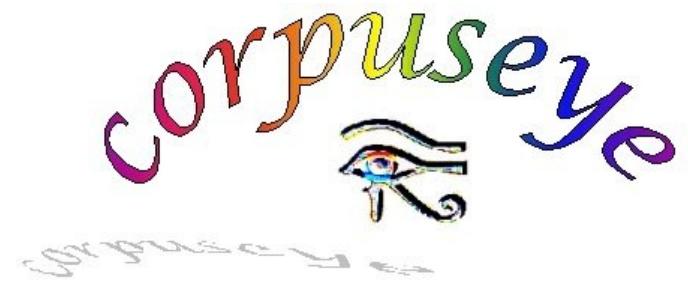


# CorpusEye:

## Supporting Your Research and Teaching with Annotated Data from Large Text Collections

Eckhard Bick



# Overview / Topics

## IKS centers & courses

### 1. Corpora & NLP

- Danish, English, German, American Studies

### 2. Corpus studies: e.g. Hatespeech, emojis

- Middle East and Muslim Studies

### 3. Word embeddings: Measuring stereotypes

- Culture & Technology

### 4. Distant reading: Danish literature

- e.g. H.C. Andersen, Pontoppidan, Comparative Literature

### 5. Gender studies

- Gender & Diversity, Uses of Literature

### 6. Corporate image profiling: Sustainability

- International Business Communication

### 7. Brand stereotypes: gender & age

- Business, Language and Culture

# Part 1: Corpora & NLP

## 1. Why Corpora?

- Corpora are curated collections of language data (written, spoken, multimodal) addressing a specific set of sources, languages or genres
- Corpora allow the extraction of structured and statistical knowledge and serve various purposes such as research, development or teaching
- Corpora may be enriched with linguistic or extra-linguistic (meta-) information

## 2. Why NLP?

- Getting knowledge out of language data (corpora) is not like reading a book
- 1 billion word corpus = 10,000 books = 300 m of bookshelves = a lifetime of reading
- you need quantitative methods (informed statistics)
  - optimally allowing targeted qualitative inspection
- you need to be able to find patterns and structures
- you need to separate information from “noise”

## 3. Linguistic processing (NLP)

- determines what you can do with a corpus and how easy it will be
- Compilation: corpus cleaning, meta data
- Annotation: Linguistic category tagging and structural analysis
- Task specific markers
  - e.g. sentiment markers, metaphor markers, task-specific ontologies

# Annotation methods

- **Manual** (for small specialized corpora and/or special features)
  - e.g. multimodal corpora with video, dialects, prosodics
  - semantic training data: e.g. sentiment analysis, hatespeech
- **Rule-based:**
  - exploiting context, linguistic knowledge (lexica, grammars)
  - generative: HPSG, LFG
  - analytic/reductionist: **Constraint Grammar (CG)**
    - transparency, traceable errors, genre-adaptable, interpretable
- **Machine-learning**
  - e.g. HMM-taggers, MALT-parser, LLM transformers and other neural network architectures capturing semantics through word embedding
  - growing context awareness (sentence/text embeddings)
  - predictive power
  - need for training data, issues with small or under-resourced languages

# Constraint Grammar (CG)

- (a) Wordform - can be a multi-word expression (MWE)
- (b) Lemma - e.g. 'sei, sabe, sabia, soube, souber' etc. -> 'saber' [know, get to know]
- (c) POS - e.g. **Noun**, **Verb**, **ADJektive**, **ADVerb**, **DETeminer**)
- (d) Inflection - e.g. **PResent**, **NOMinative**, **Singular**
- (e) Syntactic function - e.g. **@SUBJekt**, **@ADVerbiaL**
- (f) Secondary categories - e.g. <**mv**> main verb, <**dem**> demonstrative
- (g) Semantic categories - e.g. <**party**>, <**H...**> person, <**Q->- (h) Dependency tag - e.g #2->5 (token 2 attaches to head token 5)
- (i) Frames and semantic roles - e.g. <**FN:attack**>, **§AG**, **§DES**
- (j) Long-distance relations: Anaphora, discourse structure, speaker attribution, e.g.  
*SOURCE, SPEAKER* (IE verification, false news ...)
- (k) Informational structure: Cast extraction, e.g. *spouse\_of, child\_of, syn\_name, soc.role***



# CG: who & what?

Lingsoft Proofreader Finnish for Sisulizer 2008

99,00 € + VAT 23 %

[Add to cart >](#)

[More Sisulizer versions...](#)

Lingsoft Svefix 2.2 for Microsoft Office

50,00 € + VAT 23 %

[Evaluate](#) [Add to cart >](#)

Lingsoft Speech Controller 1.4 (Windows)

220,49 € + VAT 23 %

[Evaluate](#) [Add to cart >](#)



Machine translation  
Spell- and Grammar checking  
Corpus annotation  
Relational dictionaries: DeepDict

The screenshot shows a search results table for the word "voice". The columns include Part of Speech, Lemma, and Definition. Definitions for "voice" include "sound produced by the larynx", "sound produced by the larynx", "sound produced by the larynx", and "sound produced by the larynx". Other rows show definitions for "head", "deep", "grave", "husky", "articulate", "passive", "active", "quiet", "human", "replay", "gruff", and "narrative".

# NER

VISL grammar games:

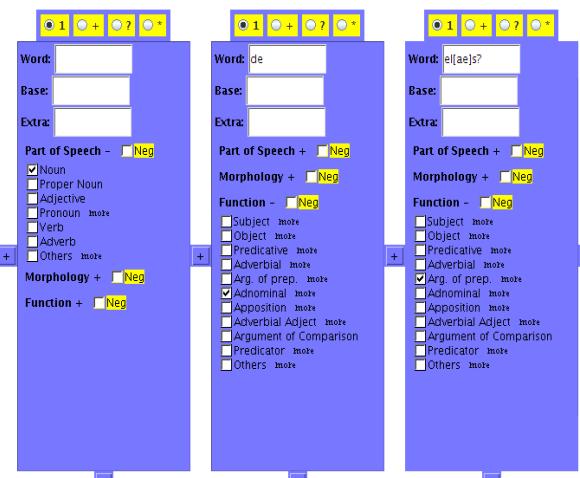


# GT

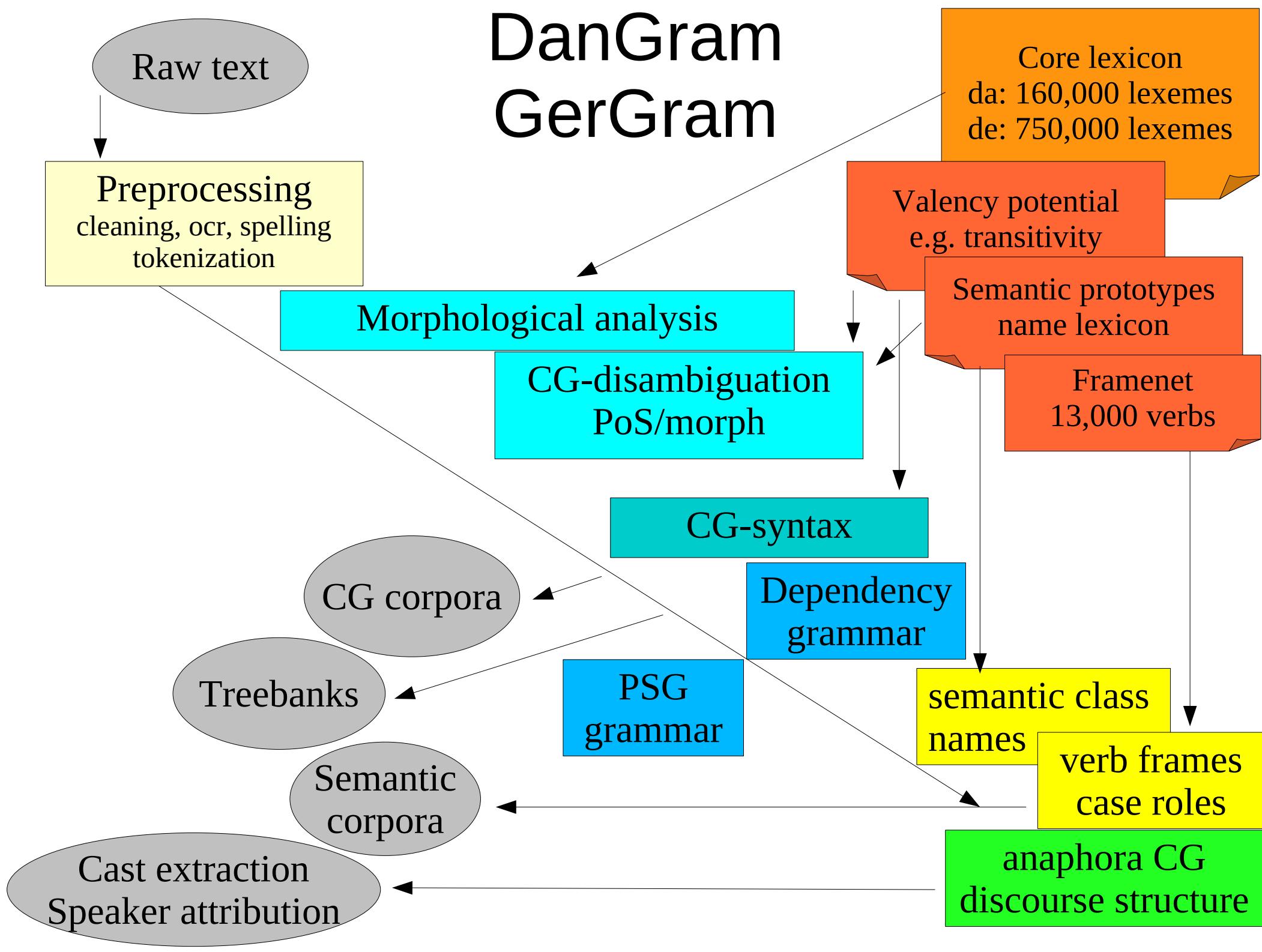
Machinese parsers  
News feed and relevance filtering  
Opinion mining in blogs  
Science publication monitoring



Annotated corpora:  
CorpusEye



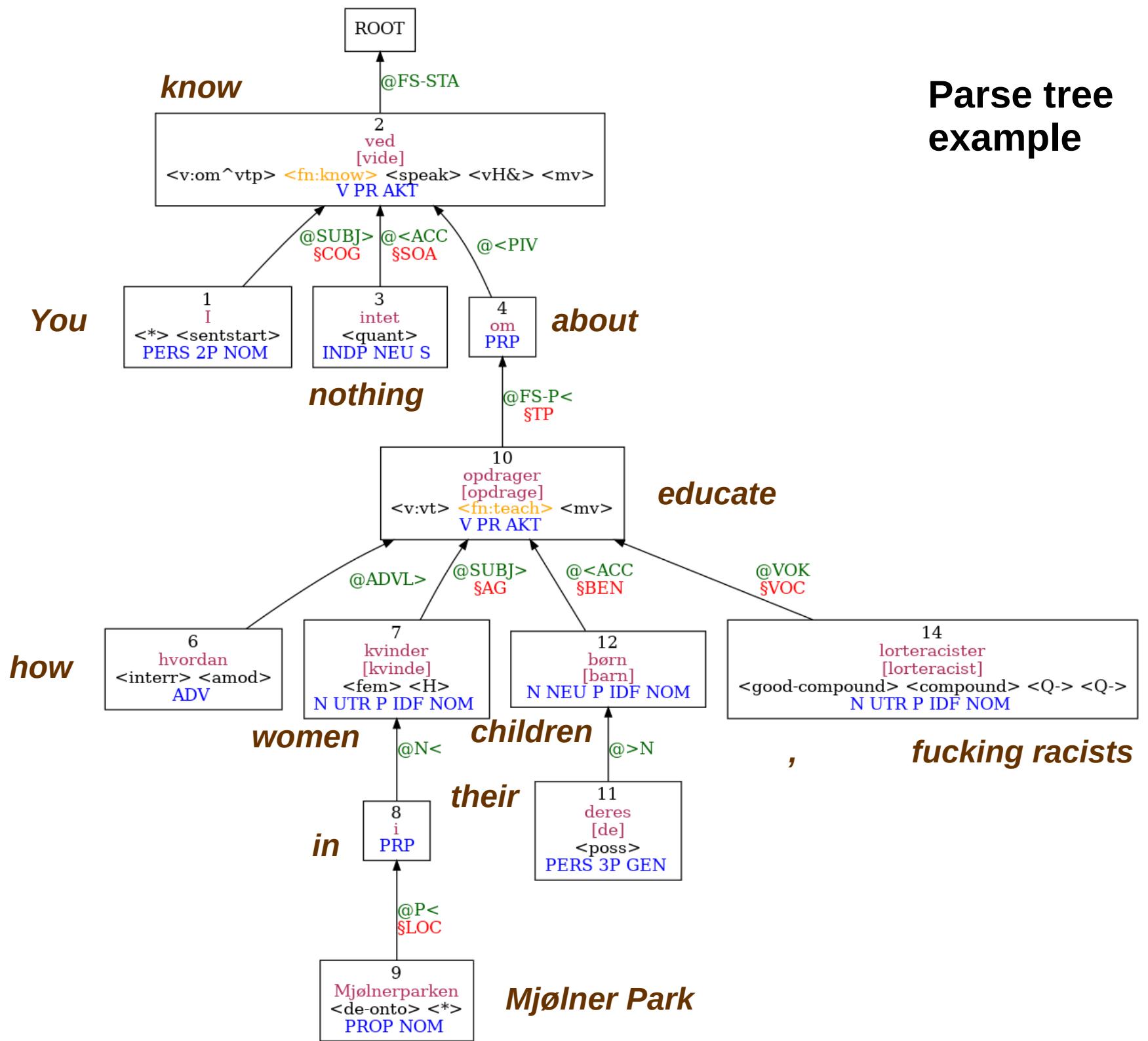
# DanGram GerGram



# Danish sentence parse

|               |                         |  |
|---------------|-------------------------|--|
| I             | <i>you</i>              | [I] PERS 2P NOM @SUBJ> §COG #1->2                                    |
| ved           | <i>know</i>             | [vide] <fn:know> <mv> V PR AKT @FS-STA #2->0                         |
| intet         | <i>nothing</i>          | [intet] <quant> INDP NEU S @<ACC SSOA #3->2                          |
| om            | <i>about</i>            | [om] PRP @<PIV #4->2   |
| \$,           | ,                       | [,] PU @PU #5->0   |
| hvordan       | <i>how</i>              | [hvordan] <interr> <amod> ADV @ADVL> #6->10                          |
| kvinder       | <i>women</i>            | [kvinde] <fem> <H> N UTR P IDF NOM @SUBJ> §AG #7->10                 |
| i             | <i>in</i>               | [i] PRP @N< #8->7  |
| Mjølnerparken | <i>Mjølner Park</i>     | [Mjølnerparken] <top> <Lh> PROP NOM @P< §LOC #9->8                   |
| opdrager      | <i>bring up</i>         | [opdrage] <fn:teach> <mv> V PR AKT @FS-P< §TP #10->4                 |
| deres         | <i>their</i>            | [de] <poss> PERS 3P GEN @>N #11->12                                  |
| børn          | <i>children</i>         | [barn] <Hbio> N NEU P IDF NOM @<ACC §PAT #12->10                     |
| \$,           | ,                       | [,] PU @PU #13->0  |
| lorteracister | <i> fucking racists</i> | [lorteracist] <N:lort~e+racist> <Hideo> N UTR P IDF NOM @VOK #14->10 |

# Parse tree example



# Annotation: semantics

## ■ semantic POS subclasses

- ◆ adverbs: <atemp>, <aloc>, <adir>, <aquant> ....
- ◆ pronouns: <rel>, <interr>, <dem>, <refl>, <quant> ...

## ■ semantic prototypes

- ◆ nouns: ~200 types: <Hprof>, <Vair>, <tool-shoot> ...
  - atomic feature bundles:  $\pm$ hum,  $\pm$ anim,  $\pm$ move,  $\pm$ loc ...
- ◆ adjectives: <jnat> <jpsych> <jcol> <jshape> <jgeo> ...

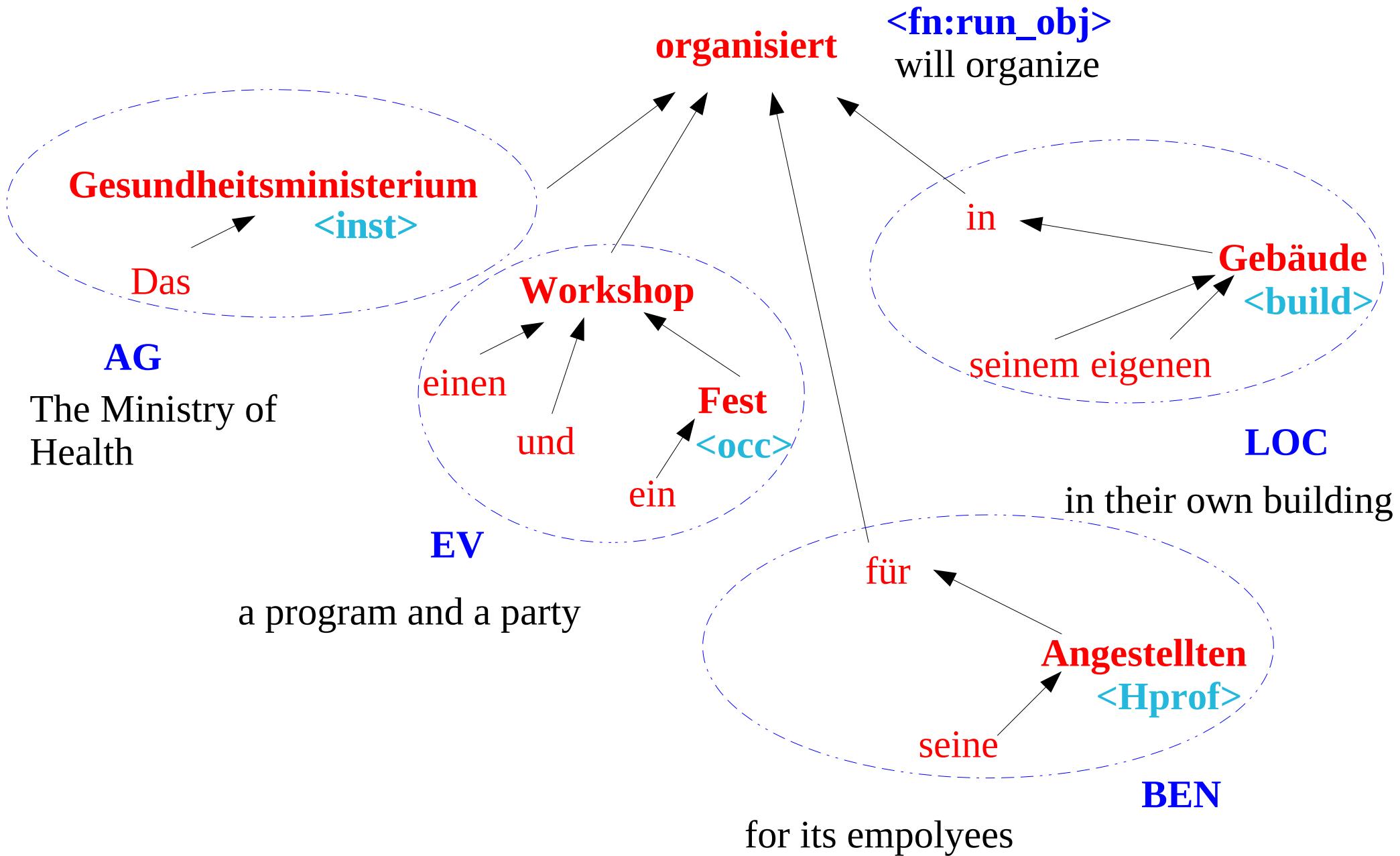
## ■ semantic roles

- ◆ 15 core roles: §AG, §PAT, §TH, §REC, §COG ...
- ◆ 35 “adverbial” and meta-roles: §DIR, §DES ...

## ■ Framenet

- ◆ ~ 500 different verb frames, covering 9-14,000 verb lemmas

# Dependency trees with semantics



# contextual rule examples

- select POS=noun, if there is an article to the left
- uniqueness principle
  - ◆ select a finite verb reading if there is no other finite verb candidate without a comma or coordinator in between
  - ◆ same for subjects, objects and other bound complements: only 1 per clause
- category variable unification
  - ◆ choose gender and number in an np depending on the gender and number of its head
  - ◆ disambiguate for same semantic class in co-ordinated nouns, lists etc.
- exploit verb frames
  - ◆ choose subject and agent role for an np left of a verb belonging to the “attack” or “build” frames.

# Semantic disambiguation: Frames

- <fn:leave> *stikke af* (to run off)
- <fn:give> *stikke ngn ngt [en øl eller en løgn]* (give sth to sb)
- <fn:reveal&identify> *stikke ngn [angive]* (inform about sb)
- <fn:poke> *stikke ngn [fx bi eller kniv]* (sting sb, stab sb)
- <fn:vanquish> *stikke et kort [trumfe]* (trump sth)

## Frame templates & dependency slot fillers

<FN:give/S§DON'H/D§REC'H/O§TH'cc>

→ triggered by ditransitive valency and a dative object / receiver role (§REC)

<FN:poke/S§AG'H | A/O§PAT'H>

→ triggered by e.g. an animal subject or a human (monotransitive) object

**Preparations:** Use semantic rather than syntactic dependencies

Mark transparent nouns:

e.g. *en slags x, en masse x, et glas x, en af x'erne, størsteparten af x'erne*



# Parser performance (GerGram)

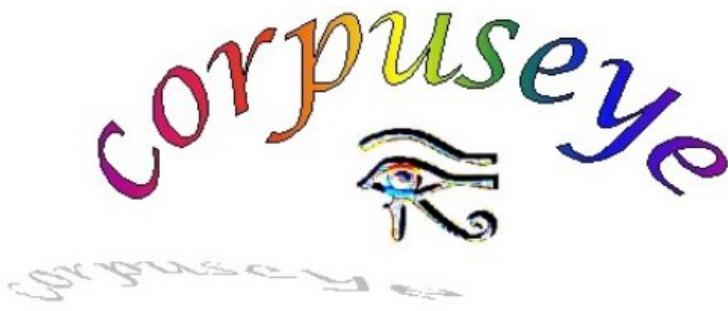
|                             | Recall<br>% | Precision<br>% | F1<br>% |
|-----------------------------|-------------|----------------|---------|
| POS (coarse)                | 98.54       | 98.54          | 98.54   |
| POS<br>+ inflection + lemma | 96.94       | 96.94          | 96.94   |
| Syntactic function          | 93.55       | 93.60          | 93.58   |
| Dependency links            | 94.18       | 94.18          | 94.18   |
| LAS (synt. + dep.)          | 92.02       | 92.08          | 92.05   |
| All errors                  | 91.32       | 91.38          | 91.35   |

# Part 2: Corpus studies & Hatespeech

Examples taken from:

- (a) **Danish literature** corpus (Gutenberg, Royal Library)
  - ca. 510 novels and short stories, mostly 1800-1940
- (b) **Information**, newspaper corpus
  - ca. 92 million words, 14,780 articles (1996-2008)
- (c) **XPEROHS** Social media corpus
  - bilingual German – Danish
  - Twitter (unabridged) & Facebook (seed pages)
  - 3 billion words
- (d) **Portuguese blog corpus** (Santos, Woloszyn & Vieira)
  - 2 billion words (2013-2018)

# Search interface



Three identical search interface panels are shown side-by-side, each with a header containing four radio buttons: (O) 1, (O) +, (O) ?, (O) \*

**Panel 1:**

- Word:
- ~case,  ~diacritics
- Base:
- ~case,  ~diacritics
- Extra:
- ~case,  ~diacritics
- Sem:
- ~case,  ~diacritics

**Panel 2:**

- Word:
- ~case,  ~diacritics
- Base:
- ~case,  ~diacritics
- Extra:
- ~case,  ~diacritics
- Sem:
- ~case,  ~diacritics

**Panel 3:**

- Word:
- ~case,  ~diacritics
- Base:
- ~case,  ~diacritics
- Extra:
- ~case,  ~diacritics
- Sem:
- ~case,  ~diacritics

Query:

SQ:  Don't detect CQL vs. plain text  Don't wrap (...) within <s/>  
 Case-insensitive  Collapse diacritics

**Manatee interface (fast, most up-to-date)**

Flags of various countries: Denmark, UK, France, Germany, Italy, Norway, Portugal, Spain, Romania, Sweden.

