ein meginundirstaða þróunar í íslenskri máltækni. Sú staða mun enn styrkjast með þessari nýju gerð.