

Röð lýsingarorða á undan nafnorði í íslensku

Gunnar Pálsson og Hrafn Loftsson
Mál- og raddtæknistofa
Háskólinn í Reykjavík
gunnar19@ru.is, hrafn@ru.is

Útdráttur

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 eðlilegar 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 ákvörðunarliður 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.

1 Inngangur

1.1 Fyrri rannsóknir

Vísindalegar rannsóknir á röð lýsingarorða eiga sér langa sögu. (Martin, 1969) er ein fyrsta greinin sem skoðar röð lýsingarorða á undan nafnorði. Þessi grein virkar sem upphafspunktur þessa ákveðna sviðs og höfund ritgerðarinnar, J.E. Martin, mætti kalla upphafsmann sviðsins. Á þessum tíma skrifaði hann ýmsar ritgerðir um lýsingarorðaröð, að mestu í geðfræðilegu samhengi og eingöngu í ensku.

(Waugh, 1976) fjallar m.a. um brót á gefnum málfræðireglum, þ.á.m. röðunarreglunni, og af hverju það gerist. Þar kemur fram að helsta ástæða fyrir broti á röðunarreglunni er til að leggja áherslu á ákveðið lýsingarorð. Það lýsingarorð, sem skal fá mesta áherslu, færast fremst í röðina þrátt fyrir þann lýsingarorðaflokk sem það tilheyrir.

(Waugh, 1977) byrjar fyrstur umfjöllun um lýsingarorðaröð utan ensku, þ.e. frönsku í þessu tilfelli. Með því að skoða frönsku ásamt ensku breytist umfangið talsvert vegna þess að þessi tvö tungumál virðast hafa mismunandi lýsingarorðaraðir. Til dæmis setur franskan ákveðin lýsingarorð á eftir nafnorði eins og í þessu dæmi:

The big, red balloon (Stóra rauða blaðran)

Le grande ballon rouge (Stóra blaðran rauða)

(Byrne, 1979) kynnti ýmsum þátttakendum spurningar um hvaða röð lýsingarorða passar best, þ.á.m. þar sem lýsingarorðin passa alls ekki saman, t.d.:

A slow fast dog (Hægur fljótur hundur)

Niðurstöðurnar studdu þá tilgátu að fólk er gjarnan sammála um hvaða röð lýsingarorða er réttust. Einnig kemur fram að fólk hefur innbyggðan hæfileika að bæta við og endurraða orðum til að skapa samhengi sem betur útskýrir tilfelli þar sem orð virðast skarast. T.d. bættu þátttakendur samtengingu við fyrra dæmið:

A slow but fast dog (Hægur en þó fljótur hundur)

(Wulff, 2003) framkvæmdi rannsókn þar sem lýsingarorðaröð í málheildum er skoðuð og línulegri aðskilnaðargreiningu var beitt á gögnin. Þær upplýsingar eru síðan notaðar til að byggja spálíkan fyrir röð lýsingarorða. Líkanið nær 73.5% nákvæmni sem gefur okkur viðmið um hversu mikill hluti dæma ætti að fylgja röðunarreglu.

1.2 Tilgátur

Verkefnið byggir á flokkun lýsingarorða í merkingarlega flokka sem ákvarðast af fyrirbærinu, sem lýsingarorðið lýsir (litur, stærð, þjóðerni, o.s.frv.). Unnið er út frá þeirri tilgátu að öll lýsingarorð geti flokkast í a.m.k. einn merkingarlegan flokk. Önnur tilgáta er að ekki þurfi fleiri en 100 mismunandi flokka þannig að öll lýsingarorð geti flokkast í viðeigandi flokk.

Rannsóknir á stærri tungumálum hafa leitt í ljós ýmsa eiginleika sem hafa áhrif á lýsingarorðaröðun þvert á tungumál, t.d. áhrif breytileika lýsingarorða á röð þeirra¹ og OSASCOMP-reglan sem er stöðluð þumalputtaregla í ensku². Tilgáta er að íslenska fylgi þessum reglum líkt og önnur tungumál og ástæður lýsingarorðaröðunar í íslensku séu sambærilegar og í öðrum tungumálum.

2 Aðferðarfræði

Vinnsla verkefnisins er þríþætt: Fyrst eru merkingarlegir flokkar lýsingarorða myndaðir, síðan eru lýsingarorð flokkuð í lýsingarorðaflokka, og að lokum eru tölfræðilegar rannsóknir gerðar á gögnunum. Allir þrír þættir eru að mestu gerðir frá grunni, þar sem ekki finnast nein dæmi um að einhverjir þeirra hafa verið framkvæmdir fyrir íslensku áður.

2.1 Skilgreining flokka

GermaNet (Hamp & Feldweg, 1997) er þýsk útgáfa af WordNet og gefur góðan grunn á merkingarlegri flokkun. Í GermaNet eru skilgreindir 76 lýsingarorðaflokkar og geta þeir nýst fyrir íslenska flokkagreiningu vegna mikilla líkinda í málfræðilegri uppbyggingu þýsku og íslensku. Hver flokkur var þýddur yfir á íslensku ásamt tveimur lýsingarorðum sem hegða sér sem prýðisdæmi á orði sem passar í flokkinn. Í GermaNet eru allir flokkarnir hluti af yfirflokki³. Tölfræðilegar rannsóknir verkefnisins munu ekki taka tillit til yfirflokka en þeir eru hluti af flokkaröðunartöflunni í Viðauka 1.

Fimm flokkanna í GermaNet eru strangt tiltekið ekki merkingarlegir flokkar, heldur almennir flokkar fyrir orð sem ekki passa í neinn annan flokk. Þessir flokkar eru: *samanburður*, *mat*, *einkenni*, *afleiðsla* og

¹Breytileiki lýsingarorða vísar í hversu auðvelt er að breyta eiginleikanum sem er lýst, t.d. hefur litur meiri breytileika en þjóðerni þar sem auðveldara er að breyta lit hlutar en þjóðerni.

²OSASCOMP-reglan segir að lýsingarorð raðast eftir skammstöfuninni OSASCOMP, hún stendur fyrir Opinion[Sköðun], Size[Stærð], Age[Aldur], Shape[Lögun], Colour[Litur], Origin[Uppruni], Material[Efni], Purpose[Tilgangur].

³Yfirflokkar eru 13 samtals: *skyn*, *staðsetning*, *tími*, *hreyfing*, *efni*, *veður*, *líkami*, *skap*, *hugur*, *hegðun*, *samfélag*, *fjöldi* og *tengsl*. Þrír yfirflokkar voru fjarlægðir í endurbættri útgáfu: *almenn*, *afleiðsla* og *neitun*.

neitun. Uppbygging algrímsins í þessu verkefni veldur því að þessir fimm flokkar eru ekki gagnlegir og vegna þess eru flokkarnir fjarlægðir í endurbættri útgáfu verkefnisins. Lokafjöldi merkingarlegra flokka í verkefninu eru þá samtals 71.

Einn þessara 71 flokka er *tala* – sá inniheldur lýsingarorð sem tákna ákveðinn fjölda eins og “tveir” og “þrír”. Í íslensku eru þessi orð gjarnan talin vera töluorð sem er undirflokkur fallorða. Fyrir þetta verkefni munu töluorð hins vegar teljast sem lýsingarorð þar sem þau eru gjarnan flokkuð sem lýsingarorð í Risamálheildinni (Steingrímsson et al., 2018) og flokkar GermaNets gera nú þegar ráð fyrir þeim.

2.2 Flokkun lýsingarorða

Vegna merkingarlegs eðlis lýsingarorðaflokkunar er þessi hluti byggður á fastText (Bojanowski et al., 2017) sem er aðferð til byggingar orðavigra (e. word embeddings) og textaflokkunar og er þróuð af gervigreindarrannsóknarstofu Facebook. fastText var valið vegna þess að það er frítt, opið og öflugt. Virkni fastText í þessu verkefni er að taka inn mikinn texta úr Risamálheildinni⁴ og hvert orð sem finnst fær vigur (af rauntölum) sem táknar merkingarlegt innihald orðsins.

Grunnvirkni flokkunarinnar er að hver flokkur inniheldur nokkur dæmisorð sem eru góð dæmi um orð sem passa í flokkinn (t.d. “rauður” og “blár” í flokknum *litur*) og öllum þessum orðum er breytt í 100-víða vigra með fastText-líkani sem er þjálfað á Risamálheildinni⁵. Þá er einfaldlega hægt að flokka hvaða lýsingarorð sem er með því að breyta því í vigur með hjálp fastText-líkansins, bera vigurinn saman við vigra dæmisorðanna með því að finna kósínus-líkendi⁶ vigranna og flokka orðið í flokk dæmisorðsins sem það líkist mest, þ.e. flokk dæmisorðsins sem átti hæstu kósínus-líkendin með orðinu.

Upprunalega voru einungis tvö dæmisorð fyrir hvern flokk, 152 dæmisorð samtals. Að bæta við fleiri dæmisorðum eru bestu og einföldustu endurbætur, sem hægt er að gera í þessu skrefi. Það bætir nákvæmni flokkunarinnar og lækkar tíðni misflokkunar. Í endurbættri útgáfu eru mismörg dæmisorð fyrir alla flokka. Flokkurinn með flest dæmisorð er *litur* með 10 dæmisorð. Nokkrir flokkar fengu þó ekki fleiri dæmisorð og halda sömu tveimur dæmisorðum. Fjöldi dæmisorða í endurbættu útgáfunni eru samtals 260.

Í raun væri hægt að hafa hundruð eða þúsund dæmisorða. Fyrir þetta verkefni er fjöldi dæmisorða hins vegar hafður í lægri kantinum til að kóðinn sé sem skilvirkastur. Með fleiri dæmisorðum hækka keyrslutími algrímsins og líkanið verður nær því að vera uppflattitafla í stað spálíkans.

2.3 Tölfræðilegar rannsóknir

Úr Risamálheildinni eru fengin öll tilvik tveggja eða fleiri lýsingarorða á undan nafnorði og þeim safnað saman. Í Risamálheildinni eru 971.560 mismunandi tilvik af þessu tagi. Öllum lýsingarorðum er síðan skipt út fyrir merkingarlegan flokk þannig að röðun þeirra geymist. Þá eru talin tilvik hvenær orð í einum flokki birtist á undan orði í öðrum flokki og þær niðurstöður settar upp í töflu. Þessa töflu mætti kalla lokaniðurstöðu algrímsins og hana er hægt að nýta til að undirbúa gögn fyrir ýmsar tölfræðilegar rannsóknir. Hluti töflunnar sést í töflu 1.

⁴Það skal tekið fram að allur texti úr Risamálheildinni er lemmaður áður en unnið er með hann svo ekki þurfi að vinna með mismunandi beygingarleg form orða.

⁵Til dæmis er vigur orðsins “góður” = [0.03521131, -0.39758083, ..., 0.11212852, -0.13969861] með 100 gildi samtals.

⁶Kósínus-líkendi er formúla sem reiknar líkendi tveggja vigranna. Niðurstaðan er gildi milli 0 og 1. Í þessu verkefni er hægt að segja að kósínus-líkendi tveggja vigranna er mælikvarði á því hversu lík orðin eru, þar sem 0 er “algjörlega ólíkt” og 1 er “algjörlega líkt”.

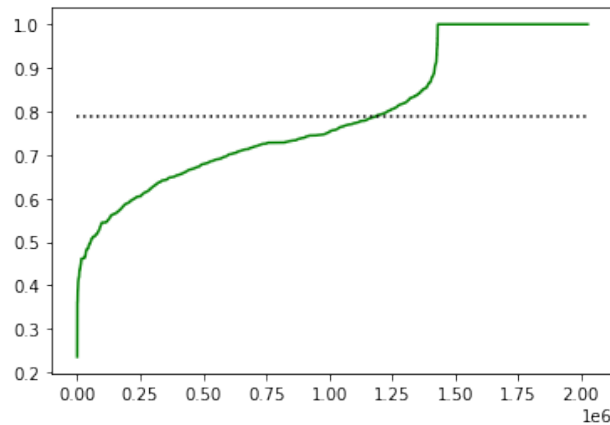
	birta	litur	hljóð	bragð
birta	806	481	68	48
litur	295	6780	28	96
hljóð	50	40	58	9
bragð	28	112	13	182
lykt	67	168	9	22
áferð	408	736	57	106
vídd	771	2999	136	465

Tafla 1: Úrtak úr niðurstöðutöflunni. Hér sést að orð í flokknum *litur* birtast 295 sinnum á undan orðum í flokknum *birta*.

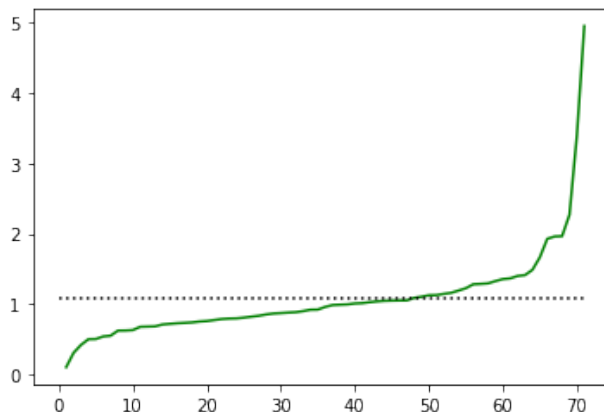
Einn helsti tilgangur verkefnisins er að byggja endanlega röð lýsingarorðaflokka og til að finna hana eru gögnin úr töflunni nýtt til að byggja röðunarstuðul fyrir hvern flokk.

3 Niðurstöður

Í heildina voru yfir tvær milljónir lýsingarorða flokkuð og gildi hæsta kósínus-líkindis allra orðanna vistað til að finna meðalgildi líkinda. Þetta gildi er hægt að líta á sem nákvæmni flokkunarinnar. Í fyrstu útgáfu verkefnisins var þetta meðalgildi 0,752. Í endurbættri útgáfu með færri flokka og fleiri dæmisorð hækkaði þetta gildi í 0,789. Dreifing allra kósínus-líkinda sést á grafi 1. Orð með kósínus-líkindi=1 eru þau orð sem eru nú þegar dæmisorð og fremur fá orð hafa kósínus-gildi á bilinu 0.9 til 1 vegna þess að flest orð með þau gildi voru nýtt sem ný dæmisorð í byggingu endurbættrar útgáfu.



Graf 1: Dreifing kósínus-líkinda orða. Punktalínan sýnir meðalgildi kósínus-líkinda (≈ 0.789). Gildin á x -ásnum eru í milljónum.



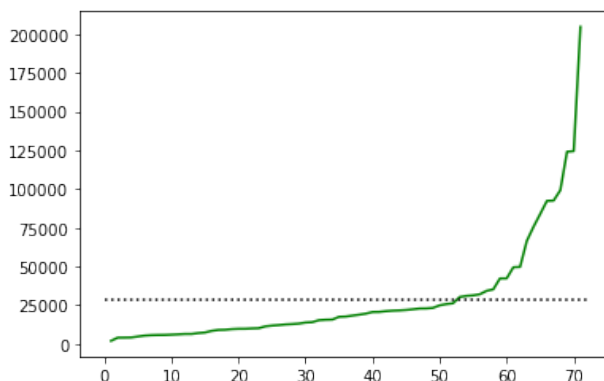
Graf 2: Dreifing stuðla flokka. Punktalínan sýnir meðalgildi stuðlanna (≈ 1.076)

Öll flokkuð orð eru nýtt í tölfræðilega greiningu og útreikningi á röðunarstuðlum. Stuðull flokks er fundinn með eftirfarandi formúlu:

$$\begin{aligned} \text{Stuðull flokks} = & \\ & \frac{\text{Fjöldi orða í öðrum flokkum sem birtast á eftir flokksorðum}}{\text{Fjöldi orða í öðrum flokkum sem birtast á undan flokksorðum}} \end{aligned}$$

Samkvæmt þessum stuðlum eiga orð í flokkum með hærri stuðul að birtast á undan orðum í flokkum með lægri stuðul til að röðun lýsingarorða hljómi sem náttúrulegust. Með stuðlunum er þar með hægt að raða flokkunum 71 í þá röð, sem hljómar best í eyrum okkar. Stuðlarnir eru á bilinu 0,102–4,954 þar sem flokkurinn *fjöldi* hefur hæsta stuðulinn, og skal þar með birtast fyrstur í röð lýsingarorða og flokkurinn *ríki* hefur þann lægsta. Þessir flokkar eru þó sérstök tilfelli og eru flestir flokkar, 37 samtals, með stuðul á bilinu $1 \pm 0,25$. Dreifing stuðla sést á grafi 2.

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. Hafa skal í huga að þetta er ekki fjöldi mismunandi orða heldur fjöldi allra orða. Dreifing orðafjölda í flokkum sést á grafi 3.



Graf 3: Dreifing orðafjölda í flokkum. Punktalínan sýnir meðalfjölda orða í flokkum (≈ 28540)

Unnið er með þá forsendu að hvert einasta lýsingarorð hafi einn ákveðinn flokk sem það passar best í. Þetta þýðir að orð með fleiri en eina merkingu eru einungis flokkuð samkvæmt einni merkingu þess. Í verkefninu er unnið með algengustu eða bókstaflegustu merkingu orða, t.d. er orðið “sætur” flokkað í *bragð*

í stað *útlit*. Með þessari aðferðafræði flokkast algeng orð með margar mismunandi skilgreiningar í einn flokk. Þetta var talið vera betri kostur en að nýta almennum flokka GermaNets, sem hefði geta valdið mikilli misflokkun lýsingarorða. T.d. var ákveðið að flokka orðið “góður” í *hegðun* þrátt fyrir margar aðrar mögulega flokka. Þetta veldur því að *hegðun* inniheldur mikið magn af orðum sem mætti réttlæt看lega vera í öðrum flokki.

Lista og röð flokka í heild sinni má sjá í Viðauka 1 í lok skýrslunnar.

4 Umræða

Tilgáturnar tvær um flokkun lýsingarorða⁷ reyndust sannar. Aðferðin, sem var notuð í verkefninu, veldur því að lýsingarorð geta ekki verið óflokkuð og ekki finnast rannsóknir sem vísa til þess að til eru lýsingarorð sem geta talist óflokkanleg. Þetta þýðir að öll lýsingarorð geti sannarlega flokkast í a.m.k. einn merkingarlegan flokk. Með GermaNet var hægt að flokka öll lýsingarorð í 76 flokka.

Ef gefið er að fengin lýsingarorðaröð sé rétt, kemur í ljós að 655.954 tilvik úr Risamálheildinni eru í réttri röð af þeim 971.560 sem fundust. Þetta eru um 67,5% sem fylgja reglunni sem búin er til af niðurstöðum tölulegrar greiningar. Þetta er sambærilegt spálíkaninu í (Wulff, 2003) sem fékk 73,5% tilvika í réttri röð fyrir ensku og gaf viðmið fyrir þetta verkefni. Þetta gæti verið merki um að til er almennt hlutfall tilvika sem fylgja lýsingarorðaröð þvert á tungumál og að það hlutfall sé í kringum 70%. Þetta styrkir líka þá tilgátu að lýsingarorðaröð í íslensku sé sambærileg öðrum tungumálum. Í töflu 2 sjást 5 orðasambönd í réttri röð og 5 orðasambönd í rangri röð, öll valin af handahófi.