Experimental: [Icelandic](#) - [Faroese](#) - [Cross-language](#)

Corpora list:

- ddk-barelgazel (i 0.05 M)
- Korpus90 (i 21.3 M)
- Korpus2000 (i 21.8 M)
- Korpus2010 (i 41.8 M)
- Corrected K90/2000 (i 0.36 M)
- Stereotype Interviews (i 0.02 M)
- ddk-folketing (i 7.75 M)
- Epicle (i 1.12 M)
- Europarl (i 20.9 M)
- Facebook (i 34.1 M)
- FBmin v.1 (i 2.35 M)
- FBmin-neg (i 0.51 M)
- FBmin-neg DR/TV2 (i 0.13 M)
- Firma\_1 (i 0.06 M)
- Firma\_2 (i 0.01 M)
- Firma\_3 (i 0.01 M)
- Information 1996-2008 (i 85.5 M)

L [Show sub-corpora](#)

Total words: 871.3 M

Legend: i Requires password, i Histogram available, i Group By available, i Semantic classes, i Corpus information link, ? Help link

- Literature
- News
- Internet
- Wikipedia
- Social Media

# Structural / dependency search

## Sequential / filtered search

Execute Search

1    +    ?    \*

Word   ↗  
 ↙-case,  ↙-diacritics

Base   ↗  
 ↙-case,  ↙-diacritics

Extra   ↗  
 ↙-case,  ↙-diacritics

Part of Speech +  ↗  
 ↙-case,  ↙-diacritics

Morphology +  ↗  
 ↙-case,  ↙-diacritics

Function +  ↗  
 ↙-case,  ↙-diacritics

Semantic Role +  ↗  
 ↙-case,  ↙-diacritics

[Dep Head](#)   [Delete](#)

Adnominal   more  
 Apposition   more  
 Adverbial Adject   more  
 Argument of Comparison  
 Predicator   more  
 Others   more

Semantic Role +  ↗

Word   ↗  
 ↙-case,  ↙-diacritics

Base   ↗  
 ↙-case,  ↙-diacritics

Extra   ↗  
 ↙-case,  ↙-diacritics

Part of Speech -  ↗  
 ↙-case,  ↙-diacritics

Noun  
 Proper Noun  
 Adjective  
 Pronoun   more  
 Article  
 Verb  
 Adverb  
 Others   more

Morphology -  ↗  
 Finity   more  
 finite verb  
 infinitive  
 present participle  
 passive past participle  
 active past participle

Tense  ↗  
 more

SQ  1    +    ?    \*

Word   ↗  
 ↙-case,  ↙-diacritics

Base   ↗  
 ↙-case,  ↙-diacritics

Extra   ↗  
 ↙-case,  ↙-diacritics

Part of Speech +  ↗  
 ↙-case,  ↙-diacritics

Morphology +  ↗  
 ↙-case,  ↙-diacritics

Function -  ↗  
 ↙-case,  ↙-diacritics

Subject   more  
Object   more  
Predicative   more

[Dep Head](#)   [Delete](#)

Verb   checked  
 Adverb  
 Others   more

Passive Past Participle   checked  
 Active Past Participle

# Sequential search (SQ):

## (1) "immigrant"-morpheme list (reg.ex.)

--> (2) animal (A[a-z]+) & subject predicative (@SC)

|        |                      |                    |  |
|--------|----------------------|--------------------|--|
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » twittername » Weiße PickUps sind nicht nur <b>Umweltsäue</b> - sie neigen bekanntermaßen auch zum m Islamismus !   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | : Die Flüchtlinge sind das neue goldene <b>Kalb</b> : Der Tanz ... » <a href="https:// goo.gl/fb/N4jfDX">https:// goo.gl/fb/N4jfDX</a>   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | : Die Flüchtlinge sind das neue goldene <b>Kalb</b> » <a href="https://t.co=WCh0dz0Ua5">https://t.co=WCh0dz0Ua5</a>  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Albert Rösti : « Der radikale Islam ist eine <b>Bestie</b> ! » » Wer könnte der #SVP da wohl widersprechen ?   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » twittername » <b>Flüchtlinge sind die Kakerlaken</b> »   |
|        |                      |                    | direct: refugees are cockroaches   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | plärren dürfen der Flüchtling ist ne notgeile <b>Sau</b> . » Aber dann rumheulen , wenn wer auch Meinungsfreiheit  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Die Bigotterie der Linken : <b>Alle Männer sind Schweine</b> , es sein denn es sind Flüchtlinge  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Das sind keine Flüchtlinge , sondern islamische <b>Bastarde</b> die alle deutschen Frauen als Nutten   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | Zuwanderern sind für den syrischen « Flüchtling » <b>Aras Bacho</b> « Pogrom » . » <a href="https://twitter.com=ArasBacho=status=927657436416266240">https://twitter.com=ArasBacho=status=927657436416266240</a> |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | finden viel Geld - er hier ist halt ein schwarzes <b>Schaf</b> . » .   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | , Unterwanderer , weil , die ficken alles , <b>Esel</b> , Schafe und was weis ich noch . » .   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | : Die Flüchtlinge sind das neue goldene <b>Kalb</b> » via twittername  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | Mann , ich mach sowas nicht ! » Das war so ein <b>Flüchtlingsschwein</b> , basta !   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | sind doch nahezu alle tot , Drachengame totaler <b>krebs</b> in Dauerwiederholung , Almangame komplett tot , Rumgeheule  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Guten Morgen , mein Chef ist auch so ein <b>Vogel</b> , hat 5 Flüchtlinge eingestellt . » Aber nicht weil  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » twittername » Islam ist <b>Krebs</b> »» 생각난김에 해보자 Thomas Shelby : « Das heutige Familientreffen  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | oder Aleviten sind kein übel . » Es sind die <b>Wölfe</b> , Moslembrüder und Wahabiziegenwemser .  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | im m Krieg alleine zurück lässt » Ist ein feiges <b>Schwein</b> und kein Flüchtling ! » Sowas machen auch nur die Muslime  |

# CQP search language:

SQ = [word="(Flüchtling|Einwanderer|Ausländer|M[uo]slem|Islam).\*"] expand to s;  
SQ; [extra="(?:.\* )?A[a-z]+(?: .\*)?" & func="(?:.\* )?<?SC>?.\*(?: .\*)?" ]

- Export INFO TW neue Art des Selbstmordes sie sind eine verlogene **Ratte** mehr nicht  
Export INFO TW « ! « Der cis-Mann ist das **Schwein** , aber der Islam ist Frieden ! « https://twitter.com/atheist :  
Export INFO TW « Jetzt versteh ich das vllt - ist das eine **Flüchtlingskuh** ? « Milch von geflohen  
Export INFO TW mal erwähnt haben , dass Studenten noch größere **Parasiten** sind , als Flüchtlinge #jug  
Export INFO TW « twittername « Der Islam ist die neue heilige **Kuh** . « Empört sind einige , gehört wollen Sie aber nicht  
Export INFO TW « Der Islamische Brückenkopf , ist der **Wolf** im m Schafspelz . « https://t.co=Z7sJJ3LDvX  
Export INFO TW « twittername « Islam ist der **krebs** der welt  
Export INFO TW ist nicht der Islam , die Täter sind « stumpfe **Bestien** » die keinen Platz in unserer Mitte haben ! « !  
Export INFO TW sind wird e d Flüchtlingen besser gehen , sowie d **Bienen** , Schmetterlingen , Vögeln etc ... Aber m keinem  
Export INFO TW auf alle Moslems einzudreschen als wären sie **Hunde** . « Ihr wollt ja auch nicht  
gezeigt , merkt nur keiner , sie ist ein **MAULWURF** des türkischen SALAFISMU  
Export INFO TW dem Polizisten in die Hand : War das ein **wildes Tier** oder ein Moslem ? « Oder gar beides ?  
Export INFO TW ... « wir sind überschwemmt von Nutzlosen und **Sozialparasiten** und Islam ... « ne bin nicht Rassist , da meine Frau  
Export INFO TW « Islam ist der **krebs** der welt  
Export INFO TW die Rechten freut das ja nur Männer sind eh **Schweine** die AfD besteht nur aus alten weißen Männern Ich würde  
Export INFO TW bezeichnen , sind keine #Andersdenkende , sondern **Schäfchen** , die blind der braunen Propagandaschinerie folgen  
Export INFO TW « twittername « Islam ist der **krebs** der welt
- indirect: ... are even bigger parasites than refugees
- ... a wild animal or a Muslim?  
Or really both?

# Mouse-over inspection

[lex="(indvandr|flygtning|udlænding|muslim|islam|arab).\*"] SQ [lex="emo-angry.\*"]

|  |  |  |   |               |      |                           |  |  |
|--|--|--|---|---------------|------|---------------------------|--|--|
|  |  |  | og splittelse og skabe en « dem-og-os-mentalitet » ...                  |               |      |                           |  | &&dkpol &&dkmedier                           |
|  |  |  | om muslimske forældre , der <u>tvangsgifter</u> deres brune <u>børn</u> |               |      | &&dkpol &&dkmedier &&eudk |  |  |
|  |  |  | og mest gemene over_for udlændingene på samfu                           | <u>børn</u>   | barn | Hbio R:sd-PAT:19          |  |  |
|  |  |  | - umuligt ... gi' mig navnet og jeg blokerer                            | N             |      |                           |  |  |
|  |  |  | skal alle os svin rejse hen , hvis Danmark blive                        | NEU P IDF NOM |      |                           | Kan vi bo hos twittername .                    | for han er jo e                              |
|  |  |  | ▫ twittername ▫ tror vores politiker har solgt os                       | <ACC          |      |                           | åber snart , der er flere dansker som vågner c |  |
|  |  |  | ? ▫ DF : ▫ Vi fejrer fx udlændingestramningen                           | PAT           |      |                           | Muslimer : ▫ cool , så prøver vi Eid-kage .    |  |
|  |  |  | er et katastrofeområde og snart en muslimsk stat                        | Hbio          |      |                           |  |  |
|  |  |  |   | 22 → 19       |      |                           |  |  |
|  |  |  | udlændinge , men du plejer ikke at være sådan et r * * hul              |               |      |                           |  |  |
|  |  |  | Han vil med sin religion bare have lov at gøre hvad han vil             |               |      |                           |  | Uuuud med ham . ▫ Min pointe PU Lad os stu   |
|  |  |  | vil med sin religion bare have lov at gøre hvad han vil                 |               |      |                           |  | Uuuud med ham . ▫ Min pointe PU Lad os straf |

[word="(Flüchtling|Einwanderer|Ausländer|M[uo]slem|Islam).\*" & func="<?SUBJ>\*" & h\_pos="V"]

In corpus: GER\_TWITTER\_20200306\_A, Found 72303 results (72303)..

Export INFO TW

» twittername » Ausländer sind keine objekte die intergriert werden sollen ,

Export INFO TW

mal die « Ausländer sind faul-Nazis » auf die « Ausländer nehmen uns die Arbeit weg-Nazis » lassen . » Und gucken .

Export INFO TW

» Ich wollte was abbeizen , aber ein Ausländer hat mir den Abbeizplatz weggenommen . » ( Anonymer

Export INFO TW

in Syrien . » Wenn Opposition gewinnt , tilgen Islamisten eine 2000jährige christliche Kultur

Export INFO TW

» Flüchtlinge müssen nicht für Polizeibegleitung zahlen . » Bundesverwaltungsgericht

Export INFO TW

. » a ) Jugendstrafrecht , lasche Strafe , abbr ) Islam prägte sie , gilt strafmildernd .

Export INFO TW

» « Aus Rumänien : Zurückkehrende Flüchtlinge werden { mit Po r Weltspiegel

Export INFO TW

, Deutschland Finanziert , Medien berichten Moslems sind Terroristen »

Muslims are terrorists

Export INFO TW

@steinbacherika » Aber sonst hieß es doch immer « Islam gehört zu Deutschland

Export INFO TW

, will die AfD genau das auch zementieren . » Einwanderung solle « der Nation » nutzen . » Whatever that means

Export INFO TW

» Wenn radikale Islamisten als « Scharia-Polizei » durch Wuppertal ziehen , wird

Export INFO TW

Claudia Mieth » » Was passiere , wenn die Flüchtlinge d.Vorschlag endgültig ablehnten , sei ungewiss » »

Export INFO TW

» Die #IS IslamistenJihadisten können genauso unfähig reimen , wie der #KKK . » #MississippiBurnin

Export INFO TW

von #Rotherham über Rotterdam bis Berlin ) : Islam versklavt junge Mädchen ! » #Pädos

Export INFO TW

Angedrohte Hinrichtung : So leidet der britische Flüchtlingshelfer im m ISIS-Knast - Politik Ausland - Bild.de » Laut

Export INFO TW

» Wie viele Flüchtlinge kommen aus Syrien und anderen Ländern ? » Wer nimmt

Export INFO TW

» Islamisten verfolgen

Export INFO TW

Personen , Zellen , Wohnorte : Wie die deutschen Islamisten vernetzt sind » http://spon.de=alpi via twittername

Export INFO TW

mit Deutschland ( 200 000 ) ja nur wenige Flüchtlinge zu versorgen . » 30 % mehr in Ö

Export INFO TW

» « Ein Ausländer hat mir den Arbeitsplatz weggenommen ! »

A foreigner has taken my job  
(away from me)

Export INFO TW

s Trinkwasser im m #Ruhrgebiet ? » Hafez : « Der Islam ist in Europa schon lange keine Religion m

Export INFO TW

» Kanzlerin #Merkel : « Der Islam gehört zu Deutschland - das ist so . » Dieser Meinung

Export INFO TW

vorher lesen , damit man versteht , wie dieses « Islamgesetz » entstanden ist .

Export INFO TW

» Islam nervt » http://wp.me=pLuGG-rb

Export INFO TW

! » » In #Leipzig beschimpften #Antideutsche Flüchtlinge als « Scheiß-Araber » » https : //www.jungewelt.de/2015/03-18/025.php

Export INFO TW

» Wäre die Germanwings-Maschine ein Flüchtlingsboot im m Mittelmeer gewesen , würde das übrigens niemanden

"immigrant" reg.ex. + subject, dep-head = verb, not passive, frame = attack

- Export INFO TW □ + Flüchtlinge sollen zwei Mädchen sexuell attackiert haben □ http://n-tv.io/KfVzwd
- Export INFO TW □ #Limbach : 16-jähriger afghanischer Flüchtling als Seriensexter überführt - attackierte mindestens
- Export INFO TW □ #Linz : 2 Ausländer überfallen junges Mädchen frühmorgens , zerren sie
- Export INFO TW □ #Langenfeld : Irakischer Flüchtling attackiert mit #Messer einen Iraner aus christenfeindlichen
- Export INFO TW □ « Flüchtling » Eric X. überfiel Paar im Zelt - Frau ( 23 ) mit
- Export INFO TW □ #Hamburg : Abgelehnter Flüchtling aus Senegal greift wg Beziehungsstress Polizei bei
- Export INFO TW □ #Öffentlichkeitsfahndung ! □ Dieser Ausländer überfällt Hannoveranerin mit einem Messer und sie aus
- Export INFO TW □ #Aarau : Türkischer Flüchtling überfällt #Billard-Lounge und b ...
- Export INFO TW □ #Hechthausen : Somalischer Flüchtling geht mit einem #Messer auf eine (... takes 19-yr-old-she on ...)
- Export INFO TW □ der versuchten Vergewaltigung - Mutmaßlicher Ausländer greift 19-Jährige von hinten an , zerrt si ... □ https://twitter.com/i/w ...
- Export INFO TW □ Unna : « Flüchtlings-Mob » greift Kneipen- und Kirmesbesucher an □ twittername
- Export INFO TW □ Wenn ein Flüchtling die Roth überfällt und vergewaltigt müssen die Kinder
- Export INFO TW □ Schweden Einwanderer mit arabischen Namen schießt nachts mit min. 20 Schüssen
- Export INFO TW □ zu tun □ Moslem greift Deutschen an □ ->
- Export INFO TW □ « Flüchtling » geht auf « Flüchtling » los
- Export INFO TW □ Flüchtlingsrudel greift Polizei an - Für Lügenpresse nur « Junge Leute
- Export INFO TW einzige Unterschied ist der , dass die « nur » Ausländer angreifen und Sie fein raus sind .
- Export INFO TW □ Flüchtlingsrudel greift Polizei an - Für Lügenpresse nur « Junge Leute
- Export INFO TW □ Tunis : Islamist attackiert 2 Polizisten □ https://wwwanton-hilft.de/tunis-islamist-attack
- Export INFO TW □ Sehe ich auch so . □ Der Islam wird Europa angreifen , dann geht 's hier rund - nur
- Export INFO TW Hauptstadt Tunis hat ein mutmaßlicher Islamist zwei Polizisten angegriffen und mit ... □ http://fb.me/2hcZ9WUbV
- Export INFO TW □ Deutschland eine Irrenanstalt . □ Ausländer schießen im Ernstfall natürlich viel leichter auf
- Export INFO TW □ Am 3. August 2014 überfiel der « Islamische Staat ( IS ) » das Hauptsiedlungsgebiet der êzidischen Minderheit
- refugee [animal-] pack
- Refugee attack refugee  
(Refugee goes refugee after)

Execute Search

1    +    ?    \*

Word    $\neg$   
  $\neg$ case,   $\neg$ diacritics  
Base    $\neg$   
  $\neg$ case,   $\neg$ diacritics  
Extra    $\neg$   
  $\neg$ case,   $\neg$ diacritics

Part of Speech -   $\neg$

- Noun
- Proper Noun
- Adjective
- Pronoun more
- Article
- Verb
- Adverb
- Others more



Morphology +   $\neg$

Function -   $\neg$

Subject more

Object more

Predicative more

Adverbial more

Arg. of prep. more

Adnominal more

Prenominal

Postnominal

Apposition more

1    +    ?    \*

Word    $\neg$   
  $\neg$ case,   $\neg$ diacritics  
Base    $\neg$   
  $\neg$ case,   $\neg$ diacritics  
Extra    $\neg$   
  $\neg$ case,   $\neg$ diacritics

Part of Speech -   $\neg$

- Noun
- Proper Noun
- Adjective
- Pronoun more
- Article
- Verb
- Adverb
- Others more



Morphology +   $\neg$

Function +   $\neg$

Subordinating Conjunction

Coordinating Conjunction

Preposition

Numeral

Article

Interjection

Punctuation

Prefix

Semantic Role +   $\neg$

1    +    ?    \*

Word  (Flüchtling|Einwa)   $\neg$   
  $\neg$ case,   $\neg$ diacritics  
Base    $\neg$   
  $\neg$ case,   $\neg$ diacritics  
Extra  [HA][a-z]+   $\neg$   
  $\neg$ case,   $\neg$ diacritics

Part of Speech -   $\neg$

- Noun
- Proper Noun
- Adjective
- Pronoun more
- Article
- Verb
- Adverb
- Others more



Morphology +   $\neg$

Function +   $\neg$

Semantic Role +   $\neg$

Dep Head

Delete

Searching for typical attributes

# searching variable, lexo-grammatically specified n-grams

## ADJ @>N (prenominal) ADJ\* (0 or more) N + "immigrant" reg.ex

### simple concordance

|        |                      |                    |  |
|--------|----------------------|--------------------|--|
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Wieder massenweise <b>unterzuckerte Muslime</b> unterwegs : « Fastenzeit beginnt - 7 strenge Regeln  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » <b>Liebe Muslime</b> . » Ihr dürft jetzt auf keinen Fall an Fleischbällchen  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | ihr die 111 Leute neulich entlassen , um diesen <b>größenwahnsinnigen Moslem</b> durchzufüttern ? » Respekt !  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | im m Irak . » Jetzt sind Christen , Jesiden , <b>moderate Muslime</b> in Gefahr . » Aufstehen !  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | den Terrorismus in #Deutschland und wirkt gegen <b>unschuldige Muslime</b> und Frauen » <a href="https://t.co=WFclVTLWt0">https://t.co=WFclVTLWt0</a>                          |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » <b>Werte Muslime</b> in Europa ! » via twittername   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » <b>Indische Muslime</b> rasten aus , weil eine Aktivistin ein Ende der Tierquälerei  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Rayhana Sultan wird von <b>wütenden Muslimen</b> vollgespampt , weil sie gestern in einer Doku über Ex-Muslime   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » twittername » SED-Parteisöhnchen zum m <b>überzeugten Islamisten</b> : Axel Mylius »   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | Der NDR-Rundfunkrat kritisiert die Einladung der <b>vollverschleierten Schweizer Muslimin</b> completely veiled Swiss Muslim-she   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | Interne Info aus Berliner Polizeiführung : Täter <b>tunesischer Moslem</b> . » das der General   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Ein <b>wahrer Muslim</b> der das Wort Gottes wirklich predigt . » <a href="https://t.co=bXMXxuGex6">https://t.co=bXMXxuGex6</a>  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Nicht jedes Magazin wie der SPIEGEL , der <b>solche Moslems</b> wie Hasnain Kazim einstellt , verliert jedes Jahr 15   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | und gleichzeitig einer von Rassismus <b>betroffenen Muslima</b> in der Bim beizustehen .   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | in Ruhe darüber nachdenken . » Gilt für jeden <b>gläubigen Moslem</b> als das wahre , und unabänderbare Wort Gottes . » <a href="https://twitter.co">https://twitter.co</a>    |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Wie kritisiere ich den Islam , ohne <b>gläubige Muslime</b> zu verletzen ? » So zum m Beispiel : <a href="http://m.youtube.com/watch?v=v">http://m.youtube.com/watch?v=v</a> |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » twittername » wieviele <b>musl Flüchtlingsfrauen</b> wurden nochmal von Deutschen vergewaltigt bzw. ermordet   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » <b>Ägyptische Muslime</b> spenden Blut für die Opfer der Anschläge auf koptische   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » #Wien : <b>16-jährige Muslima</b> sticht eine Frau nieder , weil sie sich von ihr beleidigt  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Selbst eine <b>liberale Muslima</b> aus der #SPD findet dieses Bild absurd ... » <a href="https://t.co=EXJXvHA6jp">https://t.co=EXJXvHA6jp</a>                               |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | Ich war dabei . » Wann gibt es e.Aufstand aller <b>anständigen Muslime</b> ? » #Lieberberg   |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » Österreich : Jeder <b>dritte Muslim</b> « hochfundamentalistisch » » Ich realisiere das immernoch  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | » <b>Kluge Muslime</b> laden die Menschen zum m Islam ein . » Dumme Muslime  |
| Export | <a href="#">INFO</a> | <a href="#">TW</a> | Muslime laden die Menschen zum m Islam ein . » <b>Dumme Muslime</b> möchten direkt die Scharia einführen . » Sieht das   |

# Quantitative analysis / statistics

*corpus eye* [Help](#) | [Grammatical i](#)

sort freq rel

Field Lemma ▾  
Left Context  
Right Context  
By Left Edge ▾  
Right Edge  
Dep Head

Offset 0 ▾

Freq items 100 ▾

Export



Searched for: []  
z]+(?: .\*)?" & p  
In corpus: GER  
Found 10651 re  
1 ≥ 3 4 5 6 7 »

Export All

Export [INFO](#) [TW](#)

Export [TINFO](#) [TW](#) so weltf

Field Word ▾  
Max rows 300 ▾

Count n-grams  
Export n-grams

[Refine search](#)

[New search](#)

Field 

Left Context

Right Context

By 

Right Edge

Dep Head

Offset Freq items Export Field Max rows [Refine search](#)[New search](#)**GER\_TWITTER\_20200306\_A (10651)**

frequencies:

freq<sup>2</sup>/norm (freq/corp\*10<sup>8</sup>) fi

|                               |                            |
|-------------------------------|----------------------------|
| <a href="#">radikal</a>       | 1507.55 (160.32) 9% [961]  |
| <a href="#">jung</a>          | 117.36 (77.24) 4.3% [463]  |
| <a href="#">gläubig</a>       | 8746.61 (52.55) 2.9% [315] |
| <a href="#">deutsch</a>       | 87.21 (48.38) 2.7% [290]   |
| <a href="#">liberal</a>       | 332.87 (39.71) 2.2% [238]  |
| <a href="#">ander</a>         | 4744.54 (38.71) 2.1% [232] |
| <a href="#">friedlich</a>     | 23.80 (26.86) 1.5% [161]   |
| <a href="#">lebend</a>        | 114.35 (26.2) 1.4% [157]   |
| <a href="#">gut</a>           | 0.45 (21.86) 1.2% [131]    |
| <a href="#">moderat</a>       | 133.36 (20.52) 1.1% [123]  |
| <a href="#">türkisch</a>      | 1086.08 (18.52) 1% [111]   |
| <a href="#">illegal</a>       | 7.15 (18.36) 1% [110]      |
| <a href="#">mutmaßlich</a>    | 1009.22 (17.85) 1% [107]   |
| <a href="#">gefährlich</a>    | 917.10 (17.02) 0.9% [102]  |
| <a href="#">lieb</a>          | 23.35 (16.52) 0.9% [99]    |
| <a href="#">kriminell</a>     | 39.49 (16.19) 0.9% [97]    |
| <a href="#">männlich</a>      | 812.38 (16.02) 0.9% [96]   |
| <a href="#">ganz</a>          | 0.16 (16.02) 0.9% [96]     |
| <a href="#">konservativ</a>   | 14 (15.78) 0.9% [96]       |
| <a href="#">böse</a>          | 25.3 (15.52) 0.9% [96]     |
| <a href="#">antisemitisch</a> | 4.63 (15.18) 0.9% [96]     |
| <a href="#">einzig</a>        | 11.1 (15.18) 0.9% [96]     |
| <a href="#">echt</a>          | 1.4 (15.18) 0.9% [96]      |
| <a href="#">falsch</a>        | 49.3 (15.18) 0.9% [96]     |
| <a href="#">gewaltbereit</a>  | 9.3 (15.18) 0.9% [96]      |
| <a href="#">arm</a>           | 4.19 (15.18) 0.9% [96]     |
| <a href="#">integriert</a>    | 215.3 (15.18) 0.9% [96]    |
| <a href="#">fanatisch</a>     | 124.14 (15.18) 0.9% [96]   |
| <a href="#">arabisch</a>      | 35.25 (15.18) 0.9% [96]    |
| <a href="#">fromm</a>         | 102.28 (9.83) 0.9% [96]    |
| <a href="#">militant</a>      | [96]                       |

**GER\_TWITTER\_20200306\_A (10651)**

frequencies:

freq<sup>2</sup>/norm (freq/corp\*10<sup>8</sup>) freq/conc num

|                                    |                            |
|------------------------------------|----------------------------|
| <a href="#">gläubig</a>            | 8746.61 (52.55) 2.9% [315] |
| <a href="#">ander</a>              | 4744.54 (38.71) 2.1% [232] |
| <a href="#">radikal</a>            | 1507.55 (160.32) 9% [961]  |
| <a href="#">türkisch</a>           | 1086.08 (18.52) 1% [111]   |
| <a href="#">mutmaßlich</a>         | 1009.22 (17.85) 1% [107]   |
| <a href="#">gefährlich</a>         | 917.10 (17.02) 0.9% [102]  |
| <a href="#">männlich</a>           | 812.38 (16.02) 0.9% [96]   |
| <a href="#">böse</a>               | 778.88 (15.69) 0.8% [94]   |
| <a href="#">gewaltbereit</a>       | 495.83 (12.52) 0.7% [75]   |
| <a href="#">liberal</a>            | 332.87 (39.71) 2.2% [238]  |
| <a href="#">zugewandert</a>        | 296.53 (9.68) 0.5% [58]    |
| <a href="#">antisemitisch</a>      | 254.13 (15.52) 0.8% [93]   |
| <a href="#">fanatisch</a>          | 215.96 (11.68) 0.6% [70]   |
| <a href="#">sunnitisch</a>         | 211.64 (8.18) 0.4% [49]    |
| <a href="#">gemäßigt</a>           | 186.52 (7.68) 0.4% [46]    |
| <a href="#">säkular</a>            | 18.50 (7.68) 0.4% [45]     |
| <a href="#">strenggläubig</a>      | [45]                       |
| <a href="#">praktizierend</a>      | 17.1 (7.68) 0.4% [45]      |
| <a href="#">syrisch</a>            | 162.1 (7.68) 0.4% [45]     |
| <a href="#">moderat</a>            | 13.1 (7.68) 0.4% [45]      |
| <a href="#">arabisch</a>           | 12.1 (7.68) 0.4% [45]      |
| <a href="#">jung</a>               | 11.1 (7.68) 0.4% [45]      |
| <a href="#">lebend</a>             | 11.1 (7.68) 0.4% [45]      |
| <a href="#">afghanisch</a>         | 114.1 (7.68) 0.4% [45]     |
| <a href="#">militant</a>           | 102.1 (7.68) 0.4% [45]     |
| <a href="#">orthodox</a>           | 95.99 (3.52) 0.4% [45]     |
| <a href="#">muslimisch</a>         | 95.99 (3.52) 0.4% [45]     |
| <a href="#">deutsch</a>            | 87.21 (48.38) 2.7% [290]   |
| <a href="#">gewalttätig</a>        | 84.71 (5.18) 0.2% [31]     |
| <a href="#">afrikanisch</a>        | 84.71 (5.18) 0.2% [31]     |
| <a href="#">antikapitalistisch</a> | 79.33 (5.01) 0.2% [30]     |

so-called  
dangerous  
dear  
criminal  
all  
conservative  
evil  
only  
anti-semitic  
real  
false  
violent

male  
evil  
violent  
liberal  
immigrant  
anti-semitic  
fanatic

| Concept noun                      | associated Danish adjectives (translated)   |
|-----------------------------------|---|
| <i>indvanderer</i><br>[immigrant] | <p>FB: non-Western, <b>criminal</b>, <b>illegal</b>, Western, muslim, <b>unintegrated</b></p> <p>TW: non-Western, <b>illegal</b>, arbejdsløs, Western, muslim, <b>criminal</b></p> <p><i>tyvagtig</i> [stealing], <i>hjernelam</i> [braindead], <i>pædofil</i> [pedophile], <i> fucking</i>, <i>satans</i> [damned]</p>   |
| <i>flygtning</i><br>[refugee]     | <p>FB: <b>so-called</b>, Syrian, economical, real, criminel, muslim, Palestinian</p> <p>TW: Syrian, Palestinian, vulnerable, greedy, new, so-called, under-age<br/><i> fucking</i></p>  |
| <i>udlænding</i><br>[foreigner]   | <p>FB: <b>criminal</b>, well-educated, "hardened"<br/>(aggressive), <b>unintegrated</b>, <b>so-called</b>, resident, non-Muslim</p> <p>TW: <b>criminal</b>, well-educated, homeless, <b>mediocre</b>, well-paid, non-Western<br/><i> satans</i> [damned], <i> fucking</i>, <i>fæl</i>, <i>væmmelig</i> [disgusting], <i>forpulet</i> [goddamn], <i>morderisk</i> [murderous]</p>        |
| <i>muslim</i><br>[Muslim]         | <p>FB: strict/orthodox, bad, <b>dear</b>, <b>extremist</b>, <b>unprovable</b>, non-Western</p> <p>TW: religious, <b>dear</b>, moderate, secular, strict/orthodox, renegade<br/><i> fucking</i>, <i>ulækker</i> [disgusting], <i>sindsyg</i> [insane], <i>forbandet</i>, <i>satans</i> [damned], <i>rådden</i> [rotten], <i>hjernedød</i> [brain-dead], <i>bindegal</i> [raving mad]</p> |

## Collocation search

ADJ + N

more & different (negative) stereotypes for *immigrant* and *foreigner* than for *refugee*.

distinct (negative) view on *Muslim*

## n-grams

--> inspection (quantitative-qualitative iterations)

| Count | Words                                       |
|-------|---|
| 200   | <a href="#">junge muslime</a>               |
| 182   | <a href="#">radikale islamisten</a>         |
| 177   | <a href="#">radikale muslime</a>            |
| 148   | <a href="#">radikalens islamisten</a>       |
| 125   | <a href="#">radikalens muslimen</a>         |
| 118   | <a href="#">liberale muslime</a>            |
| 68    | <a href="#">falscher islamist</a>           |
| 67    | <a href="#">deutsche muslime</a>            |
| 67    | <a href="#">lebenden muslime</a>            |
| 63    | <a href="#">gläubige muslime</a>            |
| 59    | <a href="#">andere muslime</a>              |
| 58    | <a href="#">gefährlichen islamisten</a>     |
| 56    | <a href="#">radikale moslems</a>            |
| 53    | <a href="#">männliche muslime</a>           |
| 52    | <a href="#">radikalens moslems</a>          |
| 48    | <a href="#">jungen muslimen</a>             |
| 46    | <a href="#">liebe muslime</a>               |
| 39    | <a href="#">gläubiger moslem</a>            |
| 36    | <a href="#">mutmaßliche islamisten</a>      |
| 35    | <a href="#">illegalen immigranten</a>       |
| 35    | <a href="#">moderate muslime</a>            |
| 34    | <a href="#">konservative muslime</a>        |
| 33    | <a href="#">deutsche islamistin</a>         |
| 33    | <a href="#">gläubiger muslim</a>            |
| 33    | <a href="#">radikaler islamist</a>          |
| 32    | <a href="#">gläubige moslems</a>            |
| 31    | <a href="#">illegale immigranten</a>        |
| 31    | <a href="#">radikalens muslimen</a>         |
| 30    | <a href="#">antikapitalistische muslime</a> |
| 28    | <a href="#">eingereisten muslimen</a>       |

dear Muslims  
(very often direct address)

- » Liebe Muslime . » Ihr dürft jetzt auf keinen Fall an Fleischbällchen
- » Liebe Muslime , ich bin um
- » liebe Muslime , alles gut
- Ich liebe Muslime ❤️ ❤️ ❤️
- : » Liebe Muslime , wenn ihr euch durch unsere Barmherzigkeit beleidigt
- » Liebe Muslime , ich fühle mich vom m Ramadan ausgegrenzt . » Können
- » Liebe Muslime , was machen eure Glaubensbrüder hier ? » Ich dachte
- » Liebe Muslime , liebe Araber ! » Meine Weihnachtsbotschaft an euch
- : » Liebe Muslime , bitte respektiert das Recht auf #ReligionsFREIHEIT
- » Liebe Muslime , wir Katholen haben den schönen Brauch der Exkommunikation
- » Liebe Muslime , meint ihr jetzt die Geschichten oder den Koran ?
- slime in Deutschland , die israelische Regierung ist Scheiße
- slime , wir europäer fordern euch dorthin zurück zu kehren
- » liebe muslime , wir europäer fordern euch dorthin zurück zu kehren
- ike liebe muslime , dass ihr nichts damit zu tun ... » <https://t.co/hRYAq8c9wL>
- » Liebe Muslime , ich brauche Eure Hilfe . » Ich bitte um die Telefonnummer
- : » Liebe Muslime , lasst uns mit diesen Deutschen nicht allein ? » Ihr
- inz liebe Muslime verprügeln und die Polizei nichts tut , weil es ja
- .er liebe Muslime wurden 6 Millionen Juden getötet . » » Und alle so
- » Liebe Muslime wenn ihr euch irgendwie gestört fühlt dann verschwindet
- » Liebe Muslime , dieser Mann entfernt Eure Tattoos gegen 50 Koranverse

# Searching for negative (Q-) attributes only

[Execute Search](#)

|   |                         |                         |                         |
|---|-------------------------|-------------------------|-------------------------|
| <input type="radio"/> 1   | <input type="radio"/> + | <input type="radio"/> ? | <input type="radio"/> * |
| Word <input type="text"/> <input type="checkbox"/> ↵  |                         |                         |                         |
| <input type="checkbox"/> ↵case, <input type="checkbox"/> ↵diacritics  |                         |                         |                         |
| Base <input type="text"/> <input type="checkbox"/> ↵  |                         |                         |                         |
| <input type="checkbox"/> ↵case, <input type="checkbox"/> ↵diacritics  |                         |                         |                         |
| Extra Q- <input type="text"/> <input type="checkbox"/> ↵  |                         |                         |                         |
| <input type="checkbox"/> ↵case, <input type="checkbox"/> ↵diacritics  |                         |                         |                         |
| Part of Speech - <input type="checkbox"/> ↵   |                         |                         |                         |
| <input type="checkbox"/> Noun<br><input type="checkbox"/> Proper Noun<br><input checked="" type="checkbox"/> Adjective<br><input type="checkbox"/> Pronoun more<br><input type="checkbox"/> Article<br><input type="checkbox"/> Verb<br><input type="checkbox"/> Adverb<br><input type="checkbox"/> Others more   |                         |                         |                         |
| Morphology + <input type="checkbox"/> ↵   |                         |                         |                         |
| Function - <input type="checkbox"/> ↵   |                         |                         |                         |
| <input type="checkbox"/> Subject more<br><input type="checkbox"/> Object more<br><input type="checkbox"/> Predicative more<br><input type="checkbox"/> Adverbial more<br><input type="checkbox"/> Arg. of prep. more<br><input type="checkbox"/> Adnominal more<br><input checked="" type="checkbox"/> Prenominal |                         |                         |                         |

|   |                         |                         |                         |
|---|-------------------------|-------------------------|-------------------------|
| <input type="radio"/> 1   | <input type="radio"/> + | <input type="radio"/> ? | <input type="radio"/> * |
| Word <input type="text"/> (Flüchtling Einwa) <input type="checkbox"/> ↵   |                         |                         |                         |
| <input type="checkbox"/> ↵case, <input type="checkbox"/> ↵diacritics  |                         |                         |                         |
| Base <input type="text"/> <input type="checkbox"/> ↵  |                         |                         |                         |
| <input type="checkbox"/> ↵case, <input type="checkbox"/> ↵diacritics  |                         |                         |                         |
| Extra [HA][a-z]+ <input type="checkbox"/> ↵   |                         |                         |                         |
| <input type="checkbox"/> ↵case, <input type="checkbox"/> ↵diacritics  |                         |                         |                         |
| Part of Speech - <input type="checkbox"/> ↵   |                         |                         |                         |
| <input checked="" type="checkbox"/> Noun<br><input type="checkbox"/> Proper Noun<br><input type="checkbox"/> Adjective<br><input type="checkbox"/> Pronoun more<br><input type="checkbox"/> Article<br><input type="checkbox"/> Verb<br><input type="checkbox"/> Adverb<br><input type="checkbox"/> Others more |                         |                         |                         |
| Morphology + <input type="checkbox"/> ↵   |                         |                         |                         |
| Function + <input type="checkbox"/> ↵   |                         |                         |                         |
| Semantic Role + <input type="checkbox"/> ↵  |                         |                         |                         |

[Dep Head](#) [Delete](#)

dangerous  
evil  
violence-prone  
anti-semitic  
fanatic  
militant  
violent  
anti-capitalist  
fundamentalist  
unmodern  
uneducated  
former  
criminal  
insane  
intolerant  
reactionary

|                                      |                       |       |
|--------------------------------------|-----------------------|-------|
| <a href="#">gefährlich</a>           | 19199.62 (16.85) 4.3% | [101] |
| <a href="#">böse</a>                 | 14908.36 (14.85) 3.8% | [89]  |
| <a href="#">gewaltbereit</a>         | 11746.38 (13.18) 3.4% | [79]  |
| <a href="#">antisemitisch</a>        | 7049.29 (17.69) 4.5%  | [106] |
| <a href="#">fanatisch</a>            | 5014.98 (12.18) 3.1%  | [73]  |
| <a href="#">militant</a>             | 2490.09 (10.51) 2.7%  | [63]  |
| <a href="#">gewalttätig</a>          | 2049.64 (5.51) 1.4%   | [33]  |
| <a href="#">antikapitalistisch</a>   | 1693.91 (5.01) 1.3%   | [30]  |
| <a href="#">fundamentalistisch</a>   | 1506.35 (8.18) 2.1%   | [49]  |
| <a href="#">unaufgeklärt</a>         | 1176.33 (4.18) 1%     | [25]  |
| <a href="#">ungebildet</a>           | 1084.10 (4.01) 1%     | [24]  |
| <a href="#">ehemalig</a>             | 995.64 (3.84) 0.9%    | [23]  |
| <a href="#">kriminell</a>            | 914.28 (16.85) 4.3%   | [101] |
| <a href="#">irr</a>                  | 910.95 (3.68) 0.9%    | [22]  |
| <a href="#">intolerant</a>           | 679.45 (6.34) 1.6%    | [38]  |
| <a href="#">reaktionär</a>           | 679.44 (3.17) 0.8%    | [19]  |
| <a href="#">durchgeknallt</a>        | 609.81 (3.01) 0.7%    | [18]  |
| <a href="#">link</a>                 | 543.93 (2.84) 0.7%    | [17]  |
| <a href="#">konvertiert</a>          | 497.82 (3.84) 0.9%    | [23]  |
| <a href="#">islamistisch</a>         | 423.47 (2.51) 0.6%    | [15]  |
| <a href="#">Judenhassend</a>         | 368.89 (2.34) 0.6%    | [14]  |
| <a href="#">verfassungsfeindlich</a> | 339.72 (3.17) 0.8%    | [19]  |
| <a href="#">kulturfremd</a>          | 318.08 (2.17) 0.5%    | [13]  |
| <a href="#">pädophil</a>             | 271.02 (2.01) 0.5%    | [12]  |
| <a href="#">anschlagsbereit</a>      | 271.02 (2.01) 0.5%    | [12]  |
| <a href="#">gestört</a>              | 188.21 (1.67) 0.4%    | [10]  |
| <a href="#">integrationsunwillig</a> | 188.21 (1.67) 0.4%    | [10]  |
| <a href="#">judenfeindlich</a>       | 152.45 (1.51) 0.3%    | [9]   |
| <a href="#">faschistoid</a>          | 152.45 (1.51) 0.3%    | [9]   |
| <a href="#">rückständig</a>          | 152.45 (1.51) 0.3%    | [9]   |
| <a href="#">hasserfüllt</a>          | 152.45 (1.51) 0.3%    | [9]   |
| <a href="#">arm</a>                  | 144.70 (10.35) 2.6%   | [62]  |
| <a href="#">illegal</a>              | 128.85 (16.85) 4.3%   | [101] |
| <a href="#">ausländisch</a>          | 120.45 (1.34) 0.3%    | [8]   |
| <a href="#">homophob</a>             | 120.45 (1.34) 0.3%    | [8]   |
| <a href="#">hassend</a>              | 120.45 (1.34) 0.3%    | [8]   |

# Emoticons & emojis: Sentiment & (Non-) Literality

| Emoticon category             | types by rank (other: < 1%)  |
|-------------------------------|--|
| emo-happy                     | 😊 :) 😊😺😸:-) 😊😊 😊 :3 :)) other: :-)) :]:;-]:}:-]  |
| emo-love                      | 😍😊,  😊<3 😊 😊  😊😊😊😊 |
| emo-laugh                     | 😂😂😂😂😂😂😂  |
| emo-sad                       | 😭😭:( 😢😢:-(`😢😢😢😢😢😢 other: :< :-[ :@ :[ :@   |
| emo-horror                    | 😱😱😱😱😱😱   |
| emo-angry                     | 😡😡💩😡😡😡😡 other: 😠 X(  |
| emo-surprise                  | 😲😲😲😲😲😲   |
| emo-wink                      | 😉😉😉😉😉😉   |
| emo-skeptical                 | 🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔   |
| emo-other<br>(emoticons only) | ;-)) ;)) <333 (intensifiers)<br>;-(` (combination wink+sad)<br>8D  |

# Emojis as markers for indirect HS

1. HS using key words such as ethnic slurs (e.g. *nigger*) and stereotypical attributes (*terrorist, criminal*) are easy to find / filter out, but **indirect HS is more difficult** (GermEval, Wiegand et al. 2018)
2. **Emojis can compensate for missing key words** (similar word embeddings, Pereira-Kohatsu et al. 2019), e.g.:
  - (a) **direct:** *Aber ja , holt noch mehr Musels nach Deutschland rein ... 😞*  
(But yes, go get even more Muslims to Germany)
  - (b) **indirect:** *Rentner sammeln Flaschen und Flüchtlinge leben auf großem Fuß 😞👉*  
(Pensioners collect bottles, refugees live rich)
  - (c1) **irony:** *Unsere fiese, unmenschliche Regierung ist einfach zu gemein zu den armen Flüchtlingen. :(*  
(Our nasty, inhumane government is simply sooo mean to the poor refugees)
  - (c2) **irony:** *Für die « Schutzsuchenden » gibt HH ca. 1 Milliarde pro Jahr aus . 😒 Man muss halt Prioritäten setzen . 😒*  
(For the "refugee-seekers" Hamburg spends 1 billion a year. Well, you need to have priorities)
  - (c3) **sarcasm:** *Die Asylanten sind fachkompetent, ganz lieb und es sind keine Kriminelle bzw. Terroristen mit dabei. Wurde uns so versprochen 😒* (The asylum-seekers are professionally educated, nice people. No criminals or terrorists among them. Or so we were promised.)

# Irony, search-constrained: lemma = "emo-(skeptical|laugh).\*"

- . ▫ Die Dame ist ja übrigens selbst **Immigrantin** aber stolz auf FPÖ 😂 😂 😂
- Die unbegleiteten minderjährigen **Flüchtlinge** sollten sich doch häufiger rasieren damit das glaubhaft
  - twittername ▫ ich bin Au unaccompanied under-age refugees should shave more often ...
  - Die **Flüchtlinge** sind schuld daran , dass jetzt kein Schnee fällt ...
- Ich bin für die Aufnahme dieser **Flüchtlinge** . 😊 ▫ <https://t.co/ePtYRo9vIe>
- Herr Altmaier will die Leitkultur zusammen mit **Einwanderern** erarbeiten ... 😂 😂 😂 ▫ via twittername
  - immer Konfliktpotential . ▫ Merkel glaubt das der **Islam** sich grundgesetzkonform an die Kette legen lässt . 🤔
    - **Flüchtling** kehrt erholt vom m #Heimaturlaub nach Deutschland zurück
  - Menschen zu gut , da es in deren Städten wenig **Ausländer** gibt ? 🤪 ▫ Auch lustig weil man einerseits behauptet
    - Der Grund bleibt derselbe - **Immigrantes** . 😊
  - Kommunen warnen vor Nachzug von **Flüchtlingsfamilien** ! 🤪
    - **Islamkritik** ohne Aufwärmung der #Beschneidungsdebatte scheint nicht
  - twittername ▫ Weil der **Islam** nur von ahmadiyyas besteht bleib logisch 🤪 👍 🔴 ein
- #Befristung : Wie der Staat aus den meisten **Flüchtlingen** Einwanderer macht . 🤪 ▫ #kontrollverlust #Fluechtlinge
- Wie der Staat aus den meisten Flüchtlingen **Einwanderer** macht . 🤪 ▫ #kontrollverlust #Fluechtlinge
  - leugnen . ▫ So sind sie halt , die kriminellen **Ausländer** ! 😂
  - How the state turns most refugees into immigrants

# Irony, search-constrained: lemma = "emo-(skeptical|laugh).\*" + ADJ-prenominal marked as Q+ sentiment

▫ twittername ▫ Der friedliche und **weltoffene Islam** 😂 😂 😂 😂 😂

90er , hat man sowas nicht gelesen 🤦 Wohl damals **wirkliche Flüchtlinge** .

▫ Ach , die Reisewarnung betrifft also nur **reiche Ausländer** 😂 😂 😂

▫ twittername ▫ 😂 Folgerung : **gläubige Islamisten** erzeugen Terror ( sonst wären sie nicht in Haft ) .

▫ twittername ▫ Diese Würstel suchen sich **handverlesene Flüchtlinge** aus und öffnen den Mund 😂

▫ Schäuble war ja auch d Meinung , ohne diese **grossartigen Einwanderung** würde Europa an Inzucht zugrunde gehen 😂 ▫ ht tolerant Islam: Indonesia threatens ... Der **tolerante Islam** 😂 😂 😂 ▫ Wegen obszöner Bildchen : Indonesien drohte

▫ twittername ▫ vielleicht die **strammen Moslems** unter ihnen 😁 😁 😁 😈 ? ▫ Das scheinen d Eltern zu betwittername ▫ « Da klappt es überwiegend mit den **neuen Ausländern** » 😊 in Einzelhaft bekommt man wohl nicht soviel mit

▫ twittername ▫ Ist es nicht vielleicht **besser Flüchtlinge** auf sicherem Weg in unser Land zu holen , als sie au

▫ twittername ▫ Und noch einer für unsere **gläubigen Muslime** 😂 😂 😂 ▫ Wenn 's heißt , ein Paradies mit Huris wink Schutzberechtigten strittig , nicht der der **anerkannten Flüchtlinge** ... 😊

▫ twittername ▫ Wer schon des Öfteren mit simpel **strukturierten Moslems** zu tun , wundert sich nicht . 😕

bei den #Jamaika Sondierungen : Nur 70.000 **neue Flüchtlinge** durch den #Familiennachzug . 😊 ▫ Abendessen in der

Because decent refugees keep their family together Weil bei **anständigen Flüchtlingen** die Familie zusammenbleibt 😊

▫ twittername ▫ Und die **minderqualifizierten Flüchtlinge** stellen wahrscheinlich genau den Typus

▫ Der **neue Ausländer** des twittername Jeremy Morin steht heute Abend im m

daran , dass Weihnachtsgeld eine Beleidigung für **gläubige Moslems** ist . 😈

▫ twittername ▫ Elmas die **Gläubige Muslima** 😂 😂 😂 😂 ▫ Hahahahaha die Linke Sau

zu wünschen , 🤦 mal überlegen , das stellt diese **ehrenwerten Flüchtlinge** auf eine Stufe mit dem IS .

zu wünschen , 🤦 mal überlegen , d Just think about it - that places these honourable refugees on par with ISIS

▫ twittername ▫ Freie Fahrt für **freie Islamisten** 😈 💡

# Part 3: Word embedding & Stereotypes

## Similarity ranking and clustering

1. word embeddings are

- data-based, machine-learned **word vectors**
- in a high-dimensional contextual space, with coordinates computed from co-occurring words
- where the position of a word approximates its meaning, distribution and similarity to other words.

2. we used word2vec & tensorflow software for modelling

- using not words, but lemmatized and POS-disambiguated corpus data
- CBOW for smaller subcorpora, **Skipgram** for the entire corpus

3. usage examples:

- **word similarity judgement**
- **synonym harvesting** for further corpus searches

4. e.g. human noun similarity measures for German "Moslem" [Muslim]

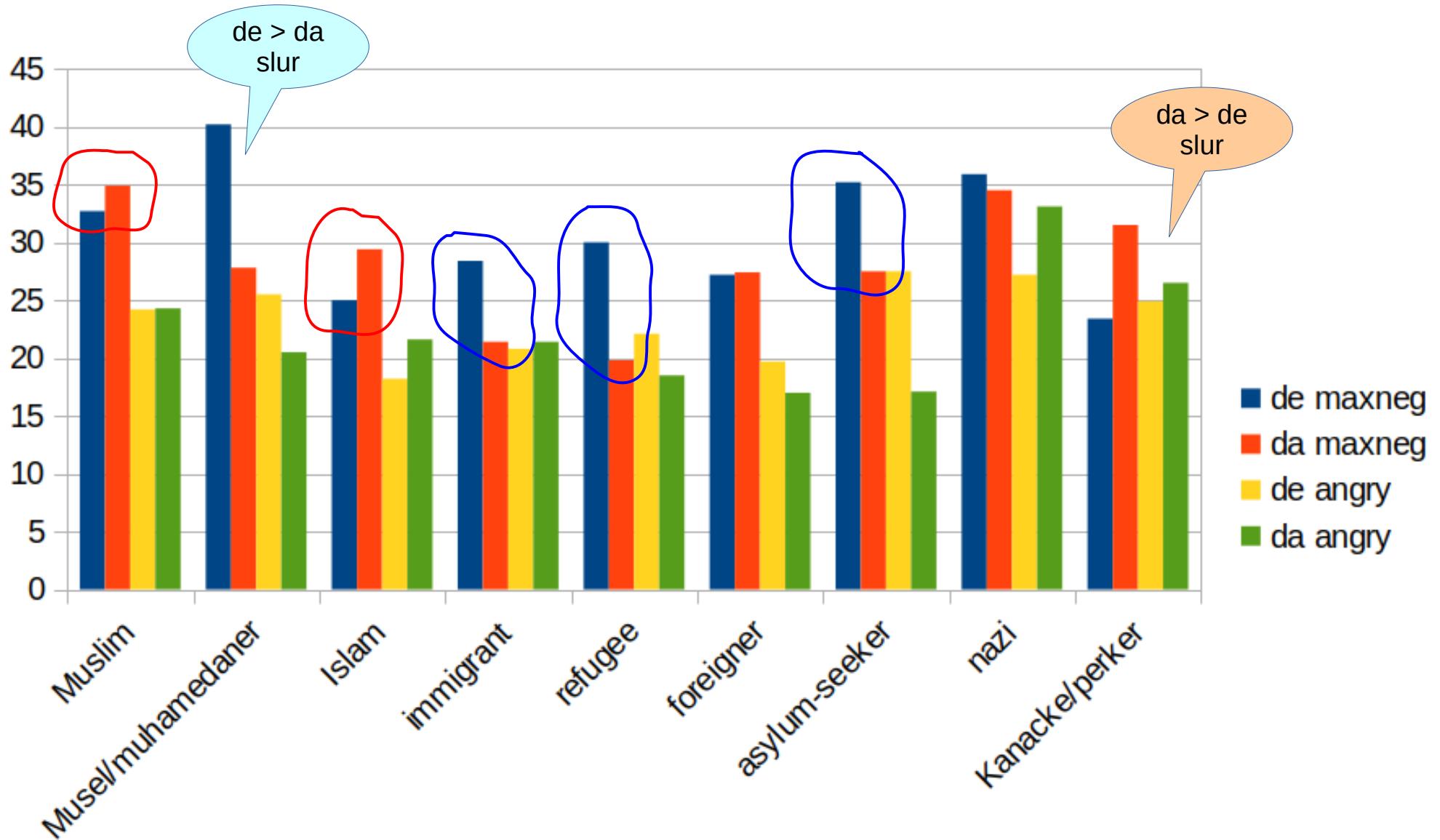
5. > 0.8: *Muslim* [Muslim, spelling variation]

- > 0.6: *Mohammedaner* [Mohammedan], *Muselman* [Mussulman]
  - *Jude* [Jew], *Hindu*
- > 0.5: *Araber* [Arab], *Islamist* [Islamist], *Glaubensbruder* [brother-in-faith], *Salafist*, *Mufti*, *Alevit*
  - *Konvertit* [convert], *Kuffar* [AR:infidel]
- > 0.4: *Schleiereule* ["veil owl" (barn owl)], *Marokkaner* [Moroccan], *Syrer* [Syrian], *Nafri* [North-Afrikan<sub>slur</sub>], *Bosnier* [Bosnian], *Dschihadist* [jihadi], *Serbe* [Serb]
  - *Israeli*, *Vietnamese*

6. Verb similarity for "Moslem"

- *durchdrehen* [loose it], *anspucken* [spit at sb.], *verkloppen* [beat up], *randalieren* [rioting], *bespucken* [spit at] (i20 skip)

# vector similarities for top negative emojis: what is worse: being muslim or an immigrant?



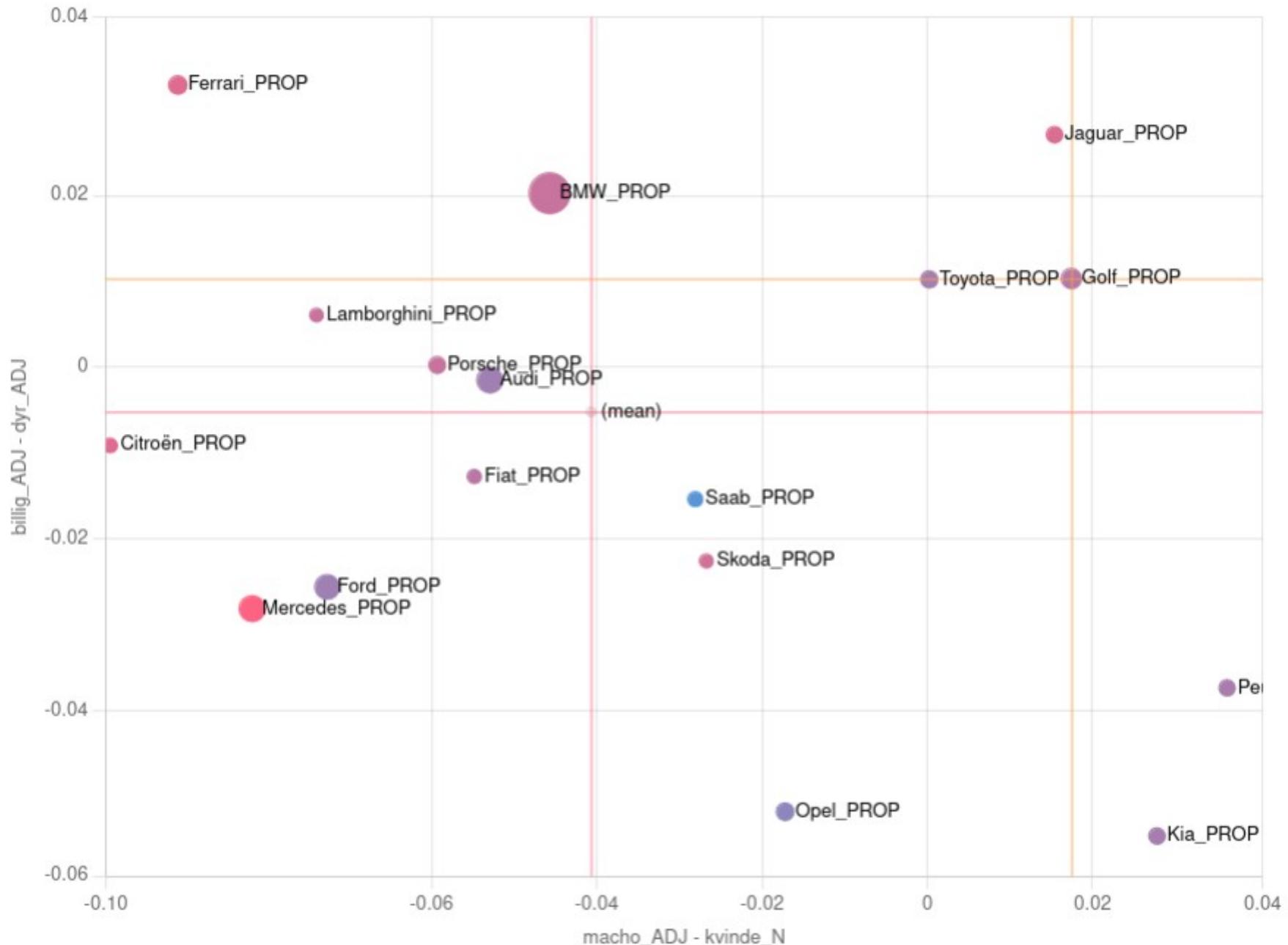
1. Vomiting emoji: **schlecht** [bad] 25.7, **böse** [evil] 27.8, **kriminell** [criminal] 43.3
2. Angry emoticon: **schlecht** [bad] 19.7, **böse** [evil] 24.8, **kriminell** [criminal] 26.5

# Stereotype interpretation of word vectors

- x- & y-axis for measuring stereotype association strength
  - ◆ polar/antonymic stereotype target for + or - values
  - ◆ defined by individual word vectors or word vector bundles
- Examples:
  - ◆ *mand - kvinde* [man - woman]
  - ◆ *ung - gammel* [young -old]
  - ◆ *forurenning; fossil - vedvarende; grøn; miljøvenlig* [polution; fossil – renewable; green; environment-friendly]
  - ◆ *sad/angry/skeptical emojis* (-) vs. *happy/love emojis* (+)
- scatter dot for potential **stereotype items** (word 1...50) computed as (x,y) with coordinates
  - ◆  $x = \text{SIM}(\text{mand}, \text{word-1}) - \text{SIM}(\text{kvinde}, \text{word-1})$
  - ◆  $y = \text{SIM}(\text{ung}, \text{word-1}) - \text{SIM}(\text{gammel}, \text{word-1})$

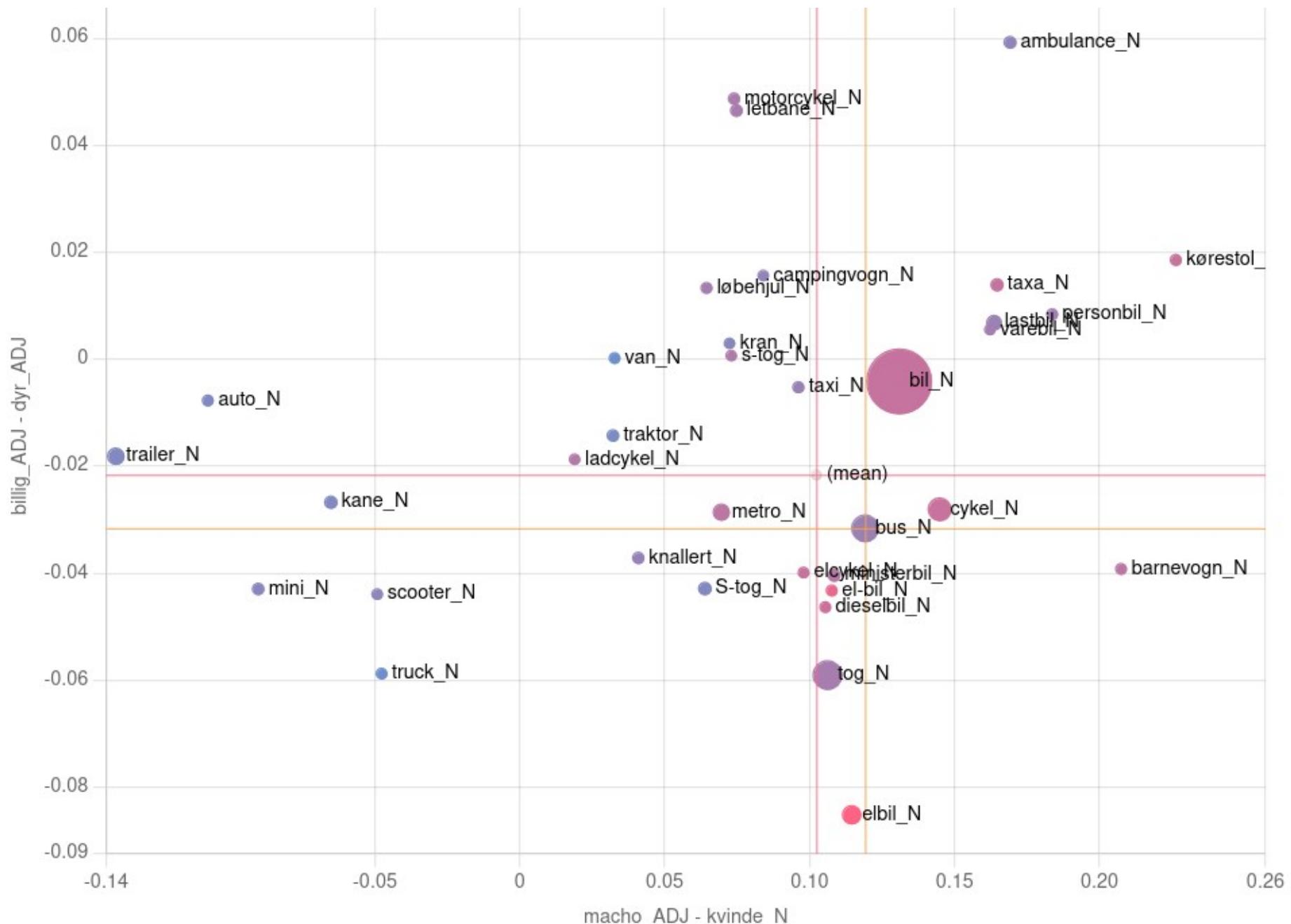
# car brands: gender and exclusivity

<Vground> PROP

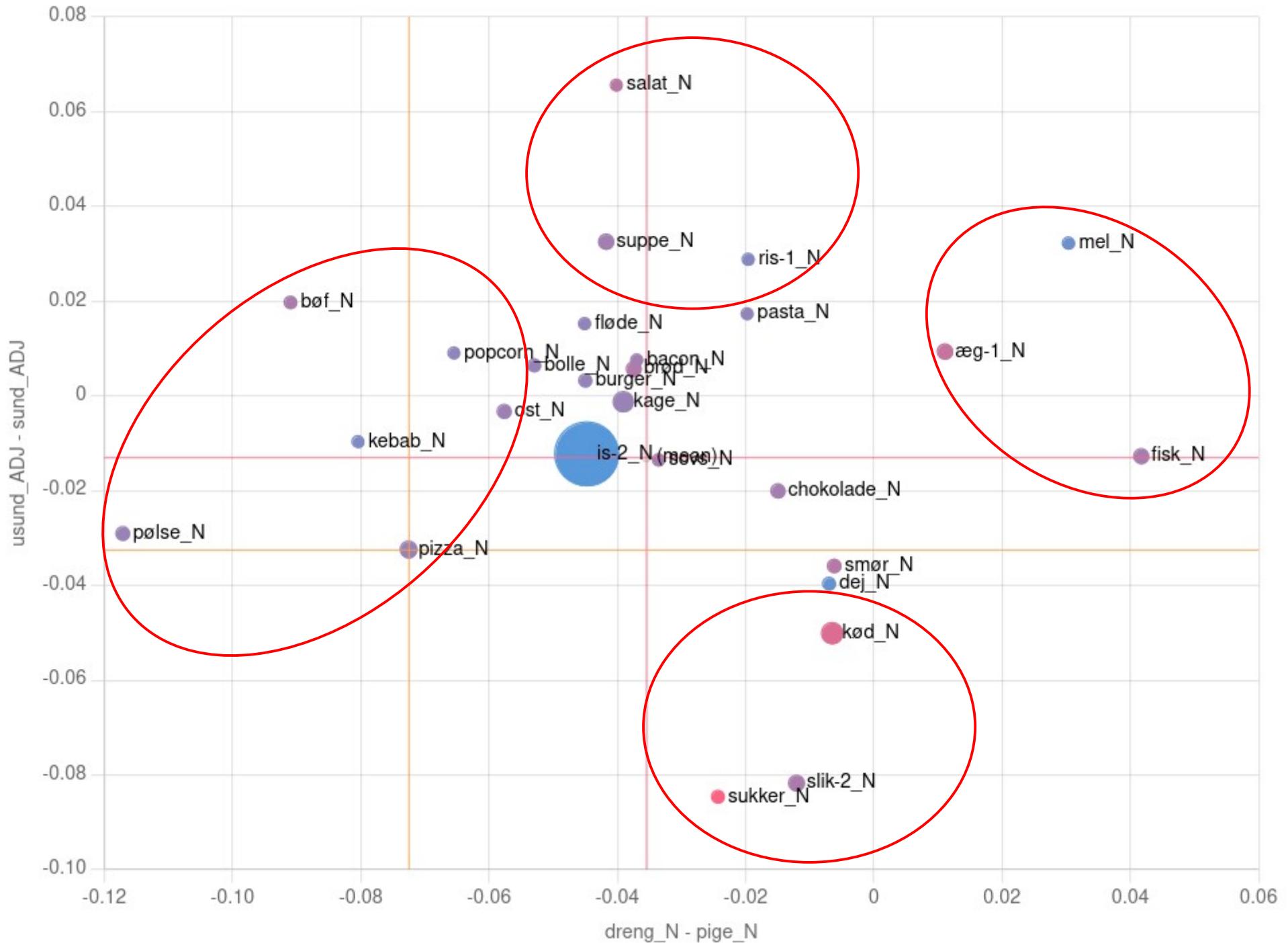


# Ground vehicles (all types): gender and exclusivity

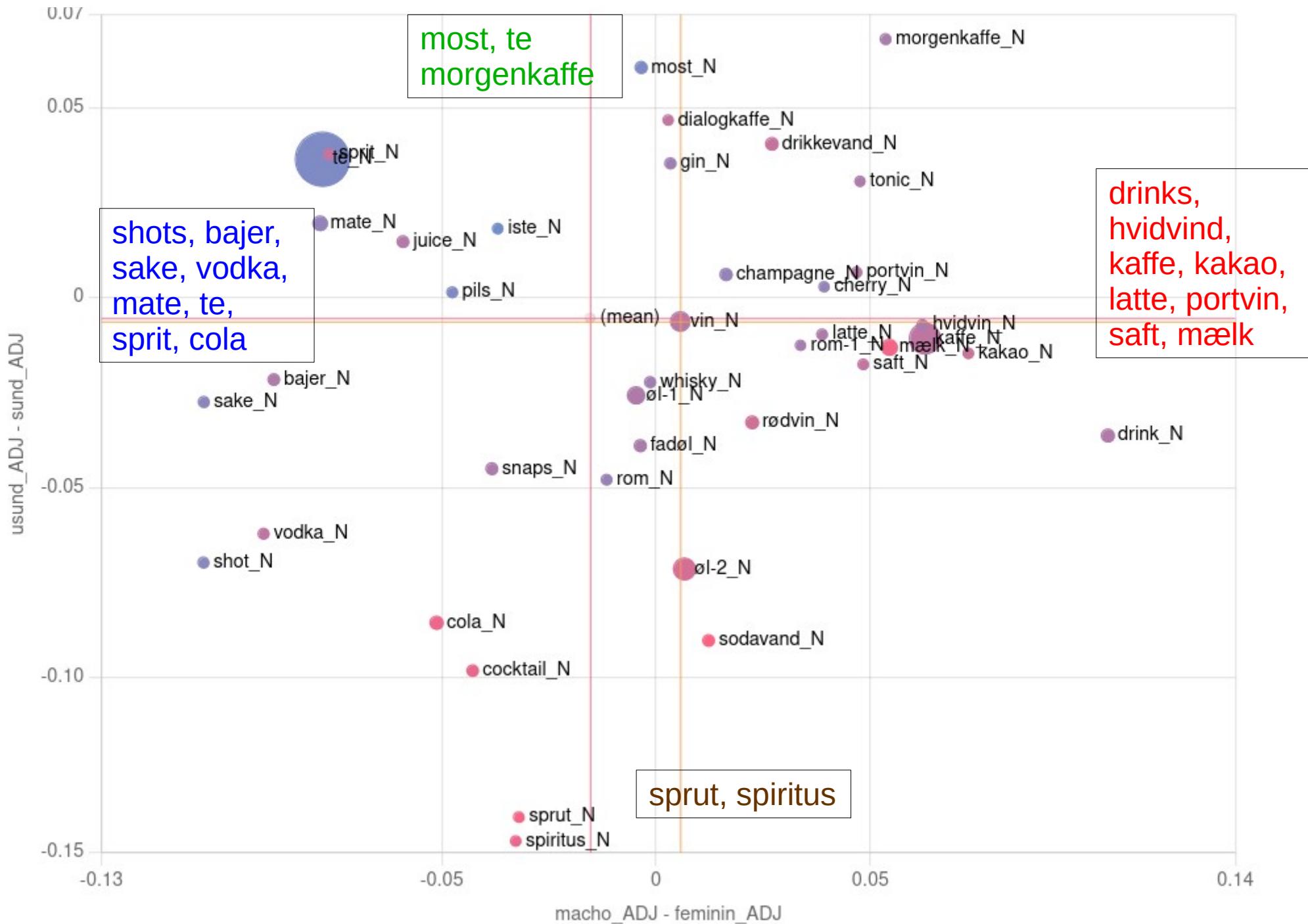
<Vground> N



# Food: gender and health

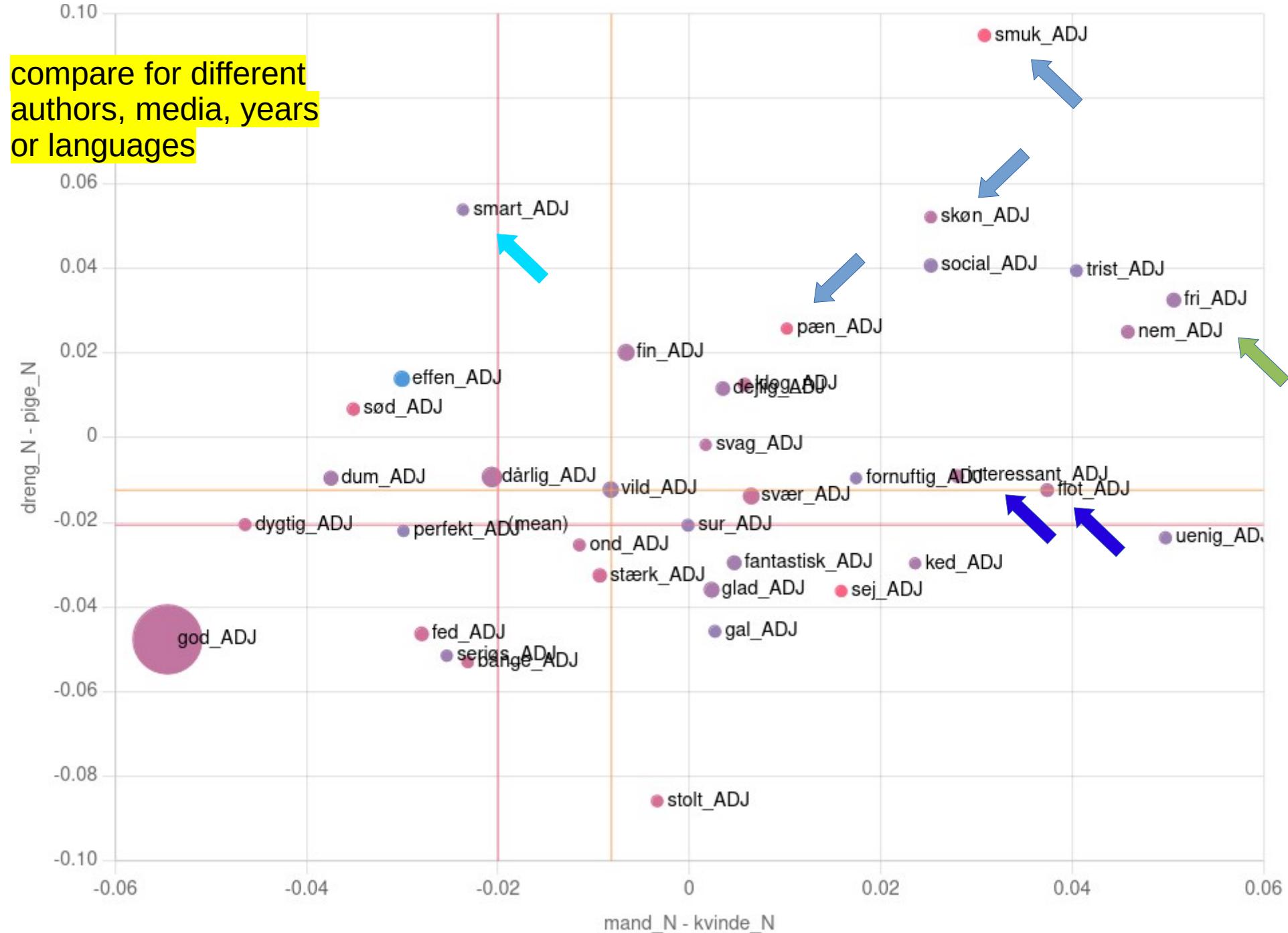


# Drink: gender and health



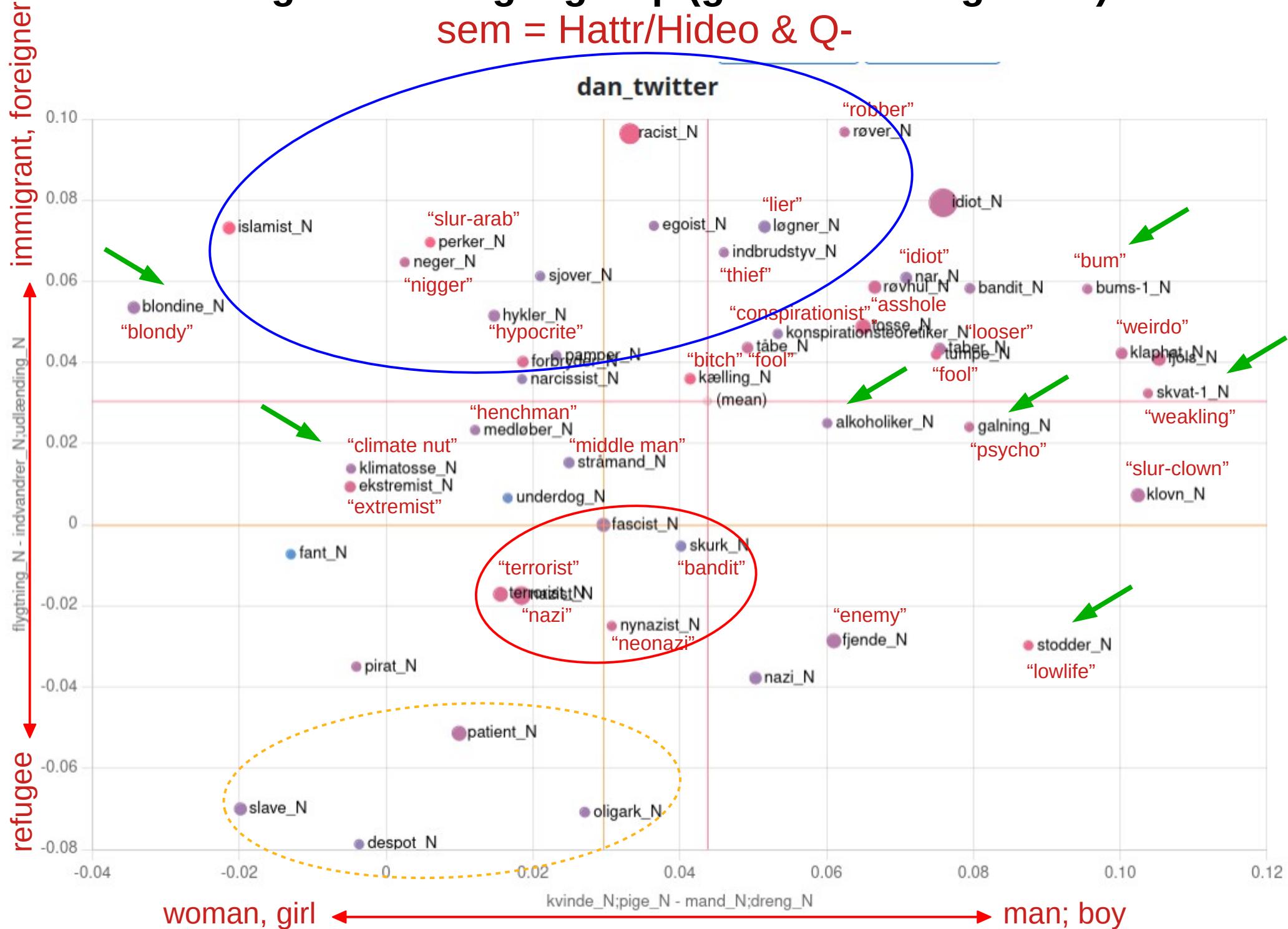
# Gender stereotypes: associated adjectives

[sem=".\*(?:|^|)(jqual|jpsych|jemo|jskill|jpower|jsoc|jcog|jcom|jsem)(?: |\$).\*" & pos="ADJ"]



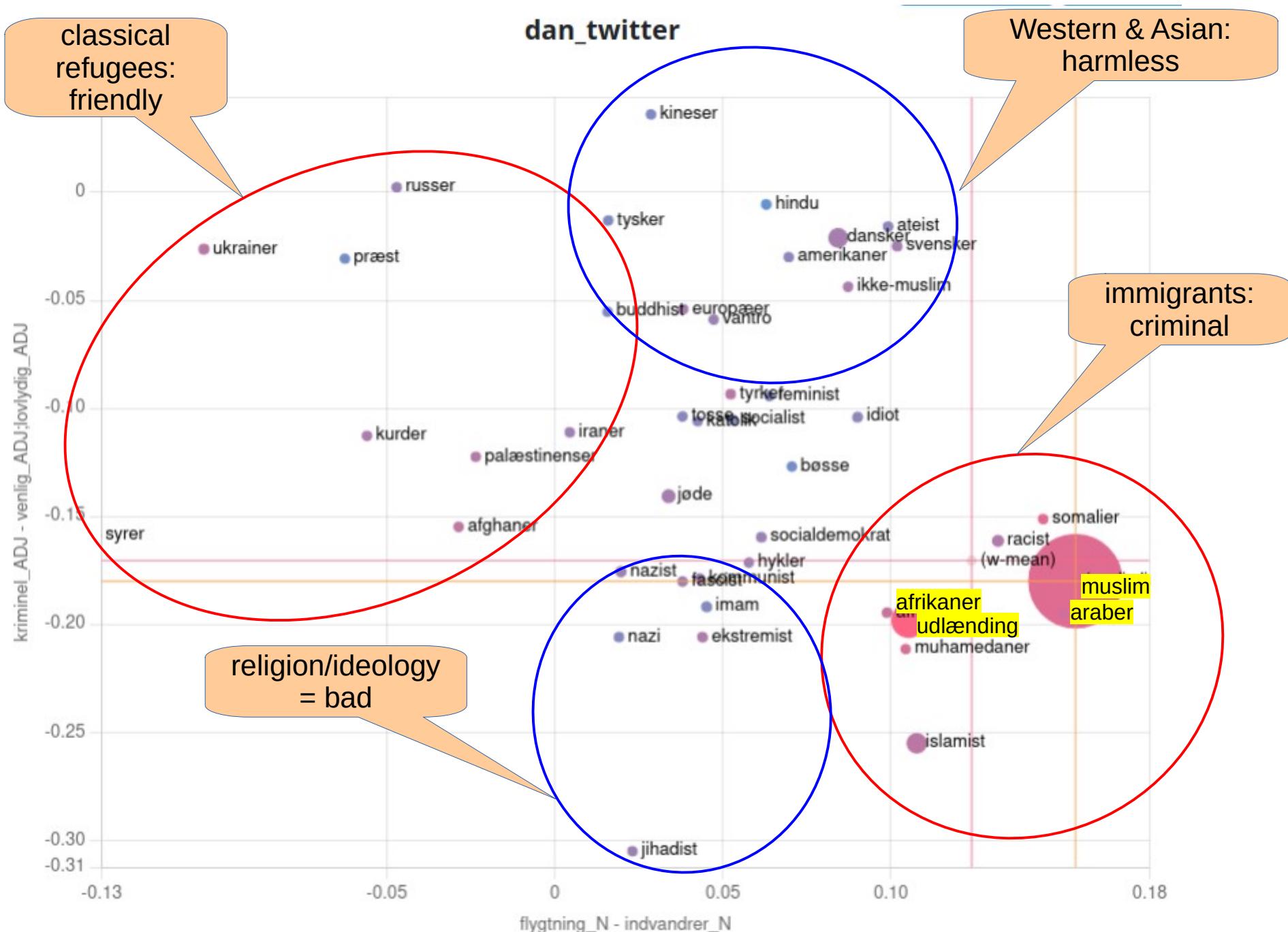
# Derogatives: Target group (gender / immigration)

sem = Hattr/Hideo & Q-



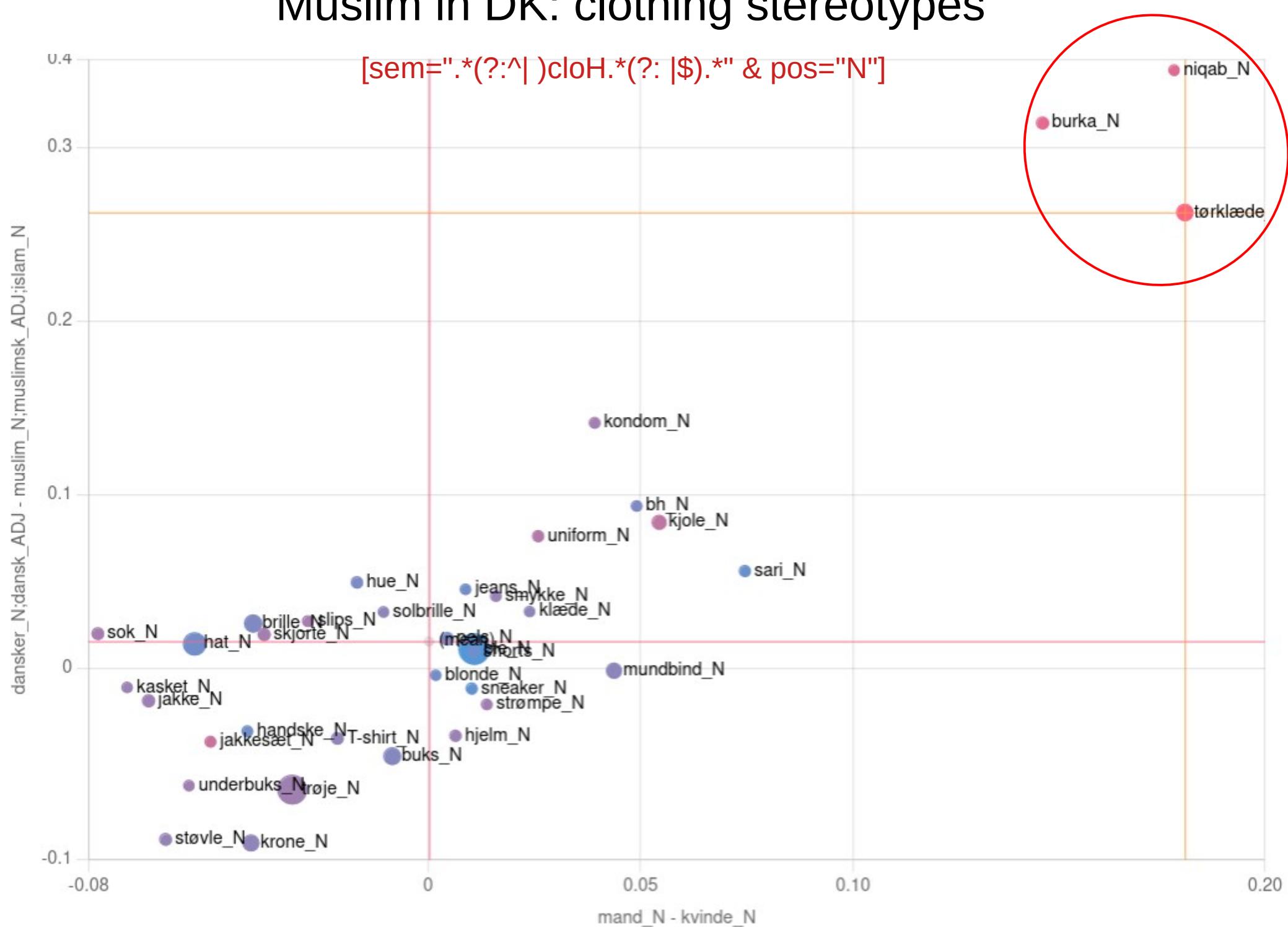
# Who are the foreigners and are they all criminal?

immi SQ sem = Hattr/Hideo/Hnat, refugee – immigrant, criminal - nice



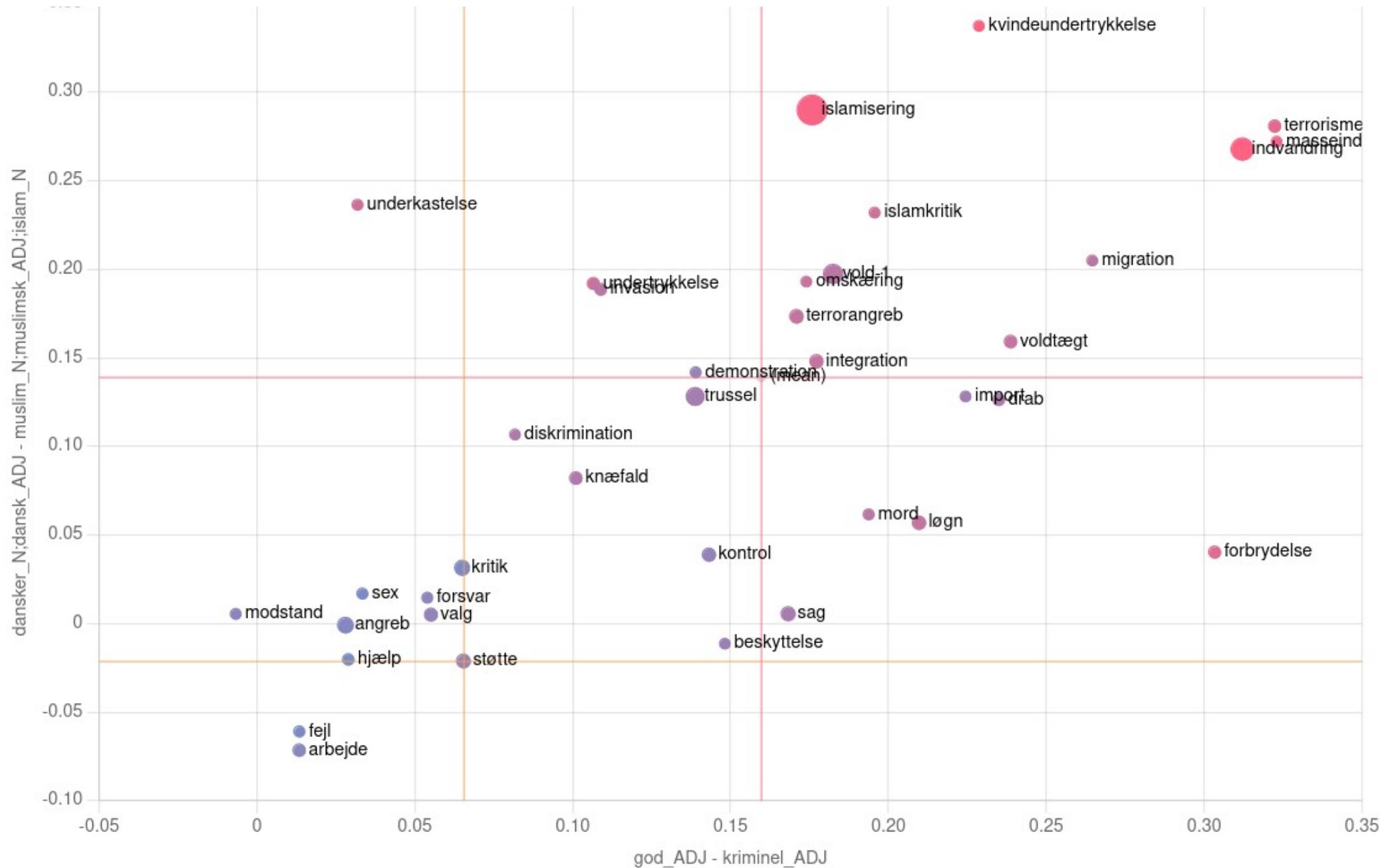
# Muslim in DK: clothing stereotypes

[sem=".\*(?:^| )cloH.(?: |\$).\*" & pos="N"]



# Muslim in DK: actions

[lex\_lc="(muslim|islam|m[ou]ham+ed|mosk[eé]).\*"] SQ [sem=".\*(?:^| )act.\*(?: |\$).\*"]



# **Part 4: Distant reading**

1. computational “overview” methods used for the interpretation of literary text or text collections
2. prepare focused inspection
3. compare authors, works or genres
4. topic identification
5. cast and character networks
6. linguistic patterns
7. vocabulary and lexical spread

# Literature: H.C.Andersen

1. find the most central human and animal protagonists in HCA texts!

- [sem=".\*(?:^| )[AH][a-z]+(?: |\$).\*"]]  
within <s author="Andersen"/>
- sort by relative frequency (global or in-corpus)

|                               |       |      |         |      |      |                                     |
|-------------------------------|-------|------|---------|------|------|-------------------------------------|
| <a href="#">dryade</a>        | 84.82 | 0.85 | 540.27  | 0.3% | 81   | <input checked="" type="checkbox"/> |
| <a href="#">jødepige</a>      | 80.42 | 0.47 | 226.78  | 0.1% | 34   | <input checked="" type="checkbox"/> |
| <a href="#">lappeskædder</a>  | 64.55 | 0.44 | 226.78  | 0.1% | 34   | <input checked="" type="checkbox"/> |
| <a href="#">bedstemoder</a>   | 60.15 | 1.74 | 1394.04 | 0.7% | 209  | <input checked="" type="checkbox"/> |
| <a href="#">røverpige</a>     | 53.09 | 0.57 | 260.13  | 0.1% | 39   | <input checked="" type="checkbox"/> |
| <a href="#">idiotbarn</a>     | 48.72 | 0.19 | 86.71   | 0.0% | 13   | <input checked="" type="checkbox"/> |
| <a href="#">spanierinde</a>   | 47.70 | 0.13 | 106.72  | 0.1% | 16   | <input checked="" type="checkbox"/> |
| <a href="#">signora</a>       | 44.40 | 0.25 | 340.17  | 0.2% | 51   | <input checked="" type="checkbox"/> |
| <a href="#">gudfader</a>      | 43.27 | 1.66 | 860.43  | 0.4% | 129  | <input checked="" type="checkbox"/> |
| <a href="#">generalinde</a>   | 35.48 | 0.08 | 233.45  | 0.1% | 35   | <input checked="" type="checkbox"/> |
| <a href="#">vægter</a>        | 35.26 | 0.94 | 980.49  | 0.5% | 147  | <input type="checkbox"/>            |
| <a href="#">moder</a>         | 32.58 | 2.17 | 7397.06 | 3.5% | 1109 | <input type="checkbox"/>            |
| <a href="#">elskede</a>       | 32.29 | 0.00 | 53.36   | 0.0% | 8    | <input type="checkbox"/>            |
| <a href="#">madam</a>         | 30.02 | 0.33 | 1200.60 | 0.6% | 180  | <input type="checkbox"/>            |
| <a href="#">signore</a>       | 28.26 | 0.11 | 146.74  | 0.1% | 22   | <input type="checkbox"/>            |
| <a href="#">rejsekammerat</a> | 27.17 | 1.38 | 733.70  | 0.3% | 110  | <input type="checkbox"/>            |

<H>

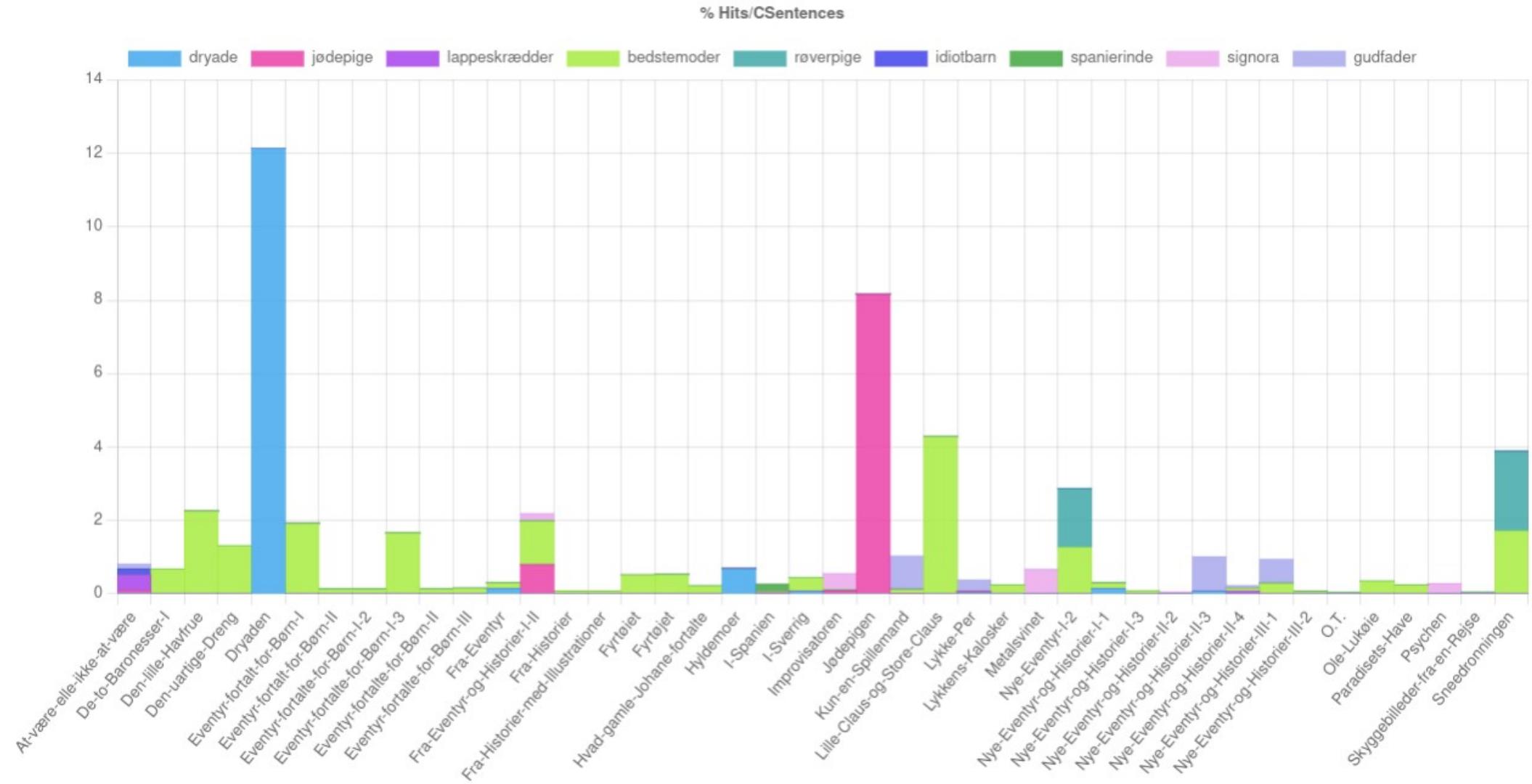
dan\_literature  
1 to 51 of 491 ( $\Sigma$  6514)

[Download TSV](#)

| Token                      | G: freq <sup>2</sup> /norm | C: freq <sup>2</sup> /norm | C: freq/corp · 10 <sup>8</sup> | freq/conc | num | Compare?                            |
|----------------------------|----------------------------|----------------------------|--------------------------------|-----------|-----|-------------------------------------|
| <a href="#">&lt;A&gt;</a>  |                            |                            |                                |           |     |                                     |
| <a href="#">skruptudse</a> | 3569.60                    | 19.92                      | 406.87                         | 0.9%      | 61  | <input checked="" type="checkbox"/> |
| <a href="#">køkkenkat</a>  | 1086.67                    | 6.36                       | 120.06                         | 0.3%      | 18  | <input checked="" type="checkbox"/> |
| <a href="#">lænkehund</a>  | 694.24                     | 10.52                      | 333.50                         | 0.8%      | 50  | <input checked="" type="checkbox"/> |
| <a href="#">skarnbasse</a> | 520.30                     | 10.52                      | 293.48                         | 0.7%      | 44  | <input checked="" type="checkbox"/> |
| <a href="#">stork</a>      | 482.03                     | 60.98                      | 1567.46                        | 3.6%      | 235 | <input checked="" type="checkbox"/> |
| <a href="#">oldenborre</a> | 459.58                     | 4.51                       | 233.45                         | 0.5%      | 35  | <input checked="" type="checkbox"/> |
| <a href="#">pengegris</a>  | 428.01                     | 9.19                       | 173.42                         | 0.4%      | 26  | <input checked="" type="checkbox"/> |
| <a href="#">spurveunge</a> | 393.27                     | 0.80                       | 33.35                          | 0.1%      | 5   | <input checked="" type="checkbox"/> |
| <a href="#">ørneunge</a>   | 334.77                     | 7.83                       | 160.08                         | 0.4%      | 24  | <input checked="" type="checkbox"/> |
| <a href="#">markmus</a>    | 297.01                     | 18.22                      | 380.19                         | 0.9%      | 57  | <input checked="" type="checkbox"/> |
| <a href="#">Anselmus</a>   | 258.61                     | 0.81                       | 26.68                          | 0.1%      | 4   | <input type="checkbox"/>            |
| <a href="#">gemse</a>      | 238.04                     | 12.91                      | 273.47                         | 0.6%      | 41  | <input type="checkbox"/>            |
| <a href="#">svale</a>      | 237.16                     | 25.66                      | 1133.90                        | 2.6%      | 170 | <input type="checkbox"/>            |
| <a href="#">svane</a>      | 221.19                     | 40.74                      | 1200.60                        | 2.8%      | 180 | <input type="checkbox"/>            |
| <a href="#">krage</a>      | 205.69                     | 22.96                      | 940.47                         | 2.2%      | 141 | <input type="checkbox"/>            |
| <a href="#">kukker</a>     | 149.76                     | 3.46                       | 93.38                          | 0.2%      | 14  | <input type="checkbox"/>            |
| <a href="#">kunstfugl</a>  | 142.60                     | 3.91                       | 80.04                          | 0.2%      | 12  | <input type="checkbox"/>            |
| <a href="#">hest</a>       | 111.41                     | 22.21                      | 3395.04                        | 7.8%      | 509 | <input type="checkbox"/>            |

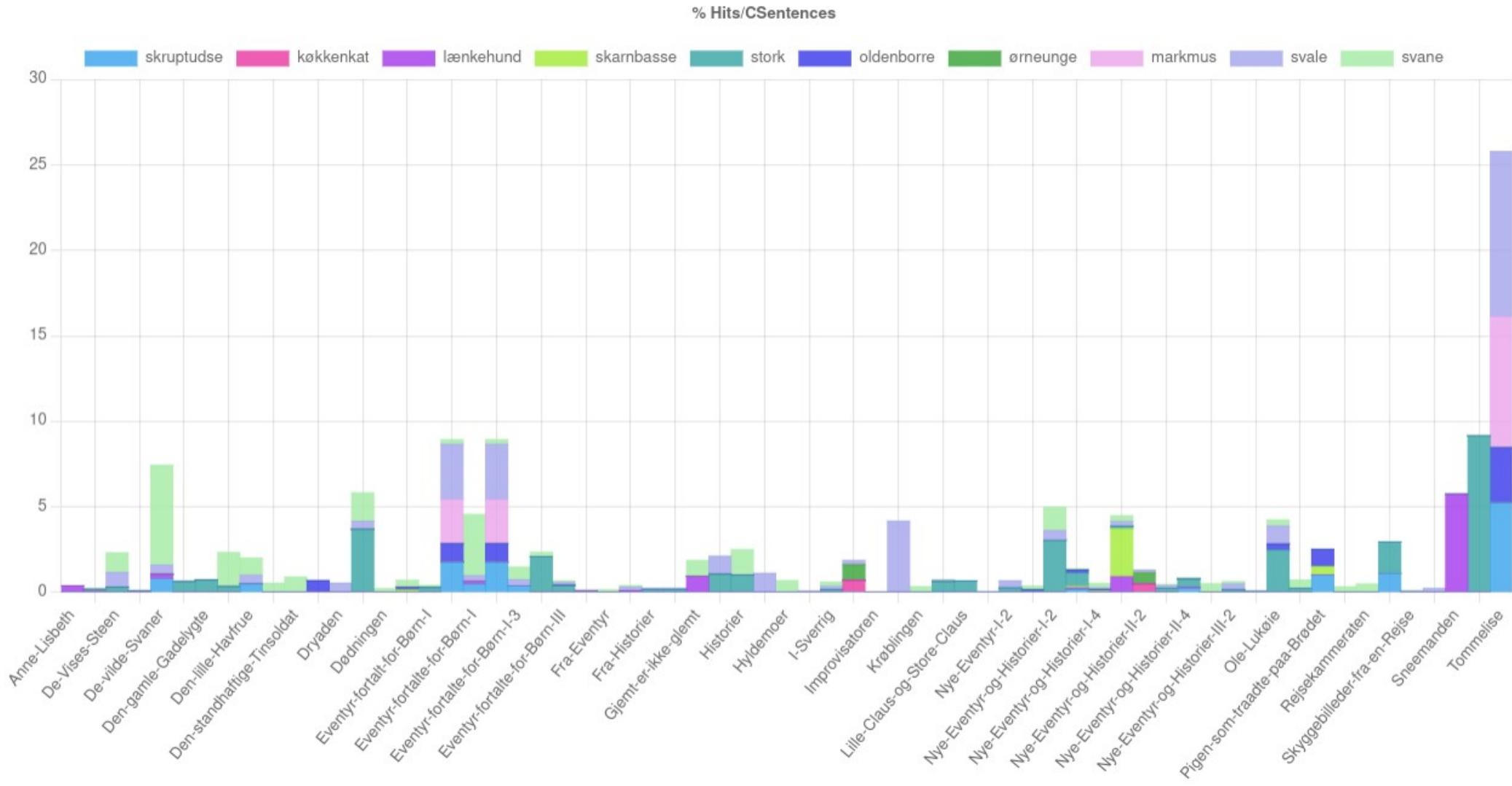
# H.C.Andersen works: human protagonists: which & where?

dan\_literature

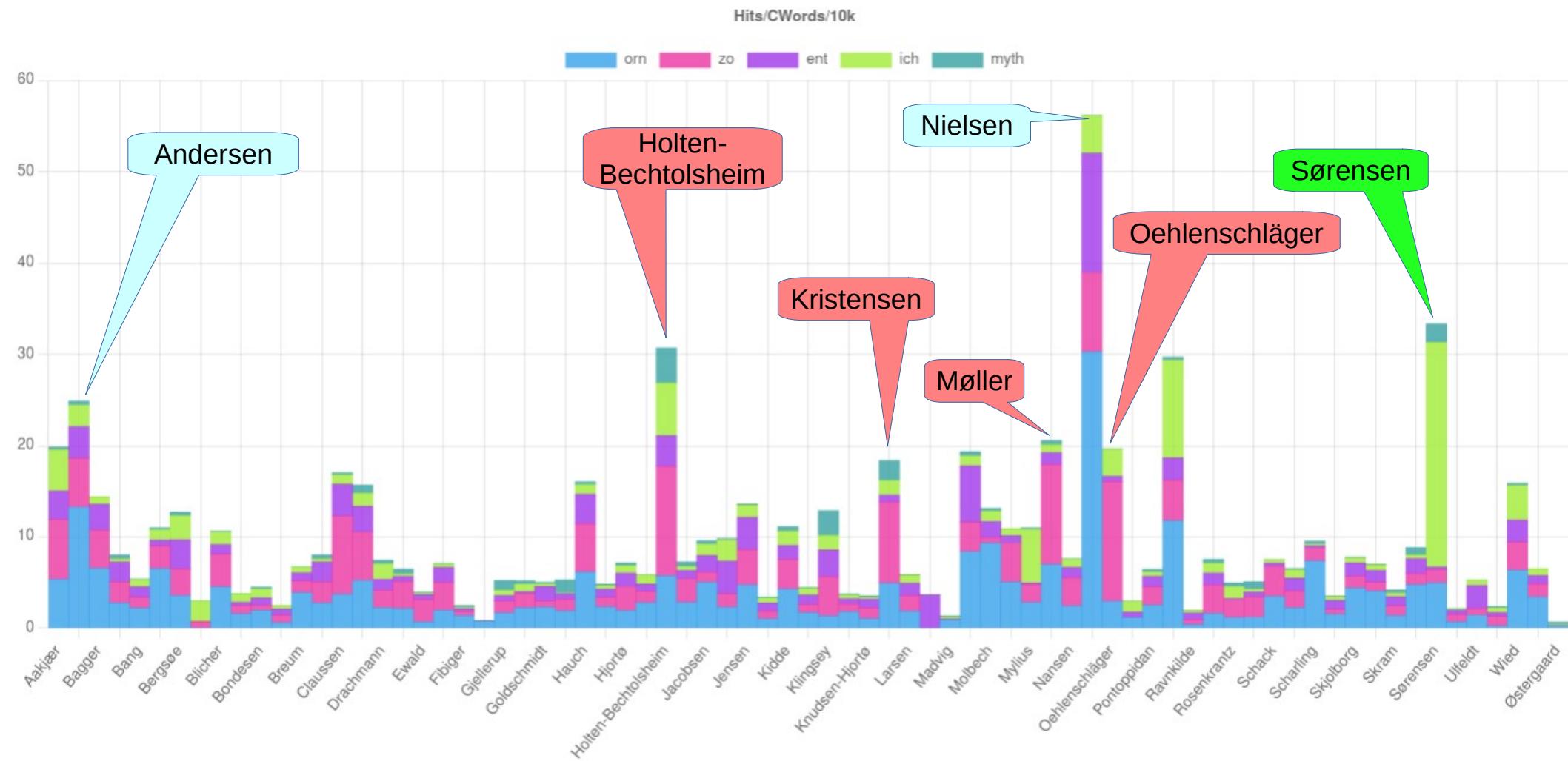


# H.C.Andersen works: animal protagonists: which & where?

## **dan\_literature**



# Non-domestic animals in Danish authors

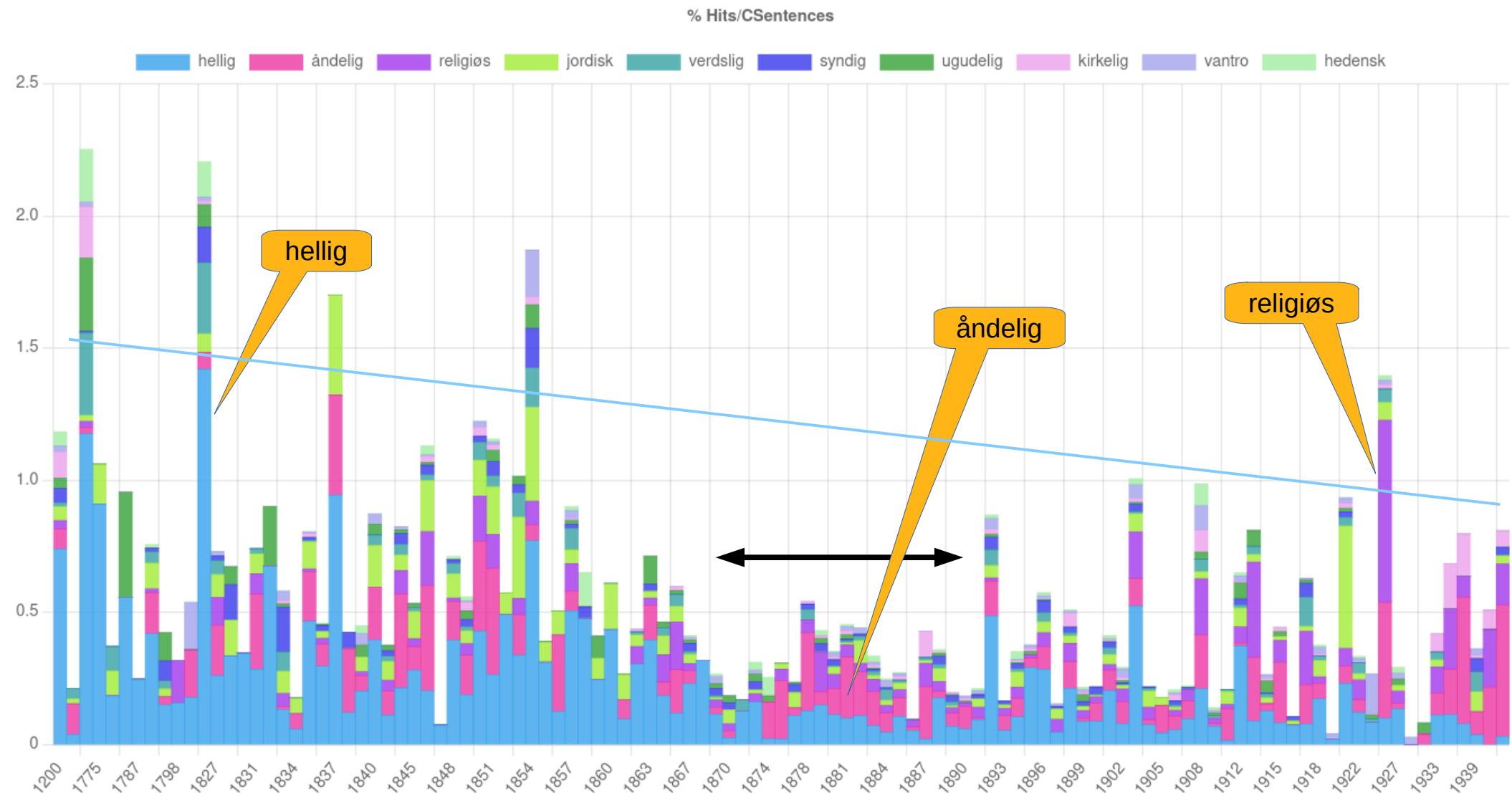


H.C. Andersen, Anton(ius) Nielsen: **birds <Aorn>**

Carl Sørensen: **fish <Aich>**

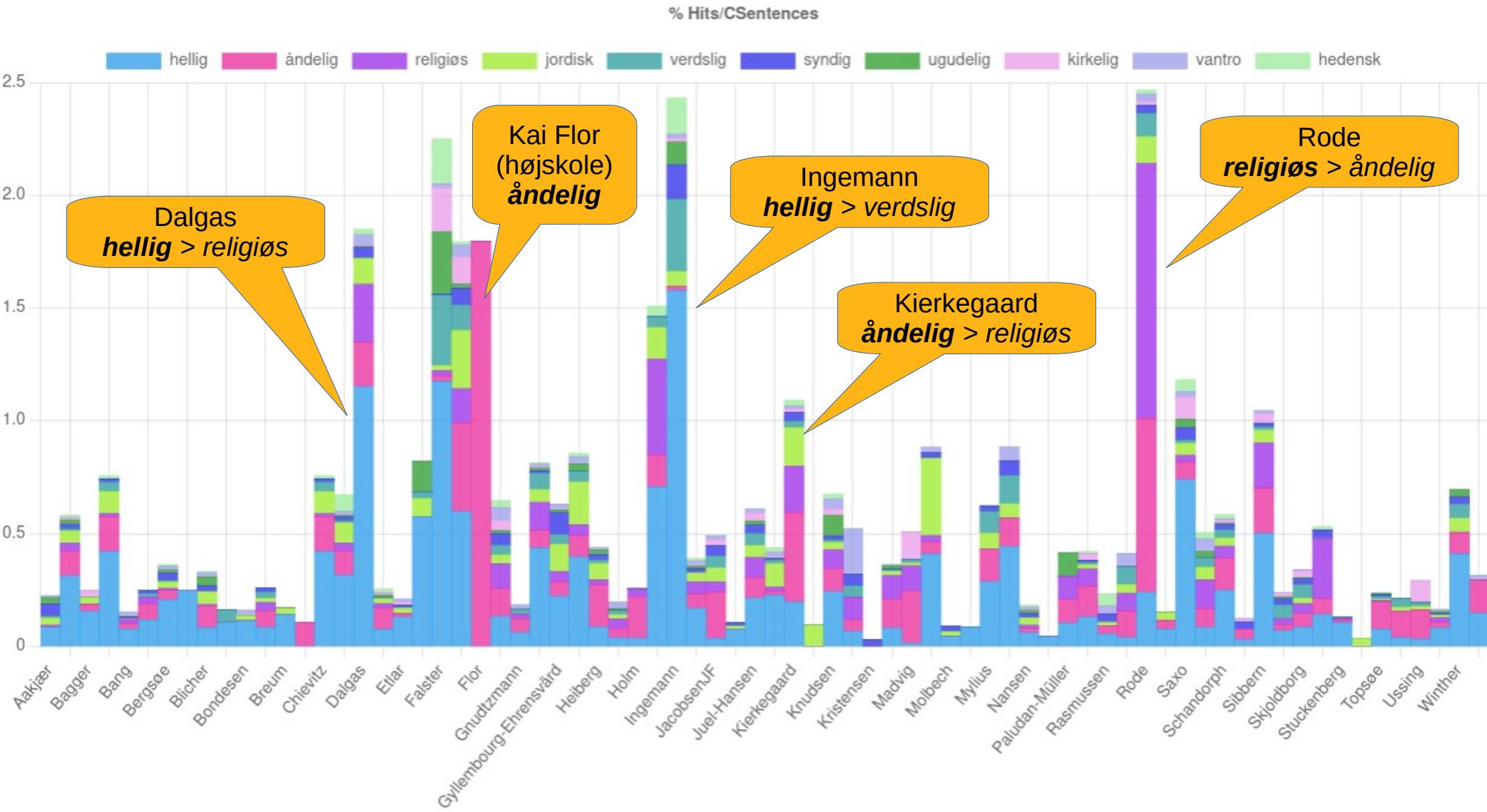
Adam Oehlenschläger, Tom Kristensen, Poul Martin Møller, Hans Holten-Bechtolsheim: **land animals <Azo>**

# Religion in Danish literature: time axis



- search: <jrel> / <Drel> adjectives
- less focus on religion over time
- lowest during modernism (“det moderne gennembrud” 1870-1890)
- ‘**hellig**’ (holy) replaced by ‘**åndelig**’ (spiritual) and the meta/topic adjective ‘**religiøs**’ (religious)

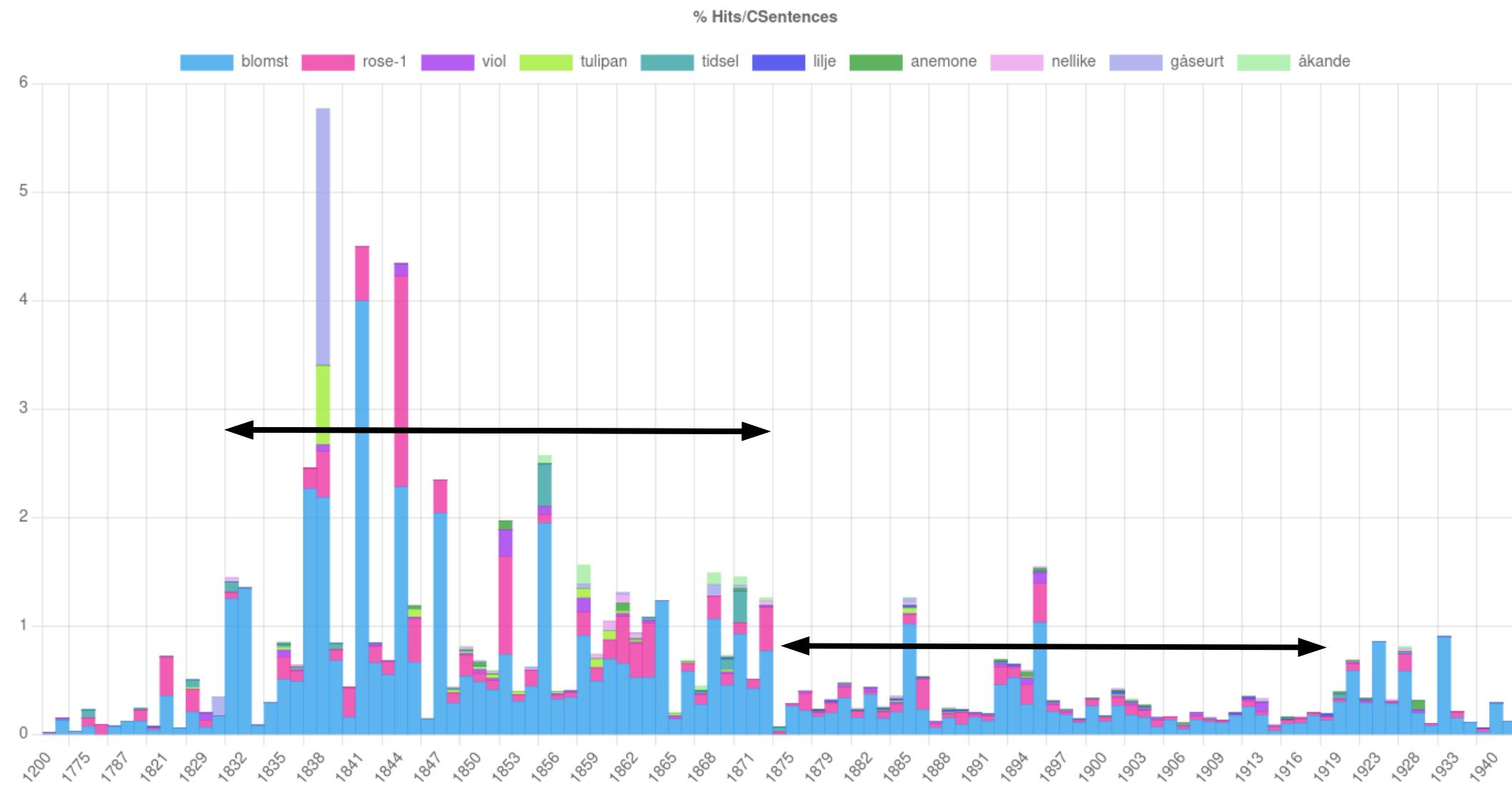
# Religion in Danish literature: authors



- **Latin & psalms:** Christian Falster 1690-1752 & B.S. Ingemann 1789-1862: *hellig > verdselig > ugudelig* (Ingemann: *syndig*)
- **the philosopher:** Ernesto Dalgas 1871-1899: *hellig > religiøs > åndelig*
- **lyric reflection:** Helge Røde 1870-1937: *religiøs > åndelig (> hellig)*

# Flowers in Danish literature: the Romance

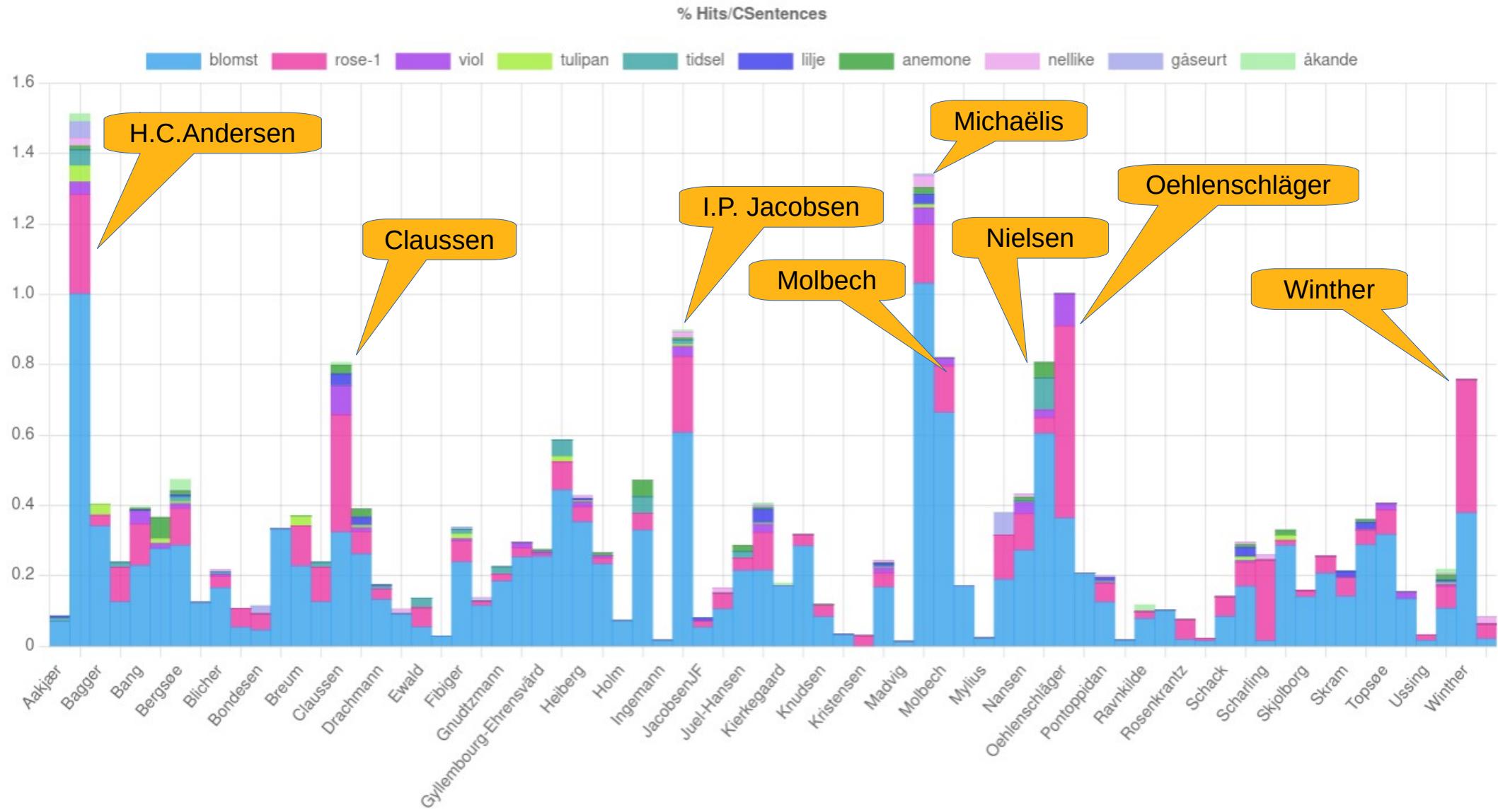
Search: Sem = <Bflo>, POS = N



- Romance period: flower power, often roses
- Modernism: fewer flowers

# Flowers in Danish literature: Authors

Search: Sem = <BFlo>, POS = N



- **Most (generic) flowers:** H.C. Andersen > Karin Michælis (> Chr. Molbeck > I.P. Jacobsen / Anton Nielsen)
- **Over 50% roses:** Adam Oehlenschläger > Chr. Winther / Sophus Claussen (> Henrik Scharling)

# quantitative → qualitative inspection

Claussen → “rose” → concordance

sin Søvn , begav de sig til Hallerne , friske som to unge Roser .

▫ Morgenens første Roser var de - den rødkindede , blomstrende Morgen , der opstaar

▫ Morgenens Roser !

▫ Maa jeg tage den ene af de blegrøde Roser ?

Farve , i Ord eller Gærning fortolket , at strø sine blege Roser duftende ud for alt Folket .

▫ Hvor lyser du af Sommer , unge Roser , muntre Kjoler !

▫ Her er Rosen , som jeg loved ( kysser ham ) .

▫ Det er Frøknerne Musman , Roser af det samme Blomsterbed

Marmortavler og Jærnkors var Ly for den solvarme Blæst og Roser , hvorhen han saa , fine , blegrøde Knopper og hvide ,

▫ Affaldne\_Purpurdynger fra store , ormædte Roser dækkede Gangene .

Et omhyggeligt dækket Aftensbord med varme Retter , Vin og Roser ved hver Kuvert var beredt de unge Gæster fra Aasum i et

og de gennemsigtige Gardiner , og hvor man , pyntet med Roserne , spiste for aabne Vinduer , mens Solen gik ned bag

af landlig Sundhed og døsende Længsel , og solgte deres Roser for erotisk høje Priser .

▫ deroppe i Alleen , hvor de mange Roser hænger ud over Plankeværket .

hendes Klædedragt en veg Varme , en Duft som af visnende Roser fromt udstrømme fra hendes Legeme ... noget , der rørte ham

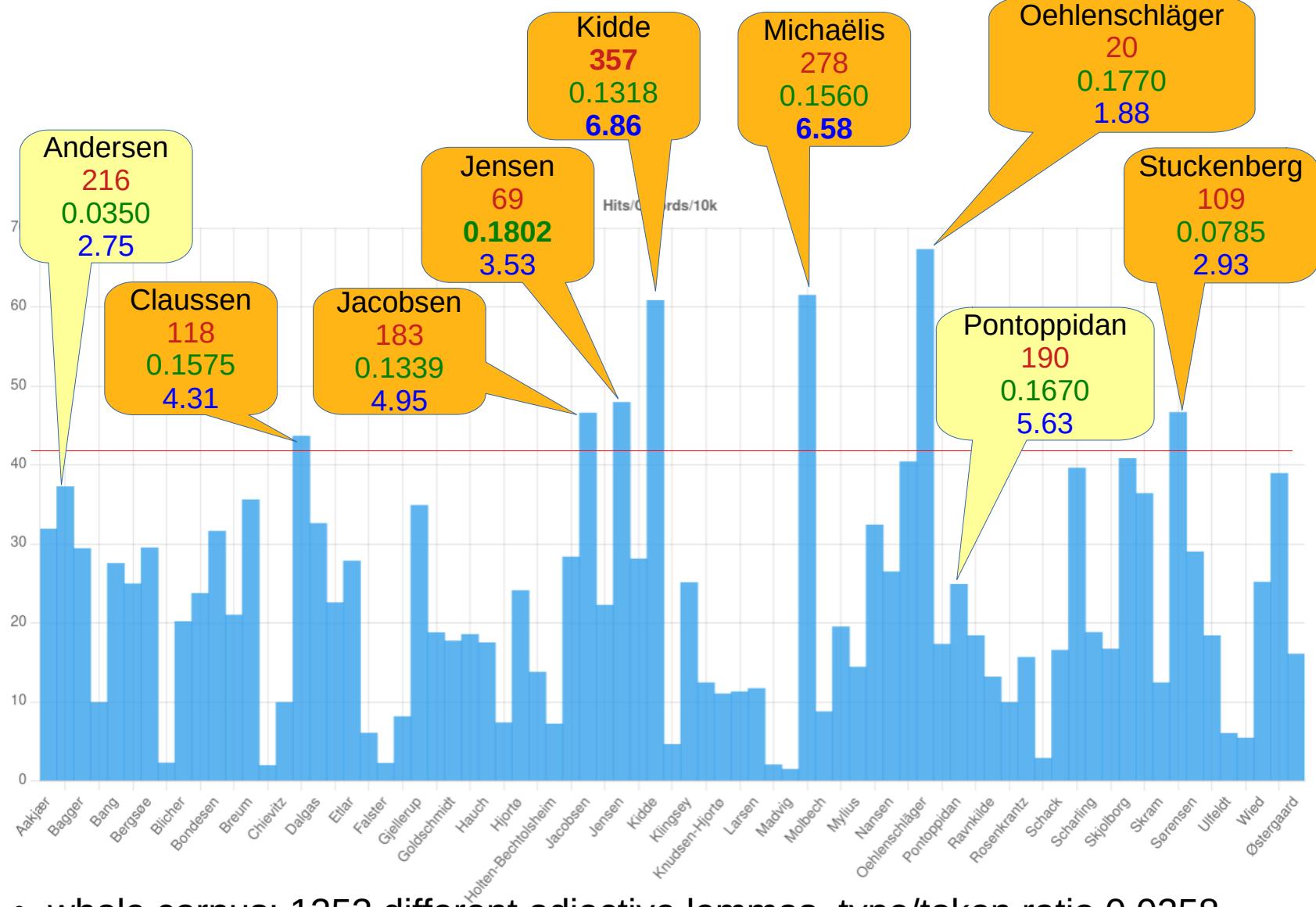
jeg saa mange foraarsklædte Skabninger med hede Øjne og Roser paa Brystet .

- often metaphorical usage: **kvinder = roser** (women = roses), “**rosen til erotisk høje priser**”
- **colour** (blood, cheeks, sunrise), **scent** (= body, fresh vs. decaying), **human conn.**: *ung*, *kys*

# use of colour adjectives:

frequency (columns), lemma types (red), TTR (green), Guiraud (blue)

|     |                 |
|-----|-----------------|
| 158 | bøgegrøn        |
| 159 | bølgeblå        |
| 160 | chokoladefarvet |
| 161 | citrongul       |
| 162 | dampgrå         |
| 163 | duegrå          |
| 164 | duggrå          |
| 165 | dunkelblå       |
| 166 | dunkelbrun      |
| 167 | dunkeltgrøn     |
| 168 | dødsblå         |
| 169 | dødsbrun        |
| 170 | dødsmørk        |
| 171 | dødssort        |
| 172 | efterårsbroget  |
| 173 | fakkelrød       |
| 174 | farvestrålende  |
| 175 | fedtplettet     |
| 176 | fjernblå        |
| 177 | flueplette      |
| 178 | fosforblå       |
| 179 | fregnat         |
| 180 | friskrød        |
| 181 | frostrød        |
| 182 | frostsort       |
| 183 | frådehvid       |
| 184 | fugtigt-sort    |
| 185 | fyrremørk       |
| 186 | fæstningsgrå    |
| 187 | gedebladsgrøn   |
| 188 | glasblå         |
| 189 | gråblå          |



- whole corpus: 1253 different adjective lemmas, type/token ratio 0.0258
- Pontoppidan: 190 types, TTR=0.1670, Guiraud: 5.63

# Part 5: Gender studies

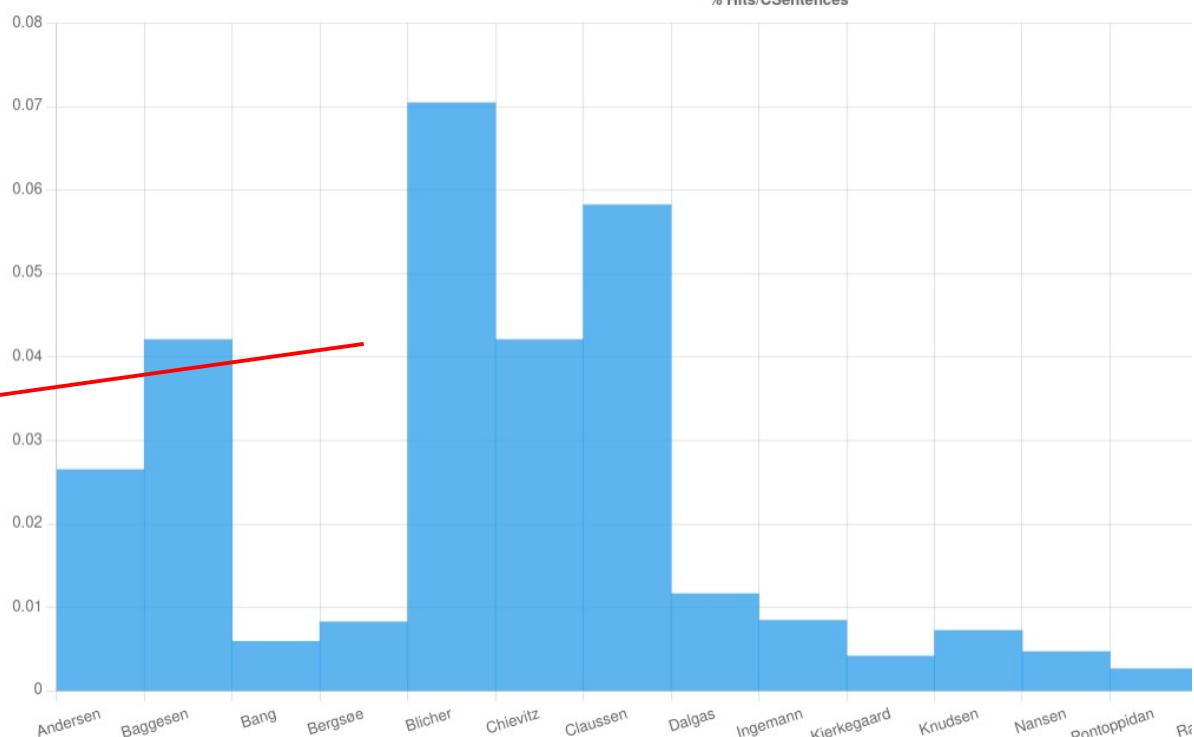
1. Corpora can identify and quantify gender stereotypes
2. Diachronic corpora can document changes over time
3. A literature corpus can provide a meta perspective on gender issues and a window into period-specific concepts of gender
4. Example: Profession & gender

|                    |      |
|--------------------|------|
| prinsesse          | 1291 |
| tjenestepige       | 386  |
| skuespillerinde    | 344  |
| generalinde        | 237  |
| husholderske       | 213  |
| lærerinde          | 196  |
| stuepige           | 151  |
| amme               | 148  |
| sangerinde         | 124  |
| kammerjomfru       | 121  |
| <b>kokkepige</b>   | 112  |
| guvernante         | 110  |
| kammerherrende     | 106  |
| herskerinde        | 95   |
| kunstnerinde       | 93   |
| husjomfru          | 77   |
| <b>hyrdinde</b>    | 77   |
| kammerpige         | 72   |
| sygeplejerske      | 64   |
| forfatterinde      | 62   |
| barnepige          | 61   |
| danserinde         | 59   |
| konsulinde         | 58   |
| ejerinde           | 52   |
| kejserinde         | 49   |
| admiralinde        | 47   |
| provstinde         | 47   |
| husbestyrerinde    | 45   |
| sypige             | 39   |
| tjenerinde         | 39   |
| oldfrue            | 38   |
| hofjægermesterinde | 35   |
| husmor             | 33   |
| præstefrue         | 33   |
| malkepige          | 31   |
| amtmandinde        | 28   |
| pastorinde         | 25   |
| borgfrue           | 24   |

# Gender: women working (lit. corpus)

search: <fem> & <Hprof>

|                   |    |
|-------------------|----|
| malkepige         | 31 |
| amtmandinde       | 28 |
| pastorinde        | 25 |
| borgfrue          | 24 |
| brudepige         | 23 |
| borgmesterinde    | 19 |
| digterinde        | 19 |
| hore              | 19 |
| priorinde         | 18 |
| hofmesterinde     | 17 |
| infantinde        | 17 |
| pianistinde       | 15 |
| syerske           | 14 |
| bestyrerinde      | 13 |
| jægermesterinde   | 13 |
| generalkonsulinde | 12 |
| herregåardsfrue   | 12 |
| plejerske         | 12 |
| vestalinde        | 12 |
| arbejderneske     | 11 |
| enkedronning      | 11 |
| operasangerinde   | 11 |
| spåkvinde         | 11 |



▫ Tescenen var endnu ikke endt , da **Kokkepigen** forsigtig traadte ind og meldte

▫ **Kokkepiger** og Købmænd og Mælkebønder !

1 , hvorledes det skal være , forbød han **Kokkepigen** paa det Strængeste at tørre Flasken af , den skulde

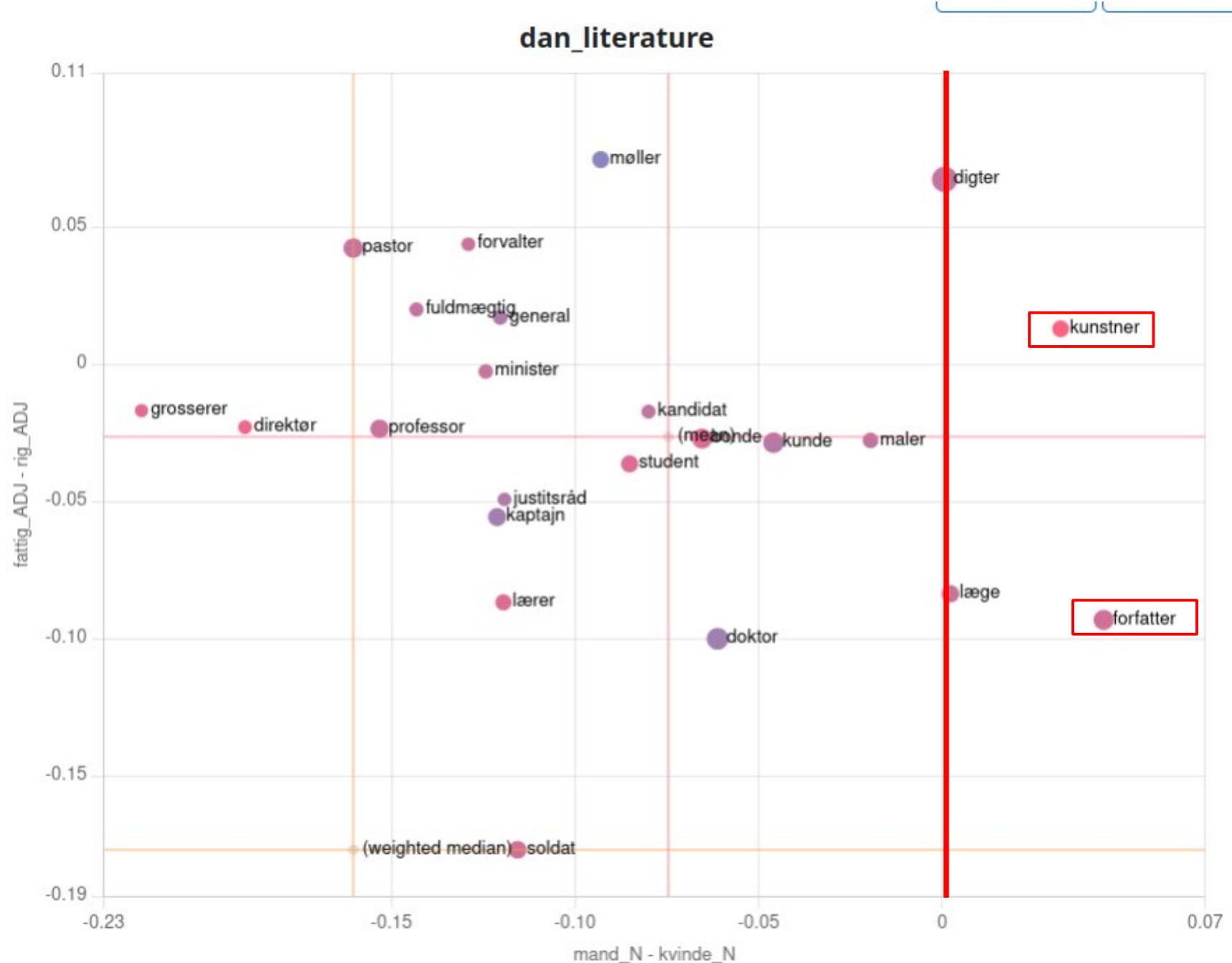
▫ Strax efter kom **Kokkepigen** med Melding fra en Sælgekone , hvorover der lige

3 , simpelthen at finde sig gift med sin **Kokkepige** ?

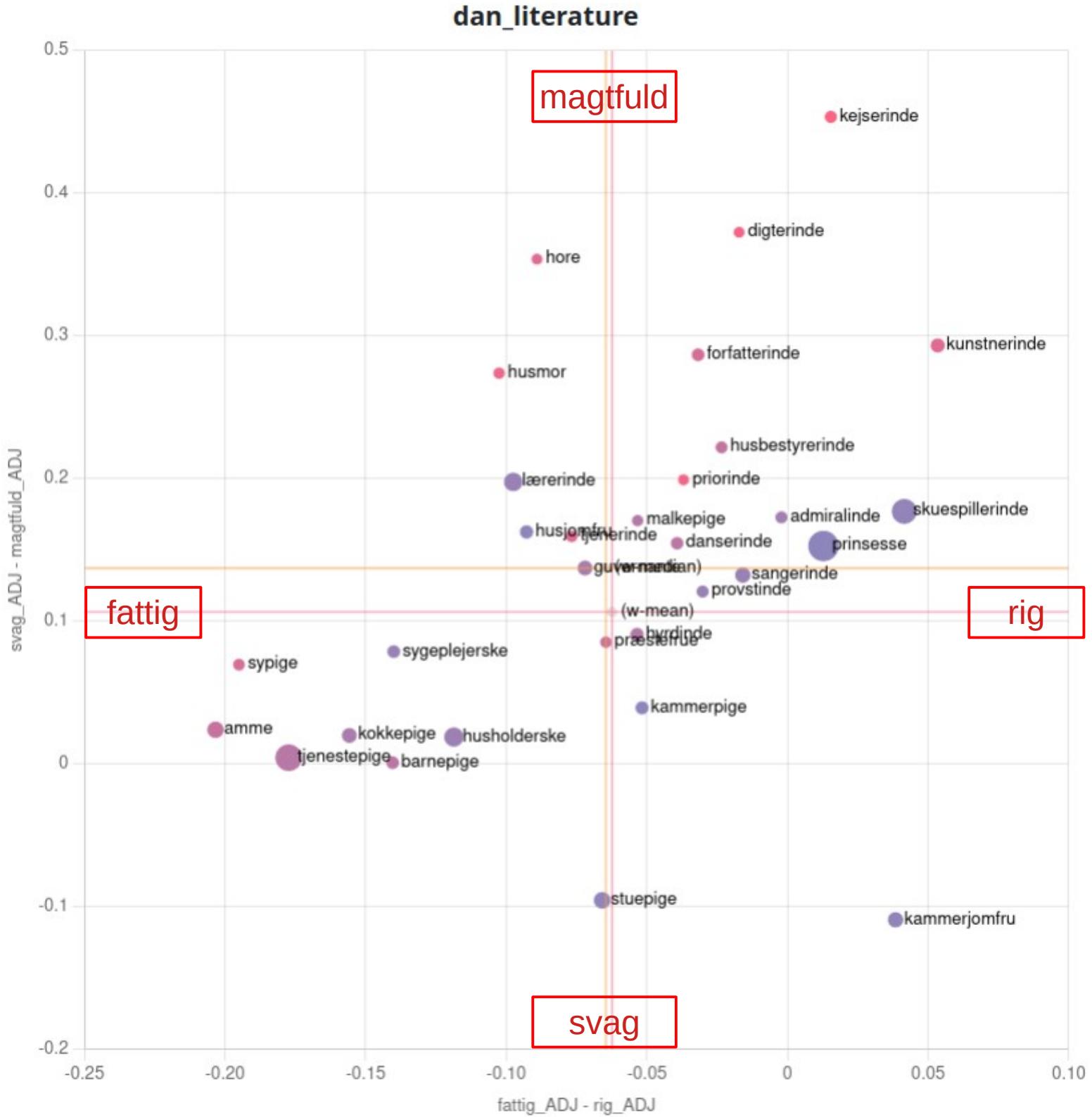
▫ Det var **Kokkepigen** « Røde\_Hanne » , som rasede over , at hun skulde

# Gender: women working (lit. corpus)

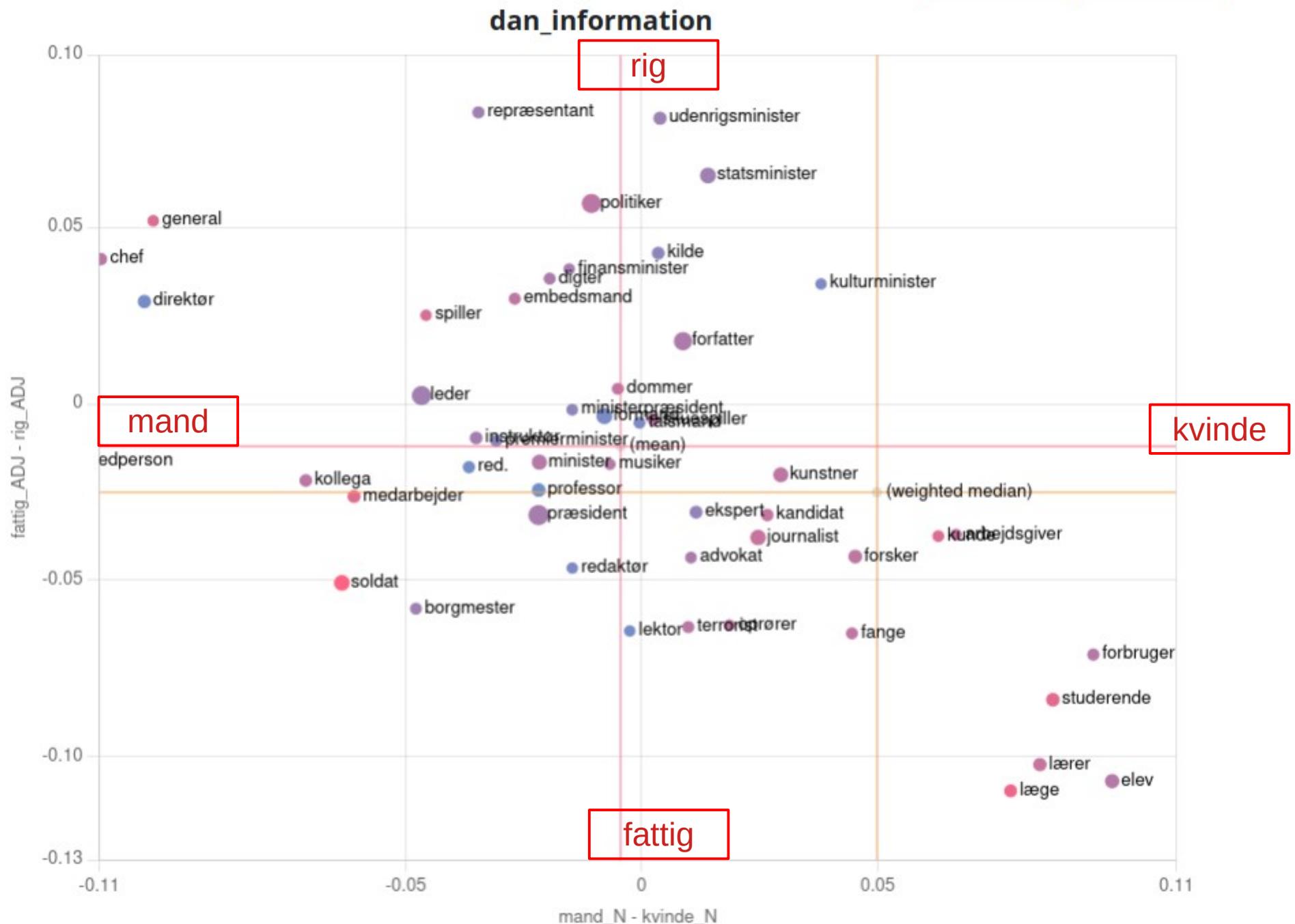
search: <Hprof> without <fem> [first 50, higher frequency than <fem>]



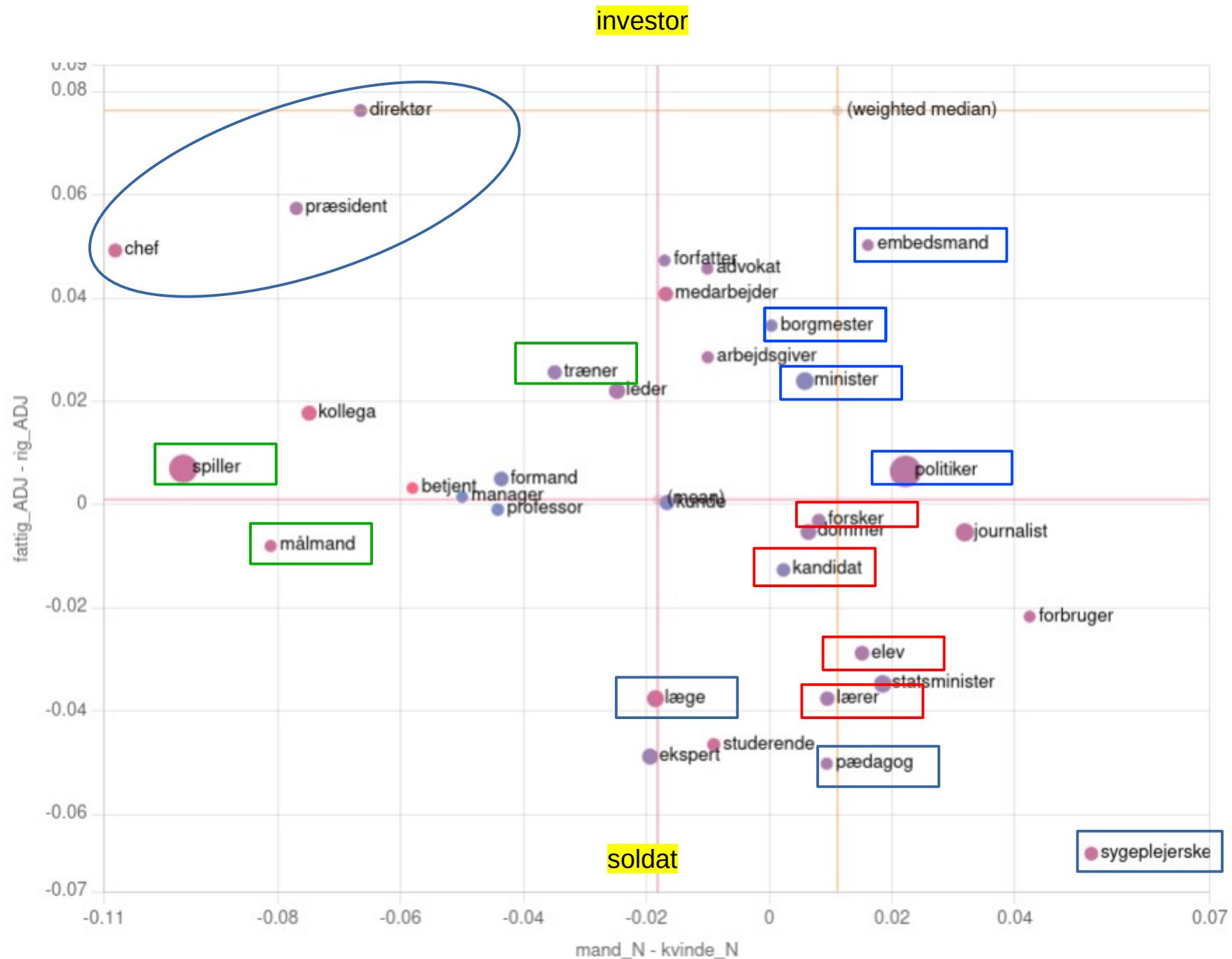
# Working women: Power & wealth



# Gender bias for professions: Information 1996-2008



# Gender bias for professions: Twitter 2017-2023



# linguistic clues to profession gender stereotypes: possessive pronouns

The figure displays three panels of a morphological analysis interface. Each panel has a header with buttons for selection (radio buttons 1, +, ?, \*), word input (Word), base (Base), extra (Extra), and semantic features (Sem). Below these are buttons for Part of Speech (-), Morphology (+), Function (+), and Semantic Role (+). At the bottom are Dep Head and Delete buttons.

- Panel 1:** Analyzes the word "lærer". The base is "lærer". The semantic feature Sem is checked. The part of speech is set to Noun (checked) and singular (checked).
- Panel 2:** Analyzes the word "og". The base is empty. The semantic feature Sem is checked. The part of speech is set to + (checked).
- Panel 3:** Analyzes the word "(hans|hendes)". The base is empty. The semantic feature Sem is checked. The part of speech is set to + (checked).

▫ twittername ▫ om , hvad en **lærer og hans** elever siger til det twittername

To 15-årige dømt : ▫ Truede med at slå **lærer og hendes** børn ihjel & utm\_medium\_twitter ▫

▫ En **lærer og hendes** 6 mdr gamle babyen skal et år og tre ▫

at du ikke har læst indlægget omkring **læreren og hans** undervisning i ytringsfrihed .....

ordan de må have været overfor en brun **lærer og hendes** brune klasse .

# linguistic clues to gender stereotypes: pronoun subjects

Diagram illustrating the analysis of the sentence "(når en) lærer SUBJ> har FS-ADVL> fri, har hun <SUBJ (fri)" using a dependency parser interface.

The interface shows three stages of analysis:

- Stage 1:** The word "lærer" is identified as the base form. The "Part of Speech - Noun" checkbox is checked. The "Function - Subject" checkbox is checked and highlighted with a green border. The "Free Adverbial (left)" checkbox is checked and highlighted with a green border.
- Stage 2:** The word "lærer" is identified as the base form. The "Function - Subject" checkbox is checked and highlighted with a red border. The "Full Subject (left)" checkbox is checked and highlighted with a red border.
- Stage 3:** The word "(han|hun)" is identified as the base form. The "Function - Subject" checkbox is checked and highlighted with a red border. The "Full Subject (right)" checkbox is checked and highlighted with a red border.

Buttons at the bottom of each stage include "Dep Head", "Sibling", and "Delete".

At the bottom, the sentence structure is shown as: "(når en) lærer SUBJ> har FS-ADVL> fri, har hun <SUBJ (fri)".

# Linguistic clues to profession gender stereotypes: pronoun subjects

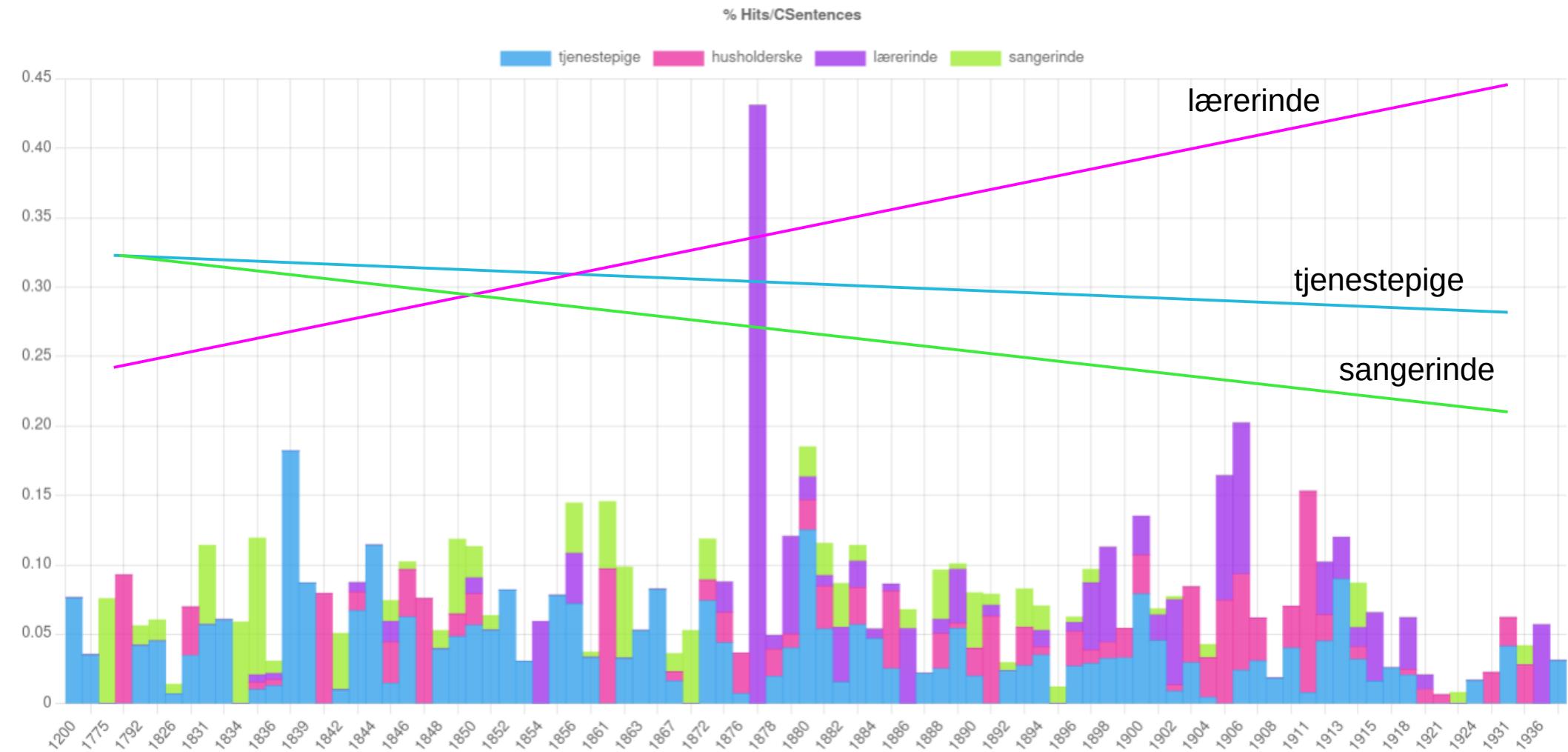
- twittername ▫ Og når læreren opdager det , så bliver hun pånt træt af moderen 😡 ▫ Tro mig !  
adder idræt . ▫ Og hvis læreren skal opfylde fælles mål , kan han med fordel bruge en mobiltelefon . ▫ Ikke i alle  
.emskolen . ▫ Hvis ikke læreren kan håndtere det , burde han aldrig have valgt den profession . ▫ Føj !  
erne i klassen . ▫ Når læreren ikke har tid , bliver han usikker på sig selv og sine omgivelser . ▫ Vi gør meget  
ername ▫ du at hvis en lærer kommer i Folketinget , og gerne vil blive ved m at undervise nogle gange , så ska hun  
ieden . ▫ Lige\_så\_snart læreren kiggede væk gjorde hun som det passede hende . ▫ 8 år gammel og megasej  

▫ Hvis bare læreren havde haft en pistol i skuffen , så havde hun kunnet forsvere sig mod den seksårige

  
ionsens med at « når en lærer har fri , så har hun fri » . ▫ Men det gør hun simpelthen . ▫ Husk lige det  
▫ Selv\_om den russiske lærer er okay , er han alligevel malplaceret , det er en skændsel at en skoleleder  
tar til krigen ? ▫ Hvis læreren er pro\_Putin , skal hun naturligvis ikke undervise ukrainske børn .  

▫ se , men når selvsamme lærer underviser i russisk historie og kultur så er hun så far off ufølsom og empati forlad

# women working: in-house or out-house



**Part 6:**

**Sustainability**

**&**

**Corporate image profiling**

# Quantitative corpus inspection: “økologi”

| Count | Words                                     |
|-------|---|
| 246   | <a href="#">økologisk landbrug</a>        |
| 129   | <a href="#">økologisk vare</a>            |
| 88    | <a href="#">økologisk produktion</a>      |
| 83    | <a href="#">økologisk landmand</a>        |
| 77    | <a href="#">økologisk produkt</a>         |
| 62    | <a href="#">økologisk fødevare</a>        |
| 61    | <a href="#">økologisk landsforening</a>   |
| 50    | <a href="#">økologisk mad</a>             |
| 43    | <a href="#">økologisk bomuld</a>          |
| 43    | <a href="#">økologisk kød</a>             |
| 39    | <a href="#">økologisk råd-1</a>           |
| 29    | <a href="#">økologisk gris</a>            |
| 29    | <a href="#">økologisk mælk</a>            |
| 28    | <a href="#">økologisk frugt</a>           |
| 26    | <a href="#">økologisk producent</a>       |
| 24    | <a href="#">økologisk drift</a>           |
| 21    | <a href="#">økologisk areal</a>           |
| 21    | <a href="#">økologisk grøntsag</a>        |
| 21    | <a href="#">økologisk svineproduktion</a> |
| 21    | <a href="#">økologisk æg-1</a>            |
| 20    | <a href="#">økologisk tilstand</a>        |
| 19    | <a href="#">økologisk kollaps</a>         |
| 19    | <a href="#">økologisk sengetøj</a>        |
| 18    | <a href="#">økologisk kylling</a>         |
| 17    | <a href="#">økologisk katastrofe</a>      |
| 17    | <a href="#">økologisk kø</a>              |
| 16    | <a href="#">økologisk æble</a>            |
| 14    | <a href="#">økologisk honning</a>         |
| 14    | <a href="#">økologisk mat-2</a>           |
| 14    | <a href="#">økologisk råvare</a>          |
| 13    | <a href="#">økologisk gård</a>            |
| 13    | <a href="#">økologisk juletræ</a>         |

n-grams:  
“økologisk” + N

relative frequency  
of correlates (N)  
(in-corpus)

(new) compounds  
<F:øko+.\*>

|                              |        |        |      |      |    |
|------------------------------|--------|--------|------|------|----|
| <a href="#">økoflubber</a>   | 178.25 | 125.45 | 5.83 | 0.3% | 30 |
| <a href="#">øko-landmand</a> | 169.20 | 121.27 | 5.64 | 0.3% | 29 |
| <a href="#">øko-landbrug</a> | 246.21 | 117.09 | 5.44 | 0.3% | 28 |
| <a href="#">øko-mælk</a>     | 82.46  | 108.73 | 5.06 | 0.2% | 26 |
| <a href="#">økomærke</a>     | 56.70  | 104.54 | 4.86 | 0.2% | 25 |
| <a href="#">økohøne</a>      | 368.01 | 92.00  | 4.28 | 0.2% | 22 |
| <a href="#">økokød</a>       | 222.20 | 92.00  | 4.28 | 0.2% | 22 |
| <a href="#">økotosse</a>     | 351.28 | 87.82  | 4.08 | 0.2% | 21 |
| <a href="#">øko-dag</a>      | 334.59 | 83.63  | 3.89 | 0.2% | 20 |
| <a href="#">økofascisme</a>  | 97.38  | 79.45  | 3.69 | 0.2% | 19 |
| <a href="#">øko-salg</a>     | 136.67 | 75.27  | 3.50 | 0.2% | 18 |
| <a href="#">øko-øl</a>       | 188.01 | 75.27  | 3.50 | 0.2% | 18 |
| <a href="#">økoko</a>        | 284.40 | 71.09  | 3.31 | 0.2% | 17 |
| <a href="#">økosvin</a>      | 97.07  | 71.09  | 3.31 | 0.2% | 17 |
| <a href="#">øko-kampagne</a> | 184.44 | 66.91  | 3.11 | 0.1% | 16 |
| <a href="#">øko-mad</a>      | 125.77 | 66.91  | 3.11 | 0.1% | 16 |
| <a href="#">økobonde</a>     | 184.44 | 66.91  | 3.11 | 0.1% | 16 |
| <a href="#">økokylling</a>   | 267.67 | 66.91  | 3.11 | 0.1% | 16 |
| <a href="#">økoko</a>        | 250.94 | 62.73  | 2.92 | 0.1% | 15 |
| <a href="#">økokollaps</a>   | 250.94 | 62.73  | 2.92 | 0.1% | 15 |
| <a href="#">økosamfund</a>   | 59.35  | 62.73  | 2.92 | 0.1% | 15 |

dan\_twitter

1 to 51 of 1615 ( $\Sigma$  6577)

Download TSV

| Token               | G: freq <sup>2</sup> /norm | C: freq <sup>2</sup> /norm | C: freq/corp · 10 <sup>8</sup> | freq/conc | num |
|---------------------|----------------------------|----------------------------|--------------------------------|-----------|-----|
| spisemærke          | 372.20                     | 178.84                     | 3.69                           | 0.3%      | 19  |
| mejeriproducent     | 563.57                     | 174.78                     | 4.08                           | 0.3%      | 21  |
| bomuld-2            | 245.38                     | 125.09                     | 10.11                          | 0.8%      | 52  |
| polykrise           | 327.29                     | 81.81                      | 3.31                           | 0.3%      | 17  |
| erhvervsteam        | 321.06                     | 80.25                      | 1.75                           | 0.1%      | 9   |
| landbrug            | 67.89                      | 79.03                      | 100.14                         | 7.8%      | 515 |
| luksus-ide          | 237.82                     | 59.45                      | 0.97                           | 0.1%      | 5   |
| babyTskjorte        | 198.19                     | 49.54                      | 0.97                           | 0.1%      | 5   |
| CBD-pasta           | 190.26                     | 47.56                      | 0.78                           | 0.1%      | 4   |
| mejerisheik         | 190.26                     | 47.56                      | 0.78                           | 0.1%      | 4   |
| gårdkylling         | 152.21                     | 38.05                      | 0.78                           | 0.1%      | 4   |
| rajsvingel          | 152.21                     | 38.05                      | 0.78                           | 0.1%      | 4   |
| cannellini-bønne    | 142.69                     | 35.67                      | 0.58                           | 0.0%      | 3   |
| landsforening       | 7.33                       | 28.93                      | 6.03                           | 0.5%      | 31  |
| æg-1                | 20.62                      | 28.45                      | 32.67                          | 2.6%      | 168 |
| landbrugsfællesskab | 7.56                       | 26.75                      | 0.58                           | 0.0%      | 3   |
| mælk                | 22.80                      | 24.34                      | 30.14                          | 2.4%      | 155 |
| bulgursegment       | 0.00                       | 23.78                      | 0.39                           | 0.0%      | 2   |

# øko-: inspection

Export all ↗

- 🔗 i 🌐 , der lovede medvind på cykelstierne ... men den hopper økoflipperne jo altid op 😊
  - 🔗 i 🌐 Det er ikke kun økoflipperne , der dropper gift - almindelige haveejere vælger det også
  - 🔗 i 🌐 Haveekspert : Det er ikke kun økoflipperne , der dropper gift - almindelige haveejere vælger det også
- 🔗 i 🌐 lad os blive ved med at vaccinere danske børn - gælder også økoflipperne ! =&dkpol =&A6slinger ? & utm\_campaign\_buffer
  - 🔗 i 🌐 twittername går GonHo hele vejen og stemmer økoflipperne .
  - 🔗 i 🌐 Kan man være økoflipper , samtidig med man har tatoveret et facebook like på
  - 🔗 i 🌐 twittername De ræverøde økoflipperne sku tvinges til at cykle< ; < . De skal ikke
  - 🔗 i 🌐 . Og at spedningen kunne være undgået hvis ikke økoflipperne havde fået forbudt afbrænding om vinteren der ellers ville
  - 🔗 i 🌐 . Det er vist kun Greenpeace der tror på den . Og så økoflipperne .
  - 🔗 i 🌐 twittername Som forsøgende værende politisk korrekt økoflipper der dog altid sætter smag over etik - øko muscavado sukker
  - 🔗 i 🌐 lige i skabet . Det dyrefri landbrug er nytænkning . Økoflipperne , klimatosser og fuglekiggere , læs ! Den danske
  - 🔗 i 🌐 Så er der lagt i kakkelovnen til alle økoflipperne , miljøtosserne , klimaklovnene og plantekostprofeterne .
  - 🔗 i 🌐 synes folk skal fratages billige flyrejser . Egoistiske økoflipperne .
  - 🔗 i 🌐 twittername består det sindssyge ? Altså udover at økoflipperne og klimatosserne ikke spise dansk ? 😊
  - 🔗 i 🌐 sig ansvaret og skælder dig ud for at være kommunist , økoflipper eller røgfifianatiker 😎
  - 🔗 i 🌐 twittername en økoflipper !
  - 🔗 i 🌐 politisk uvidende studerende , cyklende storbymennesker , økoflipperne der tror at DK kan redde verdenen , generelt naive , der
  - 🔗 i 🌐 twittername er et af verdens rigeste lande , og så påstår økoflipperne , at vi ikke har råd til de skidedyre øko varer . Sludder
  - 🔗 i 🌐 Fedt , du skal ikke tro du som økoflipper er bedre end andre & twittername

# Identifying mining companies in the corpus

[lex="mineradora" & pos="N"] [pos="PROP"]

The screenshot shows a search interface with various filters and a results table.

**Field:** Baseform / lemma

**By:** Right edge

**Offset:** 0

**Group By:** publisher

**Sort:** Freq

**por\_blogs**  
1 to 51 of 73 ( $\Sigma 329$ )

**Download TSV**

| Token            | G: freq <sup>2</sup> /norm | C: freq <sup>2</sup> /norm | C: freq/corp * 10 <sup>8</sup> | freq/conc | num | Compare?                            |
|------------------|----------------------------|----------------------------|--------------------------------|-----------|-----|-------------------------------------|
| Vale             | 66538.48                   | 56411.55                   | 5.92                           | 32.2%     | 106 | <input checked="" type="checkbox"/> |
| Samarco          | 1375749.19                 | 179434.62                  | 5.09                           | 27.7%     | 91  | <input checked="" type="checkbox"/> |
| MMX              | 23576.88                   | 19991.62                   | 0.95                           | 5.2%      | 17  | <input checked="" type="checkbox"/> |
| Anglo American   | 42552.30                   | 16339.49                   | 0.73                           | 4.0%      | 13  | <input checked="" type="checkbox"/> |
| Kinross          | 142604.60                  | 18599.47                   | 0.34                           | 1.8%      | 6   | <input checked="" type="checkbox"/> |
| Belo Sun         | 39002.97                   | 5087.03                    | 0.22                           | 1.2%      | 4   | <input checked="" type="checkbox"/> |
| SAMA             | 34968.18                   | 4560.79                    | 0.22                           | 1.2%      | 4   | <input checked="" type="checkbox"/> |
| Yamana Gold      | 59651.60                   | 7780.17                    | 0.22                           | 1.2%      | 4   | <input checked="" type="checkbox"/> |
| Alcoa            | 227.12                     | 607.33                     | 0.17                           | 0.9%      | 3   | <input checked="" type="checkbox"/> |
| BHP Billiton     | 8775.67                    | 1144.58                    | 0.17                           | 0.9%      | 3   | <input checked="" type="checkbox"/> |
| Paranapanema     | 57.40                      | 347.65                     | 0.17                           | 0.9%      | 3   | <input type="checkbox"/>            |
| Vale S.A.        | 0.00                       | 0.00                       | 0.17                           | 0.9%      | 3   | <input type="checkbox"/>            |
| Vale do Rio Doce | 455.89                     | 609.82                     | 0.17                           | 0.9%      | 3   | <input type="checkbox"/>            |

# Profiling through associated concepts: Samarco

| por_blogs  |                            |  |                                |           |     |                                     | <a href="#">Download TSV</a> |
|------------|----------------------------|--|--------------------------------|-----------|-----|-------------------------------------|------------------------------|
| Token      | G: freq <sup>2</sup> /norm | C: freq <sup>2</sup> /norm                                       | C: freq/corp · 10 <sup>8</sup> | freq/conc | num | Compare?                            |                              |
| barragem   | 107.23                     | 483.42   | 11.73                          | 4.3%      | 210 | <input checked="" type="checkbox"/> |                              |
| empresa    | 1.01                       | 2.50   | 7.60                           | 2.8%      | 136 | <input checked="" type="checkbox"/> |                              |
| mineradora | 693.70                     | 425.50   | 6.99                           | 2.5%      | 125 | <input checked="" type="checkbox"/> |                              |
| rompimento | 125.05                     | 117.43   | 6.99                           | 2.5%      | 125 | <input checked="" type="checkbox"/> |                              |
| rejeito    | 659.37                     | 356.68   | 4.53                           | 1.6%      | 81  | <input checked="" type="checkbox"/> |                              |
| lama       | 21.07                      | 20.14  | 4.14                           | 1.5%      | 74  | <input checked="" type="checkbox"/> |                              |
| dia        | 0.11                       | 0.09   | 3.58                           | 1.3%      | 64  | <input checked="" type="checkbox"/> |                              |
| desastre   | 7.9                        | dam, mining, break, wast, mud, disaster, tragedy, government ... |                                |           |     |                                     |                              |
| mG         | 5.9                        |  |                                |           |     |                                     |                              |
| vale       | 1.8                        |  |                                |           |     |                                     |                              |
| mineração  | 25.0                       |  |                                |           |     |                                     |                              |
| tragédia   | 2.72                       | 3.21   | 2.40                           | 0.9%      | 43  | <input type="checkbox"/>            |                              |
| ação       | 0.34                       | 0.21   | 2.24                           | 0.8%      | 40  | <input type="checkbox"/>            |                              |
| governo    | 0.07                       | 0.32   | 2.24                           | 0.8%      | 40  | <input type="checkbox"/>            |                              |
| rio        | 0.48                       | 0.62   | 2.12                           | 0.8%      | 38  | <input type="checkbox"/>            |                              |
| novembro   | 0.45                       | 0.73   | 2.07                           | 0.7%      | 37  | <input type="checkbox"/>            |                              |
| fundão     | 4.72                       | 50.73  | 2.01                           | 0.7%      | 36  | <input type="checkbox"/>            |                              |
| água       | 0.20                       | 0.15   | 2.01                           | 0.7%      | 36  | <input type="checkbox"/>            |                              |

|            |      |      |      |      |    |                          |
|------------|------|------|------|------|----|--------------------------|
| projeto    | 0.10 | 0.06 | 1.23 | 0.4% | 22 | <input type="checkbox"/> |
| região     | 0.05 | 0.10 | 1.23 | 0.4% | 22 | <input type="checkbox"/> |
| crime      | 0.11 | 0.21 | 1.17 | 0.4% | 21 | <input type="checkbox"/> |
| área       | 0.04 | 0.07 | 1.17 | 0.4% | 21 | <input type="checkbox"/> |
| risco      | 0.09 | 0.14 | 1.12 | 0.4% | 20 | <input type="checkbox"/> |
| acordo     | 0.05 | 0.35 | 1.06 | 0.4% | 19 | <input type="checkbox"/> |
| obra       | 0.04 | 0.06 | 1.06 | 0.4% | 19 | <input type="checkbox"/> |
| vida       | 0.02 | 0.01 |      |      |    |                          |
| valor      | 0.04 | 0.04 |      |      |    |                          |
| bilhão     | 0.10 | 0.28 |      |      |    |                          |
| comunidade | 0.07 | 0.08 | 0.95 | 0.3% | 17 | <input type="checkbox"/> |
| impacto    | 0.22 | 0.29 | 0.95 | 0.3% | 17 | <input type="checkbox"/> |
| morador    | 0.19 | 0.26 | 0.95 | 0.3% | 17 | <input type="checkbox"/> |
| parte      | 0.02 | 0.02 | 0.89 | 0.3% | 16 | <input type="checkbox"/> |
| órgão      | 0.09 | 0.15 | 0.89 | 0.3% | 16 | <input type="checkbox"/> |
| bacia      | 0.72 | 1.00 | 0.84 | 0.3% | 15 | <input type="checkbox"/> |
| multa      | 0.24 | 0.45 | 0.84 | 0.3% | 15 | <input type="checkbox"/> |
| população  | 0.03 | 0.07 | 0.84 | 0.3% | 15 | <input type="checkbox"/> |
| processo   | 0.02 | 0.03 | 0.84 | 0.3% | 15 | <input type="checkbox"/> |

Search:  
[lex="Samarco" & pos="PROP"]

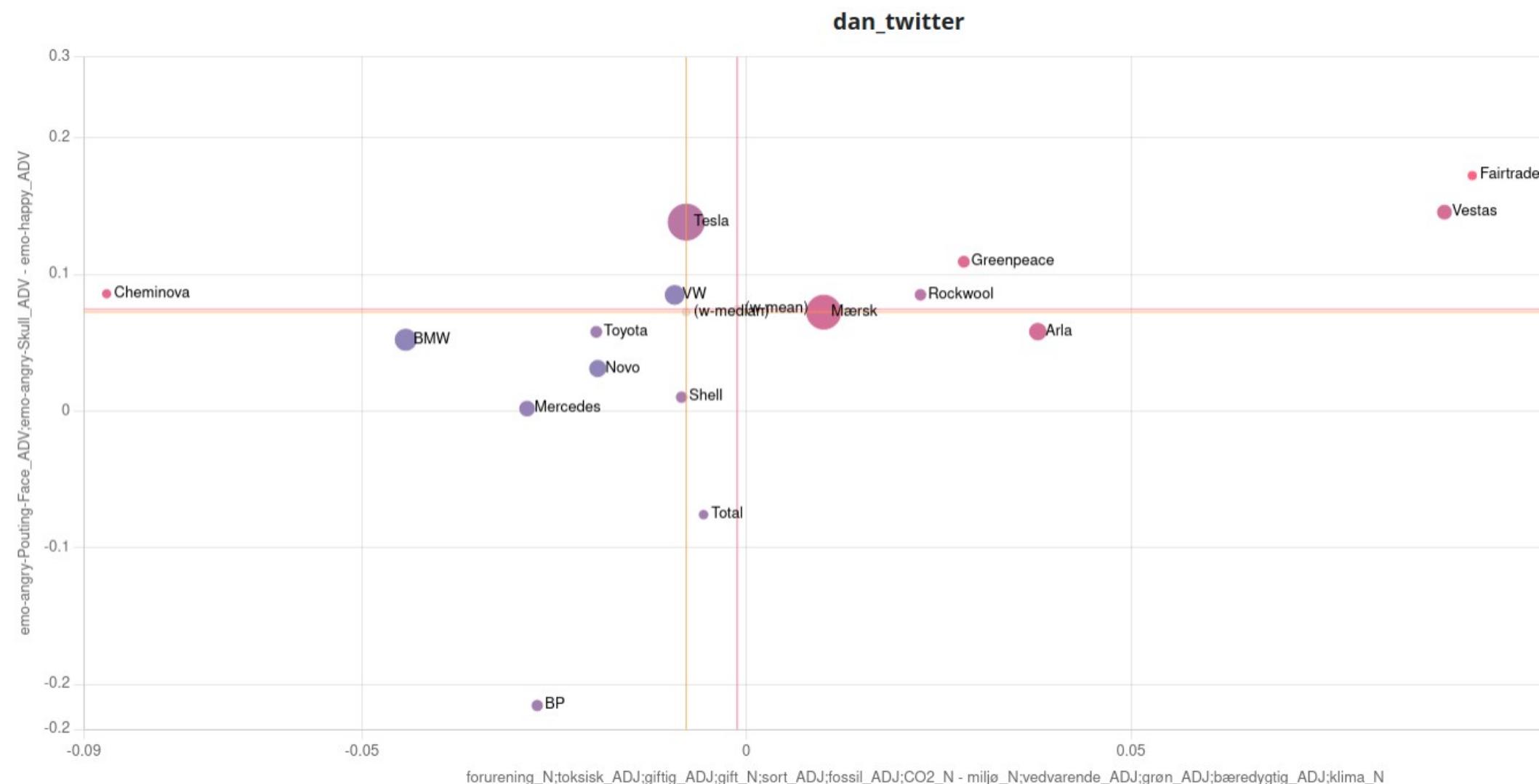
SQ: [pos="N"]

project, region, crime, risk, agreement, life, billions, impact, residents, fine, population

# Company image ranking with user-defined vector axes

- x-axis
  - ◆ toxic (-) vs. macrobiotic (+)
  - ◆ i.e. defined by individual word vectors
- y-axis
  - ◆ sad/angry/skeptical emojis (-) vs. happy/love emojis (+)
  - ◆ here: defined by vector bundles (averaged)
- scatter dot for company name CN computed as (x,y) with coordinates
  - ◆  $x = \text{SIM}(\textit{bæredygtig}, \text{CN}) - \text{SIM}(\textit{forurening}, \text{CN})$
  - ◆  $y = \text{SIM}(\textit{emo-happy.*}, \text{CN}) - \text{SIM}(\textit{emo-anxious.*}; \textit{emo-sad.*}, \text{CN})$

# Green / sustainability image: DK, cars & oil



x: **MINUS:** forurening\_N;toksisk\_ADJ;giftig\_ADJ;gift\_N;sort\_ADJ;fossil\_ADJ;CO2\_N  
**PLUS:** miljø\_N;vedvarende\_ADJ;grøn\_ADJ;bæredygtig\_ADJ;klima\_N

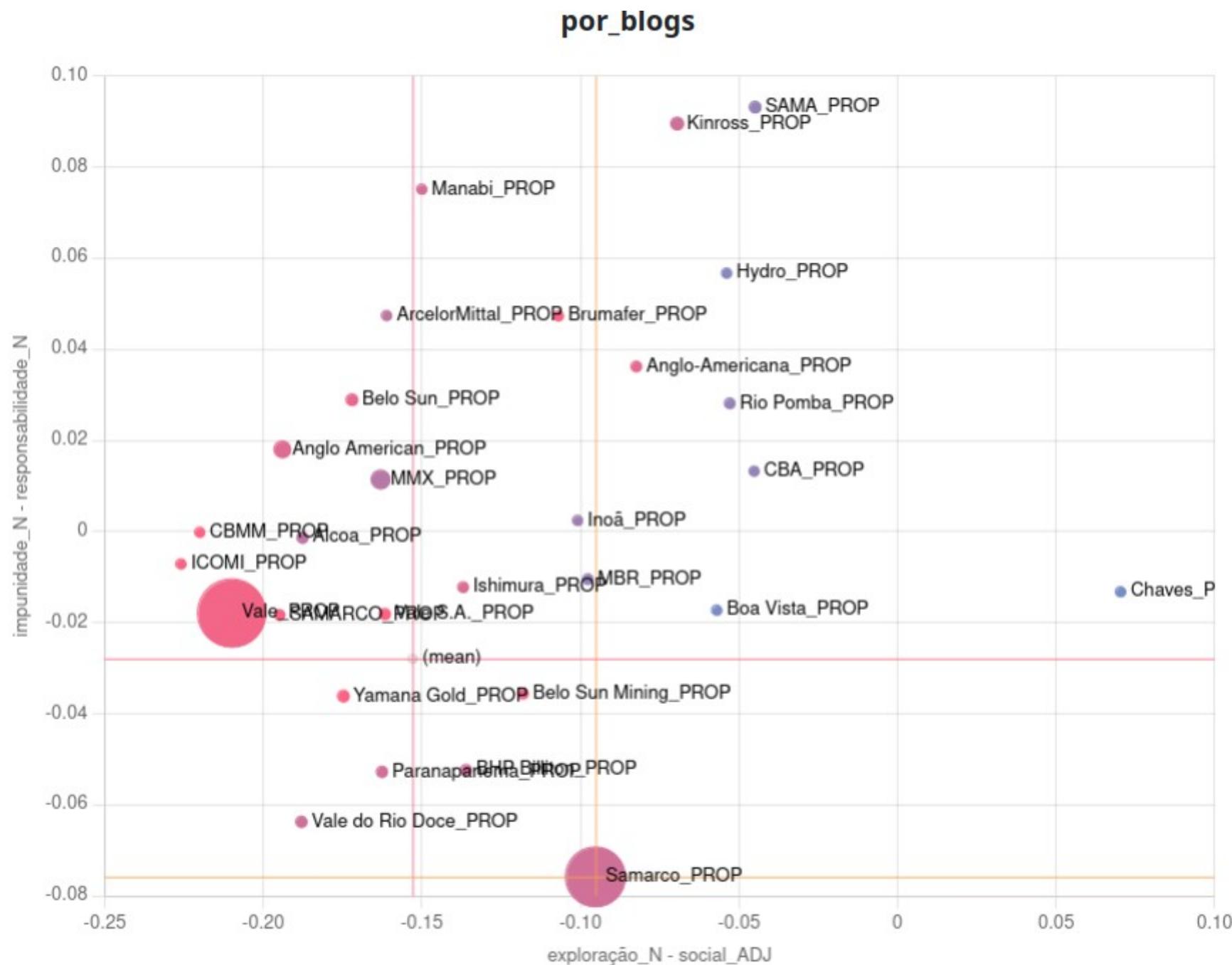
y: **MINUS:** emo-angry-Pouting-Face\_ADV --- **PLUS:** emo-happy\_ADV

# Sustainability image: Magnificent 7 / Tech – Oil companies

dan\_twitter

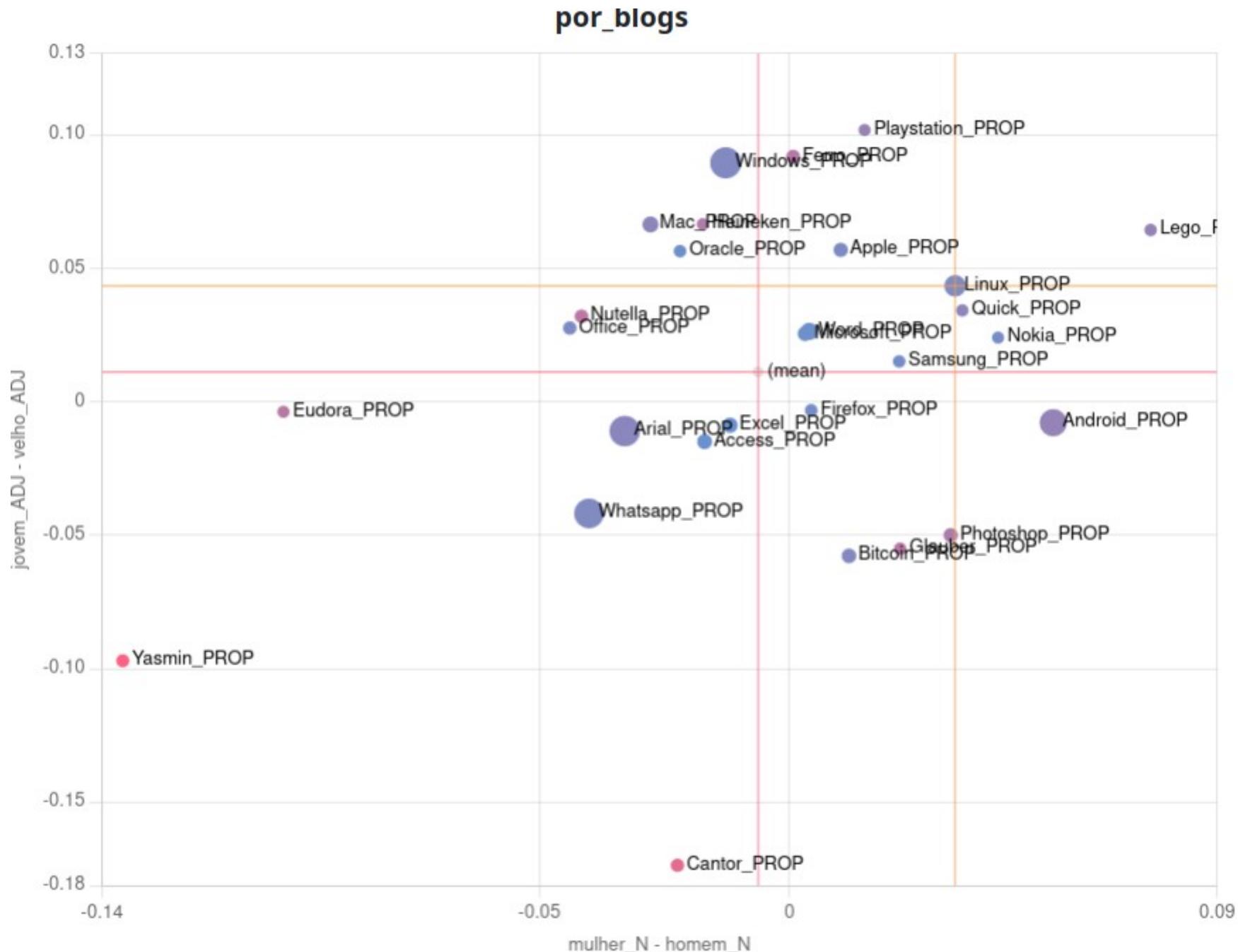


# Mining companies: Impunity vs. Accountability



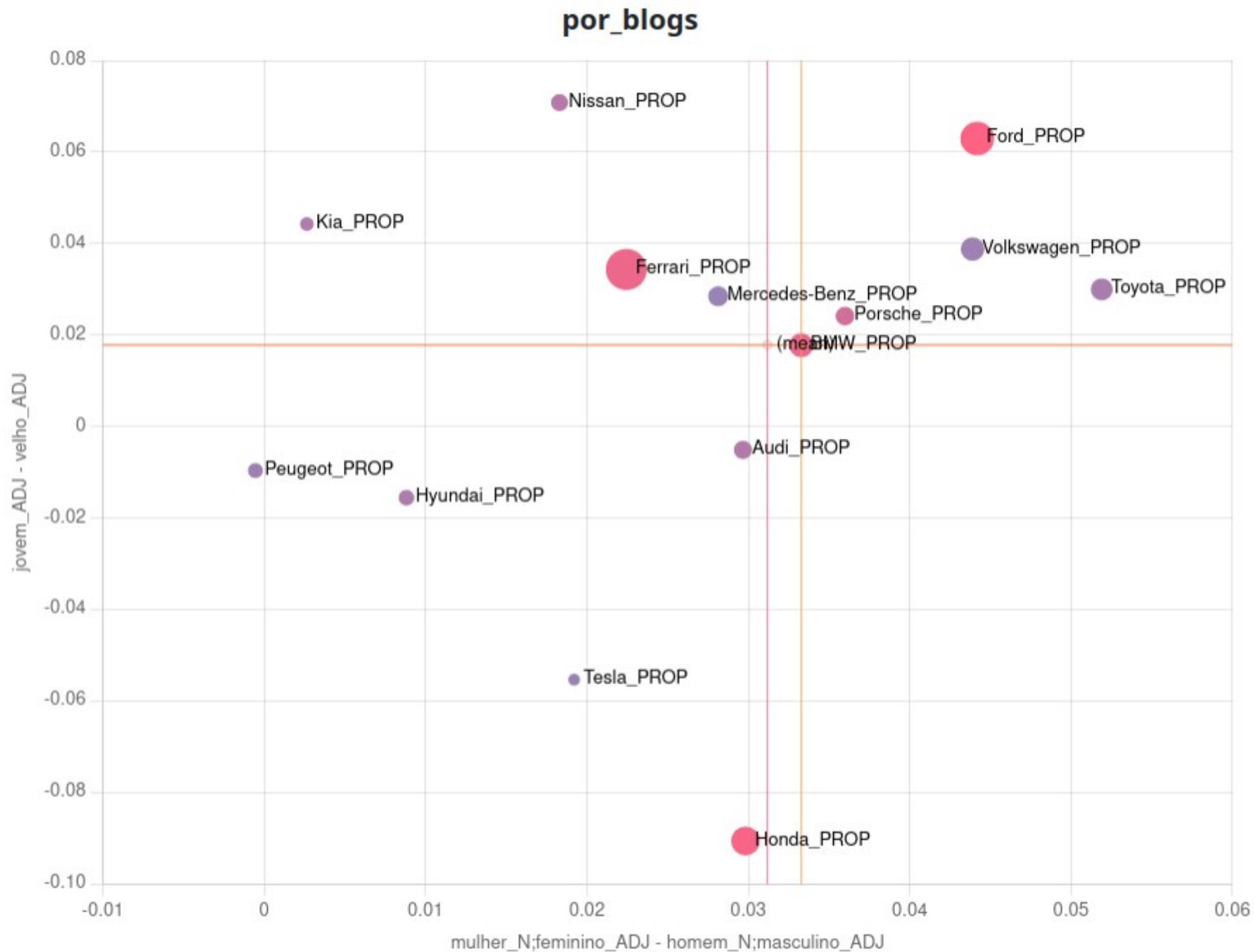
# Part 7: Social brand profiling (gender & age)

## <brand> PROP



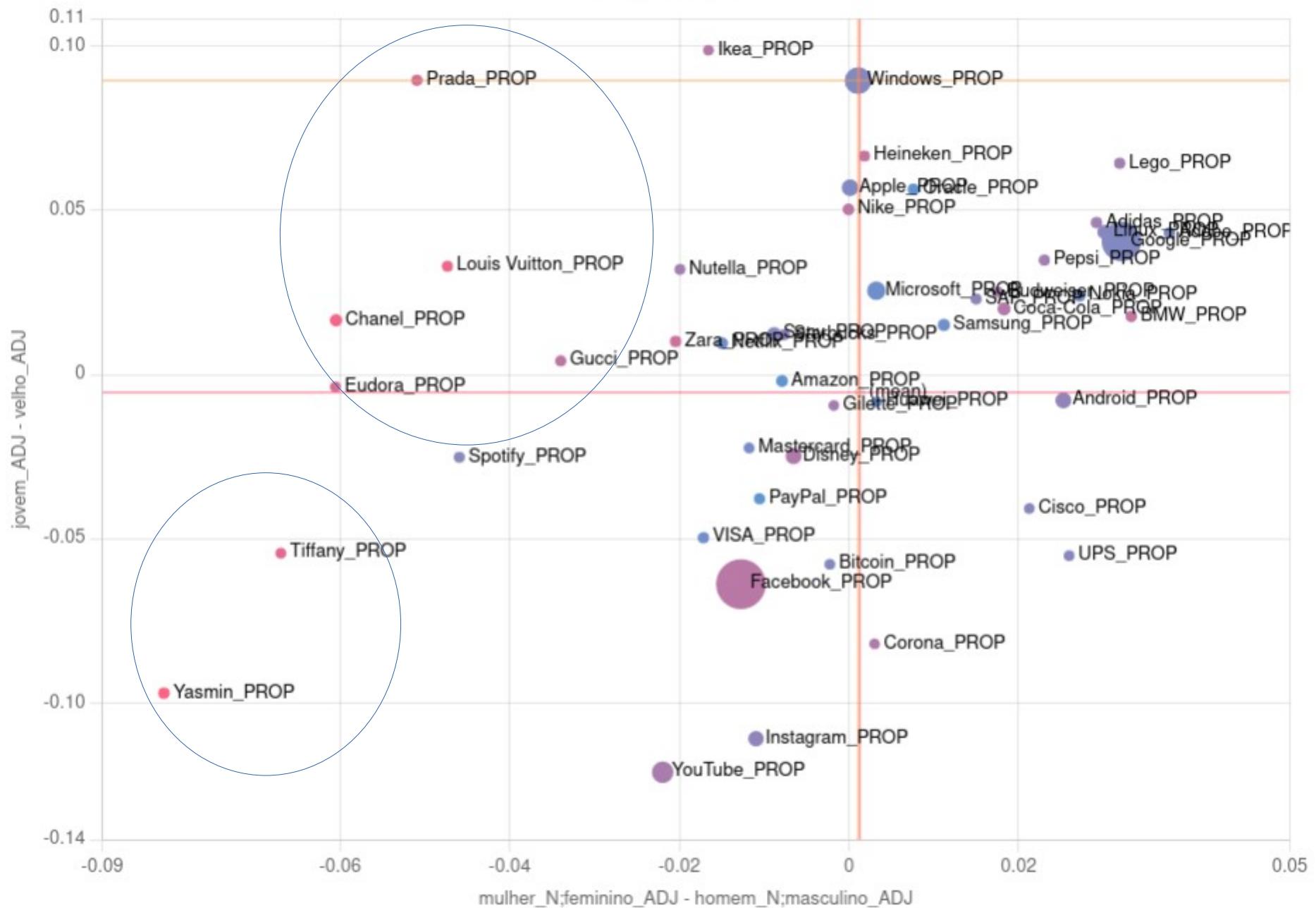
# Brand image: Cars

## gender & age



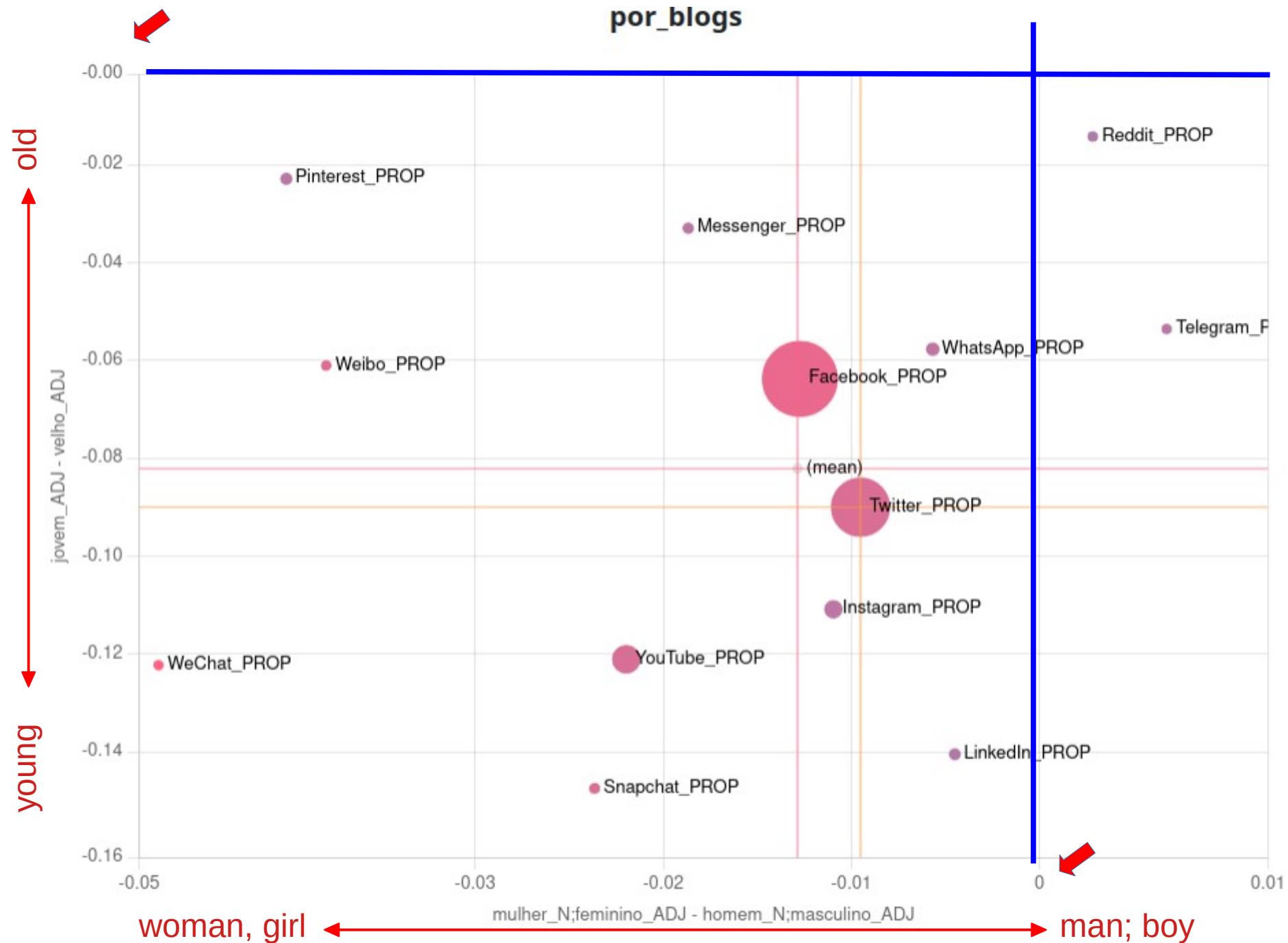
# Brand image: Mixed gender & age

por\_blogs



# Brand image: Social Media

## gender & age



# Conclusion

1. **Large corpora** can be used not only for linguistic purposes, but also for a large variety of humanistic research goals relating to literary, cultural or social topics
2. **Linguistic annotation** (and a search interface exploiting it) make it possible to identify the nature and usage of
  - the word clusters, notions and concepts involved
  - possible stereotypes related to the area
  - the association strength between concepts and perceived attributes
3. **Qualitative inspection** of large text collections is only possible in a meaningful way by first running a **quantitative analysis**: statistics, co-occurrences, visualisations ...
4. **Outlook**
  - prepare **dedicated subcorpora** for certain topics or periods
  - establish a “monitor corpus” harvesting **CMC data every day**



parsers: **edu.visl.dk** (pt,en,de,da,es,fr,it,eo,sv,no,nl)

corpus search: **corp.visl.dk** (Corpus Eye)

email: **eckhard.bick@gmail.com**

Bick, Eckhard (2023). Derogatory Linguistic Mechanisms in Danish Online Hate Speech. In: Isabel Ermida (Ed.): *Hate Speech in Social Media: A Linguistic Approach*. London: Palgrave Macmillan

Bick, Eckhard (2023). VISL & CG-3: Constraint Grammar on the Move: An application-driven paradigm. In: Arvi Hurskainen, Kimmo Koskenniemi & Tommi Pirinen (eds.), *Rule-Based Language Technology*. NEALT Monograph Series vol. 2, pp. 112-140. University of Tartu. ISSN 1736-6291

Bick, Eckhard (2022). PFN-PT: A Framenet Annotator for Portuguese. In: Heliana Mello & Fernanda Farinelli (eds.): *The computational treatment of Brazilian Portuguese*. Domínios de Linguagem. [S. I.], v. 16, n. 4, pp. 1401-1435. ISSN 1980-5799

Bick, Eckhard (2020). An Annotated Social Media Corpus for German. In: Proceedings of the 12th International Conference on Language Resources and Evaluation, LREC2020 (Marseille, May 2020), 6129-6137. ACL / ELRA.

Bick, Eckhard (2020): “Annotating Emoticons and Emojis in a German-Danish Social Media Corpus for Hate Speech Research”. In: RASK – International Journal of Language and Communication. Vol. 52. University of Southern Denmark, pp. 1-20.

Bick, Eckhard (2014). PALAVRAS, a Constraint Grammar-based Parsing System for Portuguese. In: Tony Berber Sardinha & Thelma de Lurdes São Bento Ferreira (eds.), *Working with Portuguese Corpora*, pp 279-302. London/New York: Bloomsbury Academic.