



Óviðeigandi máltækni

Anton Karl Ingason

Hugvísindaping

2020

Inngangur

Tengslin milli máltækni og óviðeigandi málfars.

- 1 Máltæknilausnir sem læra óviðeigandi fyrirbæri.

Vondar vélþýðingar.

Námskeiðið *Hverfullt tungutak í deiglu samtímans*.

- 2 Máltækni sem hjálpar notendum að forðast óviðeigandi málfar.

Óviðeigandi orð.

Verkefni: Málrýni (Máltækniáætlun); Óviðeigandi orð (Nýsköpunarsjóður námsmanna).



Vondar vélþýðingar

Vondar vélþýðingar

Hver hafa unnið að þessu?

- | Undirritaður og Lilja Björk Stefánsdóttir fjölluðu um þetta efni í námskeiðinu *Hverfult tungutak í deiglu samtímans* vorið 2020.
- | Agnes Sólmundsdóttir, nemandi í námskeiðinu og starfsmaður á rannsóknarstofunni, safnaði frekari gögnum og vann við greiningu þeirra.

Markmið:

- | Að kortleggja að ákveðna þætti í kynjahalla máltækninnar í ensk-íslenskum þýðingum og vekja athygli á þeim.

Mál, kyn og máltækni

Menning getur litað tungumál.

- | Málsamfélög með rótgróna og kerfisbundna stéttaskiptingu hafa orð og/eða beygingar til að tjá stöðu fólks.
- | Forngríska var ekki með orð yfir snjallsíma.

Útilokun og staðalmyndir.

- | Sumt fólk upplifir útilokun þegar karlkyn er notað sem sjálfkerfið kyn fyrir fólk í íslensku.
- | Við getum tengt ákveðin hlutverk við lærðar hugmyndir okkar um kyn og það hefur áhrif á hvernig við notum málið.

Áskorun: Máltækni hugbúnaður lærir af texta sem hann sér þegar hann er þjálfður og sú tækni er nógu fullkomin til að læra kynjahalla í menningunni.

Kyn, málkerfi og útilokun

Stundum er karlkyn sjálfgefið í málkerfinu sem börn tileinka sér:

(1) Allir velkomnir!

En sum vilja nota hvorugkyn til að útiloka engin:

(2) Öll velkomin!

Hagsmunaárekstur:

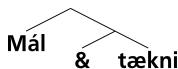
Hópur 1: Upplifir **útilokun**; (1) vísar ekki í mig.

Hópur 2: Uppl. **kjánahroll**; (2) brýtur gegn máltilfinningu.

Kallar á afstöðlun máls og umburðalyndi gagnvart breytileika.

Hópur 1: Skilja að fólk breytir ekki málkerfinu auðveldlega.

Hópur 2: Skilja að útilokun er alvarlegt mál.



Tvívæðar setningar í þýðingum

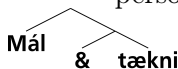
Þýðingarvélur eins og Google Translate læra af texta sem menningin framleiðir og endurspeglar hugmyndir um kyn í menningunni.

Prates, Avelar og Lamb (2019) könnuðu vélþýðingar á tvívæðum fornöfnum:

Ungverska > enska:

- (3) a. **ő** egy ápolónő > **she** is a nurse
b. **ő** egy mérnök > **he** is an engineer

Aðeins 4% tvívæðra fornafna um starfsfólk tengt tækni, vélum og stærðfræði var þýtt með kvenkyns persónufornafni.



Íslensk lýsingarorð á Google Translate

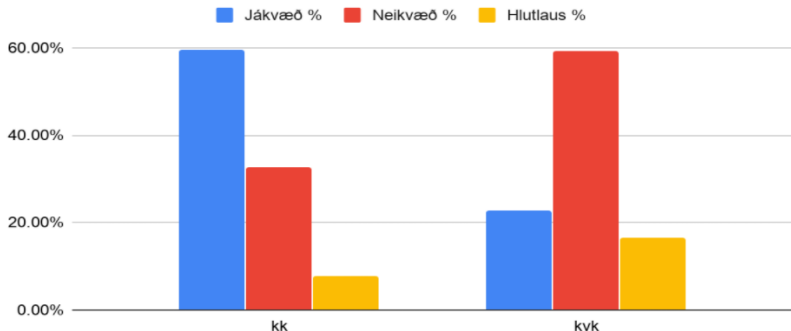
The screenshot shows the Google Translate interface. At the top left is the Google Translate logo. Below it are two buttons: 'Text' (selected) and 'Documents'. The language selection bar shows 'ENGLISH' selected on the left and 'ICELANDIC' selected on the right. The input text on the left is: 'I am vulnerable.', 'I am insecure.', 'I am weak.', 'I am strong.', 'I am confident.', and 'I am amazing.'. The output text on the right is: 'Ég er viðkvæm.', 'Ég er óörugg.', 'Ég er veik.', 'Ég er sterkur.', 'Ég er sjálfsöruggur.', and 'Ég er ótrúlegur.'. At the bottom left, there are microphone and speaker icons. At the bottom center, it shows '91/5000' and a keyboard icon. At the bottom right, there is another speaker icon.

Mál
& tækni

Persónulýsandi orð

262 lo. sem lýsa persónueinkennum og GT þýðir: *Ég er X*
GT þýðir 156 í kk en 66 í kvk.

Lo. kóðuð eftir því hvort jákvæð, neikvæð, hlutlaus:

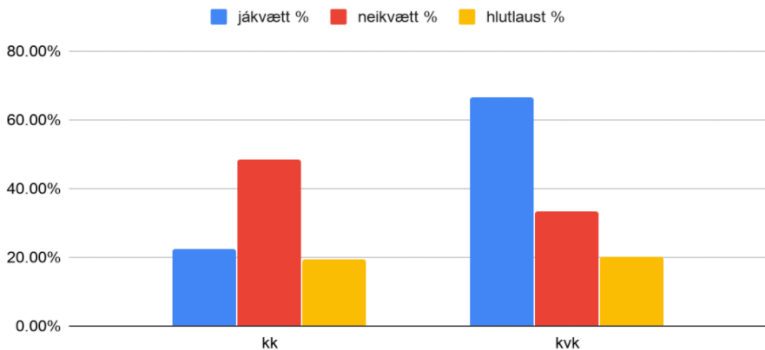


Kyn

Útlitslýsandi orð

67 lo. sem lýsa persónueinkennum og GT þýðir: *Ég er X*
GT þýðir 156 í kk en 66 í kvk.

Lo. kóðuð eftir því hvort jákvæð, neikvæð, hlutlaus:



Óviðeigandi orð

Óviðeigandi orð

Hver hafa unnið að þessu?

- | Agnes Sólmundsdóttir, starfsmaður á rannsóknarstofunni, safnaði gögnum og vann við greiningu þeirra.
- | Verkefninu stýra undirritaður og Lilja Björk Stefánsdóttir.
- | Tengist Málrýni (Máltækniáætlun) en var svo sérstakt verkefni í sumar á vegum Nýsköpunarsjóðs námsmanna.
- | Auk sjálfstæðrar söfnunar var unnið með gögn frá Miðeind og Samrómi.

Markmið:

- | Að kortleggja óviðeigandi orð í íslensku.



Óviðeigandi orð

Gagnagrunnur:

Safn um 2000 misóviðeigandi orða sem eru skráð og flokkuð. Orð sem eru almennt óviðeigandi, óviðeigandi í afmörkuðu samhengi eða gildishlaðin á einhvern hátt þannig að tilefni gæti verið til málfarsráðgjafar.

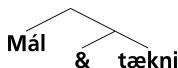
Mjög algengt að tilfinning fyrir orðum sé ólík milli fólks.

Gefið út með CC-BY 4.0 leyfi á Github.

Notagildi:

Til að þróa sjálfvirka málfarsráðgjöf (leiðréttingarforrit).

Undirstaða fyrir frekari skráningu og flokkun á óviðeigandi og gildishlöðnum orðum í íslensku.



Flokkar óviðeigandi orða

Almennt óviðeigandi

Blótsyrði

Drykkju/dóp- orð

Fötlunarorð

Heilbrigðistengt

Heimsku-orð

Kynjuð orð

Kynjuð orð bara um karla

Kynjuð orð bara um konur

Lýsingarorð (ljót um fólk)

Fagheiti

Nafnorð (ljót um fólk)

Orðasambönd

LGBTQIA+ orð

Sagnir

Trúarorð

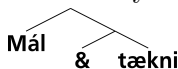
Útlit

Kynfæraorð

Áhersluforskeyti

Kynlífstengd orð

Þjóðernisorð



Dæmi um flokk: Fagheiti

barnaheimili

barnapía

bensíntittur

dagmamma

fóstra

gæslusystir

gengilbeina

glasabarn

hárgreiðslukona

hjúkka

hjúkrunarkona

kassadama

kennslukona

kuklari

skottulæknir

skúringakerling

skúringakona

vinnukona

þjónustustúlka

þrifakona

Dæmi um færslu: fóstura

Uppflettiorð: fóstura

Orðflokkur: no

Kóði: m (fyrir fagheiti)

Merking: leikskólakennari/leikskólaliði/starfsmaður á leikskóla (ATH einhverjir vilja vera kallaðir fóstrur, t.d. þau sem fóru í gamla fósturskólann)

Óviðeigandi ástæða: Telst úrelt heiti yfir fagstétt, þykir lítillækka starfið og útilokandi.

Áskoranir

Fólk hefur mjög ólíka tilfinningu fyrir því hvað telst óviðeigandi og hvað ekki.

Sumt af því sem er óviðeigandi fyrir einn er mjög virðingarvert fyrir annan.

Kóðun okkar á því sem er óviðeigandi er enn sem komið er tilraunakennd að sumu leyti og við munum halda áfram að læra á viðfangsefnið.

Github: <https://github.com/antonkarl/iceTaboo>

Mikilvægt er að vinnan er öll í opnum aðgangi. Hver sem er getur byggt á henni, breytt og snúið á hvolf.

Search or jump to... Pulls Issues Marketplace Explore

antonkarl / iceTaboo Unwatch 1 Star 0 Fork 0

Code Issues Pull requests Actions Projects Wiki Security Insights

master Go to file Add file Code

Xindan-Xu reorganized files 33 minutes ago 34

Ovideigandi_ord_flokkar...	convert to tsv files	2 hours ago
others	reorganized files	33 minutes ago
FlokkarAll.tsv	reorganized files	33 minutes ago
README.md	Update README.md	3 days ago
cc-by-4-0.txt	cc by license	6 days ago

README.md

Icelandic Taboo Database 1.0

An annotated collection of Icelandic words that are (sometimes) associated with a taboo or some sort of inappropriateness depending on context.

About: An annotated collection of Icelandic words that are (sometimes) associated with a taboo or some sort of inappropriateness in context.

Releases: No releases published. [Create a new release](#)

Packages: No packages published. [Publish your first package](#)

Samantekt

Máltækni getur lært óviðeigandi hluti úr menningunni.

Máltækni getur líka hjálpað fólki að forðast óviðeigandi orð í riti.

Hér var lýst nokkrum skrefum í rannsóknum og þróun sem þetta varðar.

| Margt er þó enn óunnið!

