

# Nýsköpunarsjóður námsmanna

## Tæknileg lokaskýrsla

Tæknilegri lokaskýrslu skal skila rafrænt, ásamt lokaskýrslu verkefnis, til Rannís, á tölvupóstfangið [nsn@rannis.is](mailto:nsn@rannis.is).

Styrknúmer: 207514-0091

Heiti verkefnis: Röð lýsingarorða á undan nafnorði í íslensku

Nemandi/nemendur og kt.: Gunnar Pálsson, 011293-2639

Umsjónarmaður/menn og kt.: Hrafn Loftsson, 140465-5509

### Undirskrift umsjónarmanns:

Námsmaður/námsmenn hafa náð settum markmiðum. Niðurstöður liggja fyrir og eru fram settar í skýrslunni sem unnin hefur verið af námsmanni/námsmönnum.

Undirskrift umsjónarmanns: \_ Reykjavík, 21.09.2020 \_

Staður og dagsetning



Umsjónarmaður

### Afgreiðsla Rannís:

Móttekin þann		Skýrsla afgreidd þann	
Er lokaskýrsla fullnægjandi?		Er heimilt að borga út samkvæmt samningi?	
UNDIRSKRIFT STARFSMANNS RANNÍS			

## Tæknileg lokaskýrsla – útfyllt af nema/nemum:

Í lokaskýrslu skulu koma fram upplýsingar um verkefnið, nemendur, sjálfstæði þeirra og hlutverk í verkefninu, markmið verkefnis og hagnýtingu þess, nýsköpunargildi, hvernig samstarfi í verkefninu var háttað (ef við á) og hverjar niðurstöður þess eru.

Tæknileg lokaskýrsla má að hámarki vera 5 bls., fjalla skal um alla neðangreinda þætti og skal hver þáttur vera ½ - 1 bls. að lengd.

### 1. Útdráttur (hámark ½ bls.)

Þegar fleiri en eitt lýsingarorð koma á undan nafnorði þurfa þau að mynda röð og svo virðist sem að ákveðin röð lýsingarorða hljómi eðlilegri en önnur. T.d. finnst flestum að “Stóri, rauði bíllinn” hljómi eðlilegra en “Rauði, stóri bíllinn”. Þessi röð lýsingarorða virðist vera innbyggð í okkur öllum og er röðin hin sama hjá langflestum.

Engin rannsókn hefur verið framkvæmd á röð lýsingarorða í íslensku, einungis á stærri tungumálum eins og ensku og frönsku. Í þeim rannsóknum kemur fram að merkingarlegur flokkur lýsingarorðs (fyrirbærið sem orðið lýsir) er helsti ákvörðunaliður varðandi röð lýsingarorðanna.

Í þessu verkefni er gerð fyrsta ítarlega rannsókn á röð lýsingarorða í íslensku. Afurðin er í formi tölfræðilegra gagna sem geta nýst í ýmsar frekari rannsóknir, s.s. í samanburði á lýsingarorðaröð í íslensku og öðrum tungumálum.

### 2. Þáttur nema og sjálfstæði

Nemi vann mjög sjálfstætt að verkefninu og hugmynd að verkefninu kom frá honum. Nemi vann sjálfstætt í forritun, uppsetningu málfanga, framsetningu gagna, og túlkun niðurstaðna. Helsta hlutverk umsjónarmanns var að leiðbeina nema á reglulegum stöðufundum, m.a. að svara tæknilegum spurningum, staðfesta stöðu verkefnisins, mæla með málföngum, og leggja fram tillögur um hvað nemi skyldi einbeita sér að. Lokaskýrslan var skrifuð af nema og yfirfarin og endurbætt af umsjónarmanni.

### 3. Hagnýting og markmið

Samkvæmt sambærilegum rannsóknum á öðrum tungumálum, s.s. J.E. Martin (1969) fyrir ensku og L.R. Waugh (1977) fyrir frönsku, er hægt að skilgreina ákveðna röð lýsingarorða á undan nafnorði og gildir hún að mestu leyti þvert á tungumál. Tilgangur þessa verkefnis er að sýna að þessi röð eigi einnig við fyrir íslensku.