Orðasambönd í réttri röð		
fyrstur	fyndinn	myndband
sterkur	hugrakkur	frænka
nýr	íslenskur	efnahagskerfi
margvíslegur	óbeinn	afleiðing
lítill	náttúrulegur	auðlind
Orðasambönd í rangri röð		
evrópskur	listrænn	kvikmynd
finn	fyrstur	verk
magnaður	feitur	ilmur
grillaður	brenndur	matur
samsettur	áhugaverður	maður

Tafla 2: Handahófsvalin orðasambönd í réttri og rangri röð. Ath. að búíð er að breyta orðunum í lemmur

Eins og sést í dreifingu stuðla er mismikið bil milli samliggjandi flokka í röðuninni. Bil milli stuðla tveggja flokka er túlkað sem mæling á því hversu óeðlileg röðin hljómar ef röð flokkanna er röng. T.d. eru orðasamböndin “kringlóttur sléttur bolti” og “súr fersk sítróna” bæði í rangri röð samkvæmt röðun en hægt er að áætla að “súr fersk sítróna” sé óeðlilegri en “kringlóttur sléttur bolti” með því að skoða mismun stuðla. “kringlóttur” er í flokknum *lögun* sem hefur stuðulinn 0,539 og “sléttur” er í flokknum *áferð* með 1,051 – mismunur flokkana er þá 0,512. “súr” er í flokknum *bragð* sem hefur stuðulinn 0,884 og “fersk” er í flokknum *þroski* sem hefur stuðulinn 1,027 – mismunur flokkana er þá 0,143.

Erfitt er að sýna fram á hvort gefin röðun raðast samkvæmt breytileika lýsingarorða vegna þess að erfitt er að sýna fram á breytileika merkingarlegu flokkana, þar sem þeir lýsa mjög mismunandi fyrirbærum. Ekki

⁷Öll lýsingarorð geta flokkast í a.m.k. einn merkingarlegan flokk; það þarf færri en 100 merkingarlega flokka til að flokka öll lýsingarorð vel

verður reynt að sýna fram á þetta í þessu verkefni. Auðveldara er að sýna að OSASCOMP-reglan gildir ekki jafnvel fyrir íslensku og ensku. Fyrsti flokkurinn í OSASCOMP-röðinni, *Opinion/Skoðun*, gildir um marga íslensku flokkana á við og dreif í röðinni og síðustu tveir, *Material/Efni* og *Purpose/Tilgangur* hegða sér öðruvísi í íslensku. Þau eru lýsingarorð í ensku en oftast sem hluti samsetts nafnorðs í íslensku (swimming pool = sundlaug; paper envelope = pappírsumslag).

Formúlan sem finnur stuðla flokka er mjög einföld og er það viljandi. Með góðum skilyrðum gerir hún það sem hún á að gera: að gefa flokkum stuðul sem stendur fyrir stöðu flokksins í lýsingarorðaröðinni. Þetta þýðir ekki að formúlan sé fullkomin. Það varð skýrt að formúlan er næm fyrir fjölda orða í flokksnum svo að mikið magn orða í flokki getur sprengt stuðulinn. Þetta sést með flokkunum með hæstu og lægstu stuðlana: *fjöldi* og *ríki*. Fjöldi orða í báðum flokkum eru langt yfir meðaltali og eru stuðlar beggja flokka jáðartilvik. Gert er ráð fyrir að þessir eiginleikar séu beintengdir. Tilgáta varð til úr þessu að flóknari stuðlaformúla gæti verið nákvæmari og öruggari fyrir jáðartilvikum.

Að lokum er vert að benda á að röð lýsingarorða er skoðuð í algjöru samhengisleysi. Þetta þýðir að í sumum tilvikum er röð lýsingarorða tæknilega vitlaus en er rétt í samhengi setningarinnar. T.d. finnst orðaröðin “nýkjörni fyrsti þingmaður” í Risamálheildinni. Með engu samhengi ætti að vera augljóst að “fyrsti nýkjörni þingmaður” hljómi eðlilegar. Í pólitísku samhengi hins vegar er upprunalega röðin réttari vegna þess að “fyrsti þingmaður” er nafn á ákveðnu fyrirbæri og orðin tvö eiga að standa saman. Það er ómögulegt að gera ráð fyrir öllum mögulegum samhengjum svo unnið var með algjört samhengisleysi. Þetta veldur því að þau 32,49% tilvika sem teljast vera í rangri röð innihalda líka þau tilvik þar sem samhengi breytir röð lýsingarorða.

Heimildir

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135-146.

Byrne, B. (1979). Rules of prenominal adjective order and the interpretation of “incompatible” adjective pairs. *Journal of Verbal Learning and Verbal Behavior*, 18(1), 73-78.

Hamp, B., & Feldweg, H. (1997). Germanet-a lexical-semantic net for german. In *Automatic information extraction and building of lexical semantic resources for NLP applications*.

Martin, J. E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior*, 8(6), 697-704.

Steingrímsson, S., Helgadóttir, S., Rögnvaldsson, E., Barkarson, S., & Guðnason, J. (2018, May). Risamálheild: A very large Icelandic text corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Wagh, L. R. (1976). The semantics and paradigmatics of word order. *Language*, 82-107.

Wagh, L. R. (Ed.). (1977). A semantic analysis of word order: Position of the Adjective in French (Vol. 1). Brill Archive.

Wulff, S. (2003). A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics*, 8(2), 245-282.

Viðauki 1: Merkingarlegir flokkar

Eftirfarandi listi inniheldur alla merkingarlega flokka lýsingarorða í minnkandi röð stuðla þeirra, þ.e.a.s. í þeirri röð sem hljómar sem eðlilegust. Ásamt flokkunum sjálfum eru yfirflokkar þeirra, stuðlar námundaðir að þriðja aukastaf, fjöldi orða sem voru flokkuð í hvern flokk og tvö orð sem voru notuð sem dæmisorð hvers flokks. Neðsta línan inniheldur meðalgildi stuðlanna og samtals fjölda flokkaðra orða.

<i>Staða</i>	<i>Yfirflokkur</i>	<i>Flokkur</i>	<i>Stuðull</i>	<i>Fjöldi orða</i>	<i>Dæmisorð</i>
1.	fjöldi	fjöldi	4.954	83828	margur, tvöfaldur
2.	fjöldi	tala	3.396	34152	tveir, þrennur
3.	tengsl	vissa	2.270	9017	vís, augljós
4.	tími	aldur	1.966	92295	ungur, tvítugur
5.	tími	hraði	1.961	26118	hægur, snöggur
6.	tengsl	tenging	1.928	12682	tilheyrandi, bundinn
7.	staðsetning	vídd	1.666	99178	langur, stór
8.	tengsl	samræmi	1.488	15599	svipaður, öðruvísi
9.	tengsl	erfiði	1.412	30190	auðveldur, flókinn
10.	tími	venja	1.400	124068	eðlilegur, skráttinn
11.	efni	ástand	1.366	9863	opinn, brotinn
12.	líkami	líkamsástand	1.355	6309	ófrískur, kynþroska
13.	efni	mengun	1.327	12451	skítugur, hreinn
14.	hegðun	samúð	1.294	21520	vinsæll, hataður
15.	skyn	hljóð	1.287	3953	hávær, hljóður
16.	staðsetning	staðsetning	1.282	12032	nálægur, fjarinn
17.	skap	tilfinning	1.225	15482	sorgmæddur, reiður
18.	tengsl	tilvísun	1.192	9429	táknrænn, beintengdur
19.	líkami	líkamstilfinning	1.159	6297	hungraður, syfjaður
20.	samfélag	félagsstaða	1.145	17367	giftur, starfandi
21.	tengsl	nákvæmni	1.127	12994	skýr, óljós
22.	skyn	birta	1.125	17559	bjartur, dimmur
23.	hegðun	hegðun	1.104	204798	kurteis, latur
24.	líkami	líkamsgerð	1.087	30896	vöðvastæltur, feitur
25.	skyn	áferð	1.051	23140	sléttur, mjúkur
26.	staðsetning	tilvera	1.051	13693	núverandi, yfirstandandi
27.	líkami	heilsa	1.049	20558	heilbrigður, lasinn
28.	veður	veður	1.046	1851	skýjaður, heiðskír
29.	tengsl	nauðsyn	1.038	35124	mikilvægur, óþarfa
30.	efni	þroski	1.027	3912	ferskur, rotinn
31.	staðsetning	rúmdreifing	1.013	21356	tómur, þéttur
32.	tengsl	virkni	1.008	6871	ósnortinn, skemmdur
33.	efni	þyngd	0.996	8899	massífur, þungur
34.	efni	raki	0.991	4035	þurr, blautur
35.	líkami	útlit	0.986	66388	rauðhærður, fallegur
36.	líkami	líf	0.960	8296	látinn, lifandi
37.	efni	jafnvægi	0.922	9656	stöðugur, skjálíandi
38.	hugur	gáfur	0.919	42203	heimskur, klókur
39.	efni	hitastig	0.899	11210	kaldur, heitur
40.	skyn	bragð	0.884	5658	súr, sætur
41.	tengsl	heild	0.879	19443	algjör, hálfur
42.	staðsetning	uppruni	0.872	49629	frumlegur, náttúrulegur
43.	fjöldi	uppskera	0.866	15192	arðbær, frjósamur
44.	hegðun	tilhneiging	0.852	21038	málgláður, rólegur
45.	tengsl	gildi	0.833	92494	mögulegur, gildur
46.	hreyfing	hreyfing	0.821	7232	klaufalegur, reiprennandi

47.	skyn	lykt	0.807	5510	arómatískur, angandi
48.	skap	örvun	0.796	75540	leiðinlegur, pirrandi
49.	tími	tími	0.794	9992	seinn, tímafrekur
50.	hegðun	samband	0.787	18761	vínalegur, samkynhneigður
51.	hugur	reynsla	0.773	21812	reynslulaus, lærður
52.	efni	eðliseiginleiki	0.758	24792	sveigjanlegur, segulmagnaður
53.	tengsl	öryggi	0.752	20489	hættulegur, öruggur
54.	hegðun	hæfni	0.739	31958	flinkur, hæfileikaríkur
55.	líkami	kyn	0.733	5249	karlkyns, kvenkyns
56.	samfélag	svæði	0.727	4647	þéttbýll, afskekktur
57.	fjöldi	verð	0.719	49432	dýrmætur, verðmikill
58.	staðsetning	stefna	0.709	11708	norðlægur, vestanverður
59.	efni	eðli	0.682	18181	gylltur, rafrænn
60.	tengsl	gagn	0.680	5849	skaðlegur, eitradur
61.	tengsl	hagnýtni	0.676	13921	áhrifaríkur, árangursríkur
62.	skyn	litur	0.630	42190	rauður, blár
63.	samfélag	trú	0.624	22745	kristinn, trúlaus
64.	samfélag	stjórnmal	0.623	31225	íhaldssamur, demókratískur
65.	hegðun	dýrseðli	0.550	22237	taminn, villtur
66.	staðsetning	lögur	0.539	9635	kringlóttur, hyrndur
67.	efni	efniseiginleiki	0.503	5621	járnríkur, súrefnismettaður
68.	efni	hamur	0.500	6051	slímugur, frosinn
69.	tengsl	regla	0.418	22800	skipulagður, kaotískur
70.	tengsl	ástæða	0.306	25600	sálfræðilegur, veðurtengdur
71.	samfélag	ríki	0.102	124473	íslenskur, alþjóðlegur
			/=1.076	+ =2026374	