Miðað við önnur stærri tungumál er íslenska ekki langt komin þegar kemur að tölfræðilegri gagnavinnslu. Helsta markmið þessa verkefnis er að vera skref í átt að betri auðlindaöflun fyrir íslenska tungu. Verkefnið skiptist í þrjú hlutverkefni og birtast hagnýtar niðurstöður í hverju þeirra sem nýtast öðrum verkefnum í framtíðinni. Sem dæmi um hagnýtt gildi verkefnisins er hægt að nefna fólk sem hefur íslensku ekki að móðurmáli. Niðurstöður verkefnisins gefa skýrari mynd af því hvernig skal beita tungunni á náttúrulegri hátt. Eins og stendur er engin góð leið til að útskýra fyrir íslenskunemum af hverju ein lýsingarorðaröð hljómar eðlilegri en önnur.

### 4. Nýsköpunargildi

Öll þrjú hlutverkefni skila niðurstöðum sem er hægt að nýta á ýmsa vegu. Fyrsta og annað verkefnið skila tæmandi lista af merkingarlegum flokkum sem ná yfir allan núverandi lýsingarorðaforða og hvaða



flokki hvert og eitt lýsingarorð tilheyrir. Þessi greining hefur ekki verið framkvæmd fyrir íslensku og reyndar í litlum mæli fyrir önnur tungumál. Eina íslenska verkefnið sem gerir sambærilega greiningu er Íslenskt orðanet sem flokkar saman íslensk orð miðað við merkingu að hluta til.

Í þriðja hlutverkefninu koma fram nýjar tölfræðilegar upplýsingar um málnotkun Íslendinga. Í þessu verkefni eru þær notaðar til að svara ýmsum forvitnilegum spurningum en auðvelt væri að nýta gögnin sem fást úr öllum þremur hlutverkefnunum í aðrar fjölbreytilegar málfræðisóknir s.s. í sögulegu eða mannfræðilegu samhengi.

## 5. Samstarf, ef við á (hámark ½ bls.)

Þetta verkefni var einstaklingsverkefni unnið af einum nema. Þessi þáttur á því ekki við.

## 6. Niðurstöður

71 merkingarlegir flokkar lýsingarorða voru fengnir með því að þýða niðurstöður úr GermaNet, þýsku verkefni sem byggir á merkingarlegri flokkun orða. Yfir tvær milljónir lýsingarorða voru flokkuð í þessa 71 flokka. Orðin eru flokkuð samkvæmt samanburði við valin orð sem eiga heima í hverjum flokki. Sérhvert orð flokkast í þann flokk sem það á mest sameiginlegt með, m.t.t. hinna völdu orða.

Með tölfræðilegri greiningu er hægt að gefa hverjum flokki stuðul sem ákvarðar hvernig orðið raðast. Flokkar með hærri stuðul birtast fyrr í röðinni og lægri stuðul seinna. Stuðlarnir eru á bilinu 0,102–4,954 þar sem flokkurinn *fjöldi* hefur hæsta stuðulinn, og skal þar með birtast fyrst í röð lýsingarorða, og flokkurinn *ríki* hefur þann lægsta. Þessir flokkar eru þó einstök tilvik og eru flestir flokkar, 37 samtals, með stuðul á bilinu  $1 \pm 0,25$ . Heildarfjöldi orða sem flokkuðust í hvern flokk er fremur breytilegur: Flokkurinn með lægsta fjölda orða er *veður* með 1851 orð og flokkurinn með hæsta fjölda orða er *hegðun* með 204.798 orð.

Í lokin er byggð endanleg röð á öllum 71 flokki og byggist röðin á stuðlunum. Þessi röð á að virka þannig að lýsingarorð sem eru í þessari röð hljóma eðlilegri en þau sem eru ekki í þessari röð. Með því að beita þessari röð á öll 971.560 tilvik sem finnast í Risamálheildinni finnst að 655.954 þeirra eða 67,5% eru í réttri röð. Þetta er sambærilegt spálíkaninu í (Wulff, 2003) sem fékk 73,5% tilvika í réttri röð fyrir ensku og gaf viðmið fyrir þetta verkefni

## 7. Undirskrift nema/nemenda

Staður og dagsetning: Reykjavík, 21.09.2020

Undirskrift nema/nemenda: