Gepimpte data! From English TV show to Dutch verb in less than three years

Stefano De Pascale, Dirk Pijpops, Freek Van de Velde & Eline Zenner

Lilith RU, University of Liège

QLVL RU, University of Leuven

Overview

1. *Pimp my ride*: background and previous studies

2. Onomasiological competition: *pimpen* vs. *opleuken*

3. Morphological productivity of *pimpen*

4. Conclusions

Pimp my ride: background and previous studies

- MTV show launched in 2004

- rich, multi-ethnic youth from Los Angeles

- cosmetic makeover of a shabby car provided to the show by a participant

- US cultural ideologies surrounding pimps and ghetto-style to a mass audience

< 2000 (in English):
1. *pimp* (noun): 'person who controls prostitutes'
2. *pimp* (verb): 'to act as a pimp' and 'to prostitute someone'

> 2000 (English & Dutch):
3. *pimp(en)* (verb): 'to fancify'

Pimp my ride: background and previous studies

- **Van de Velde & Zenner (2010, 2015)**
  - — introduction via international TV show → role of media?
  - — taboo meaning, social stigma of the referents
  - — unstraightforward denominal verbalization
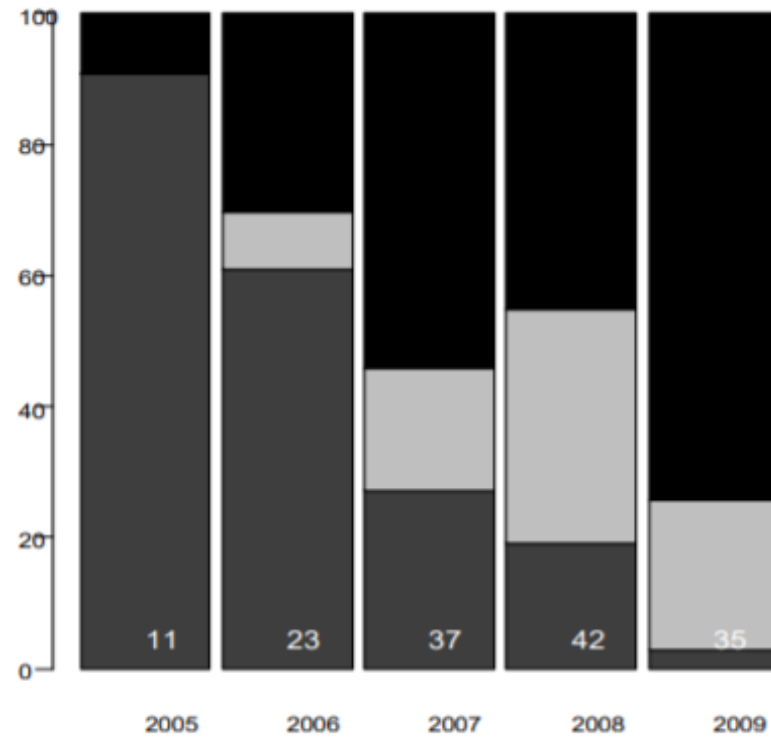 - — proper name → road to common use?

  case of constructional change!
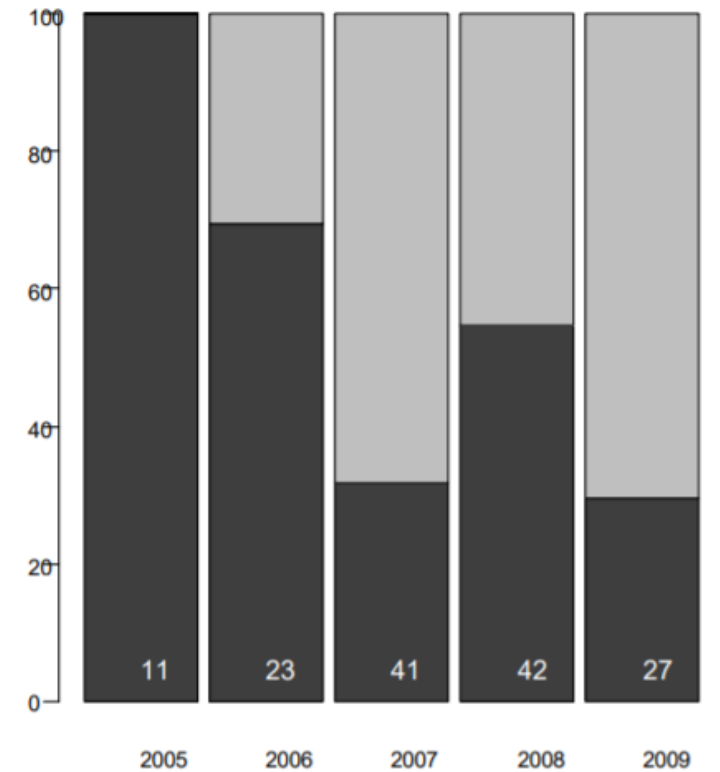 from borrowed English proper name to Dutch transitive verb

Pimp my ride: background and previous studies
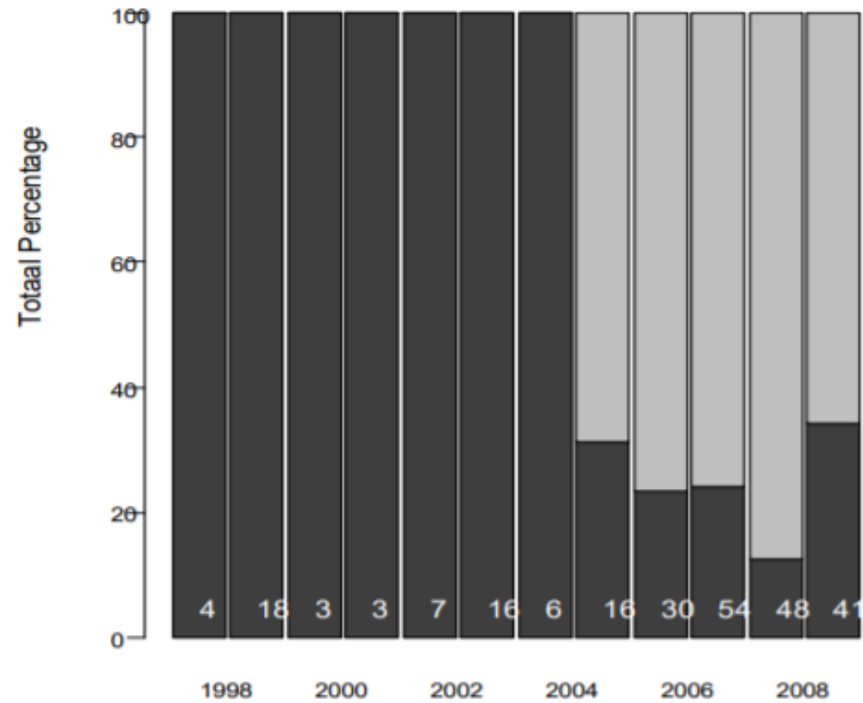
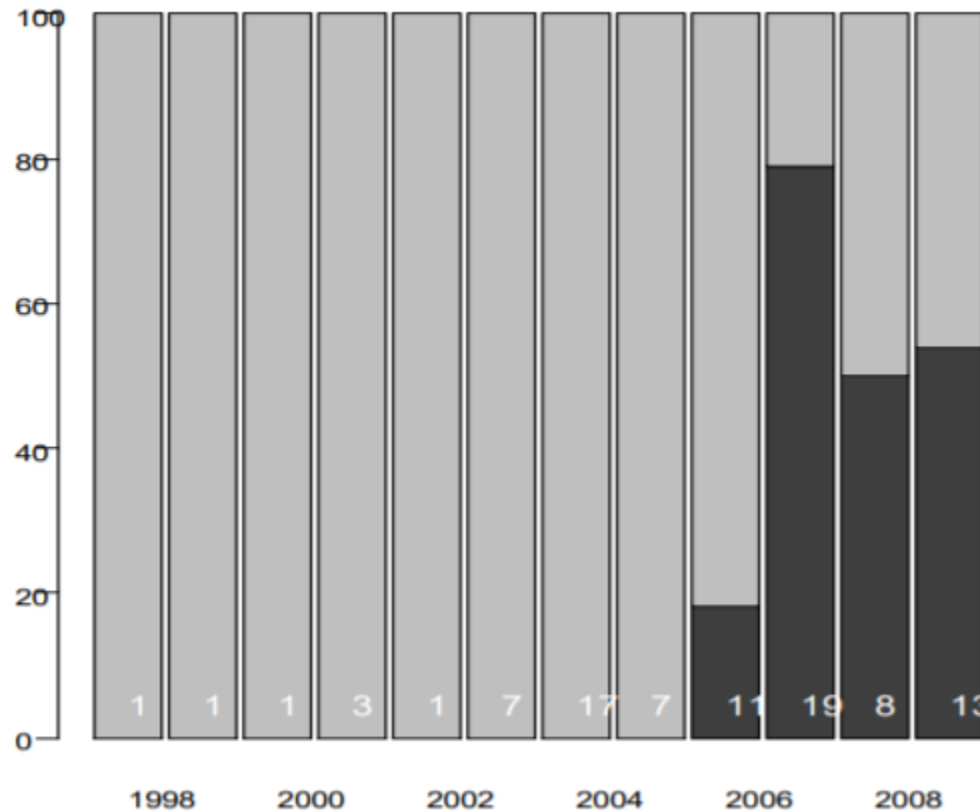sharp increase of uses of *pimp(en)* in the meaning 'to fancify' (light grey)

semantically broader range of pimped objects (light grey: other vehicles; **black: other objects**)

relaxing the construction *pimp POSS N* through open slots (light grey: other POSS, other N)

Pimp my ride: background and previous studies

onomasiological competition between *pimpen* (dark grey) and *opleuken* (light grey)

drawbacks

- small dataset (246 occurrences)

- suboptimal data source (newspapers)

Pimp my ride: background and previous studies

- De Pascale, Pijpops, Van de Velde & Zenner (under review)

  — data source: Twitter; time span: 2007 – 2020;

  — 4,561 manually annotated occurrences

  — new: deconstructionalization score (aggregate over semantic/formal diagnostics [Petré & Van de Velde 2019])

  — results

    ▪ within the *pimp POSS N* template

    ▪ with the use of the Dutch *pimpen*

    no further increase in deconstructionalization, only broader range of pimped objects

Pimp my ride: background and previous studies

---

today, we go big data

- 246 → 4,561 → 220,367 occurrences

  — i.e. all tweets ever tweeted in Dutch in which some form of *pimp** occurred

- pick up the thread left in previous studies

  1. onomasiological competition: *pimpen* vs. *opleuken*

  2. morphological productivity: *pimpen* and its derivations

Onomasiological competition: *pimpen* vs. *opleuken*

Hypotheses

- Sex: Women typically lead a language change

  → Women *pimp*, men *leuken op*.

- Country: English influences both Belgian and Netherlandic Dutch, but *leuk* is a Netherlandic adjective

  → Belgians *pimp*, Dutchmen *leuken op*.

- Year: *pimp* is young

  → We used to *opleuken*, now we *pimp*

- Tweet length: longer tweets tend to be more formal

  → *pimp* short, in longer sentences however we would rather use *opleuken*

- Activity tweeter: *pimp* is hip en cool

  → Kabouter Kwebbel *pimps*, Kabouter Lui *leukt op*

Onomasiological competition: *pimpen* vs. *opleuken*

Sex

- Based on username

- Cleaned

- Database from naamkunde.net

- If no exact match: Levenshtein distance, unless distance relative to the length of the name is too long, or the difference between the closest male name and the closest female name is too small

- 96% accuracy (98% precision 'female', 96% precision 'male')

Onomasiological competition: *pimpen* vs. *opleuken*

Country

- User location (added by tweeter)

- Cleaned

- Database from CBS and Statbel, manually extended (*Netherlands, Belgique, Vlaanderen, Achterhoek,…*)

- If no exact match: match longest substring of at least 6 characters (e.g. *aalsmeer netherlandsholland*)

- If no match: place country, place name (place where tweet was tweeted)

- 98% accuracy (77% precision 'Belgium', 100% precision 'Netherlands')

Onomasiological competition: *pimpen* vs. *opleuken*

- Year: centred

- Tweetlength: logarithm base 10

  (Palliera et al. 2011: 2524)

- Activity tweeter: root base 10

  — Number of tweets

  — Number of followers

  — Number of people the tweeter follows

- Regression model based on

  — only uninflected past participles:

    24,242 occurrences

  — of which only the occurrences with an

    attributed sex and country label :

    4532 *pimpen* vs. 1201 *opleuken*

Onomasiological competition: *pimpen* vs. *opleuken*

- Sex: Women typically lead a language change → Women *pimp*, men *leuken op.*

Coefficient = 0.23 (Reference level: women)

P-value = 0.0010

Confirmed
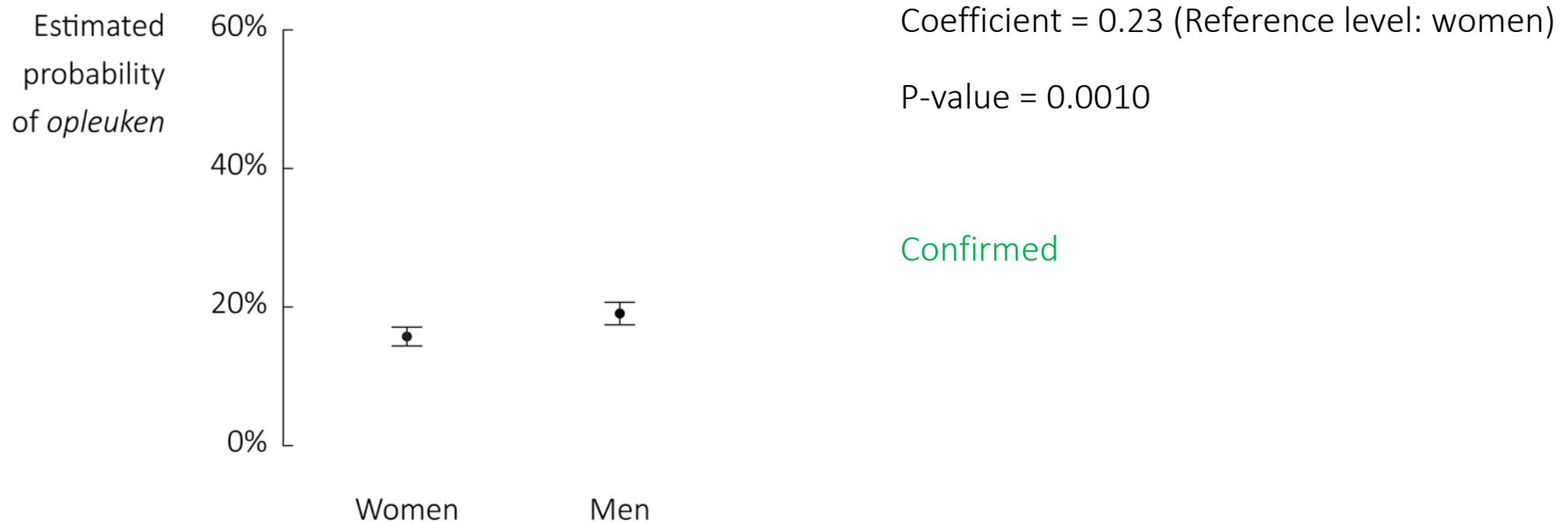
Onomasiological competition: *pimpen* vs. *opleuken*

- Country: English influences both Belgian and Netherlandic Dutch, but *leuk* is a Netherlandic adjective

  → Belgians *pimp*, Dutchmen *leuken op*.
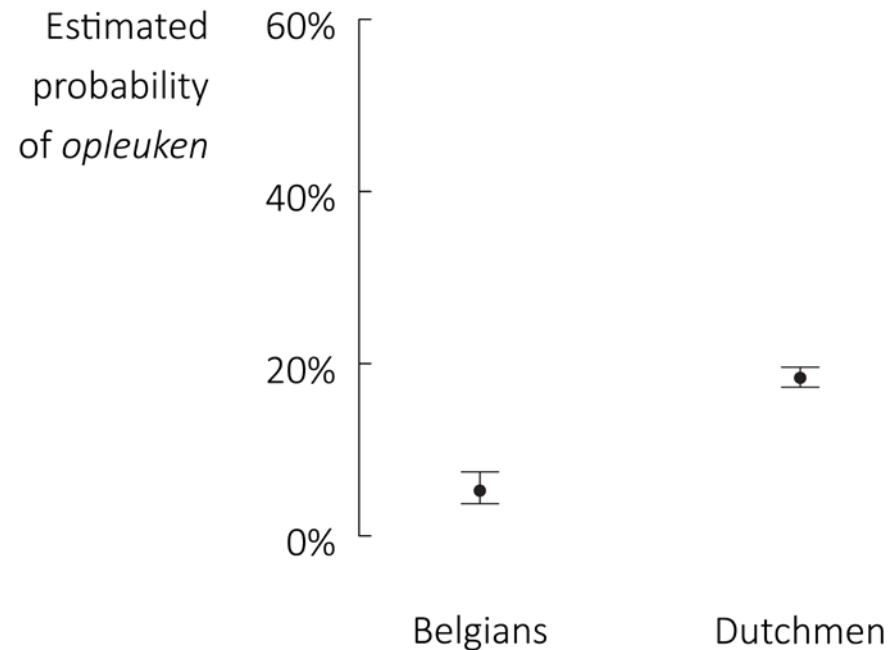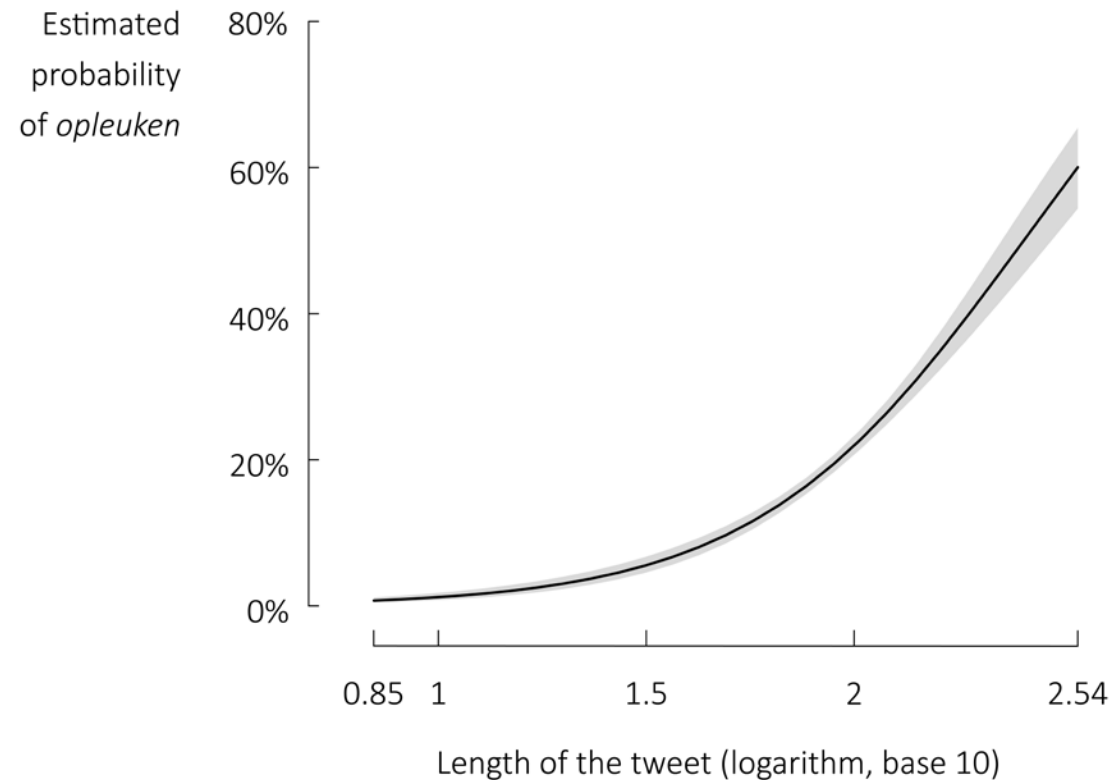

Coefficient = 1.40 (Reference level: Belgians)

P-value < 0.0001

Confirmed

Onomasiological competition: *pimpen* vs. *opleuken*

- Tweet length: longer tweets tend to be more formal

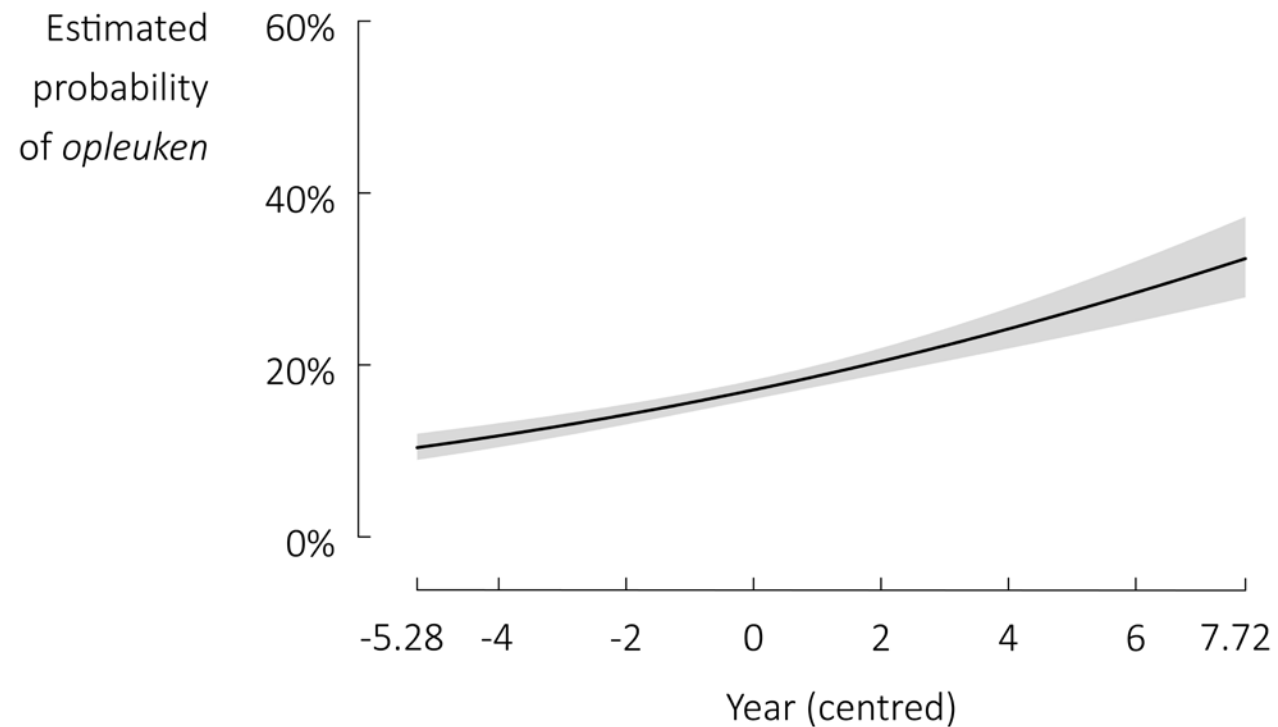 → *pimp* short, in longer sentences however we would rather use *opleuken*

Coefficient = 3.12

P-value < 0.0001

Confirmed

Onomasiological competition: *pimpen* vs. *opleuken*

- Year: Pimp is young → We used to *opleuken*, now we *pimp*
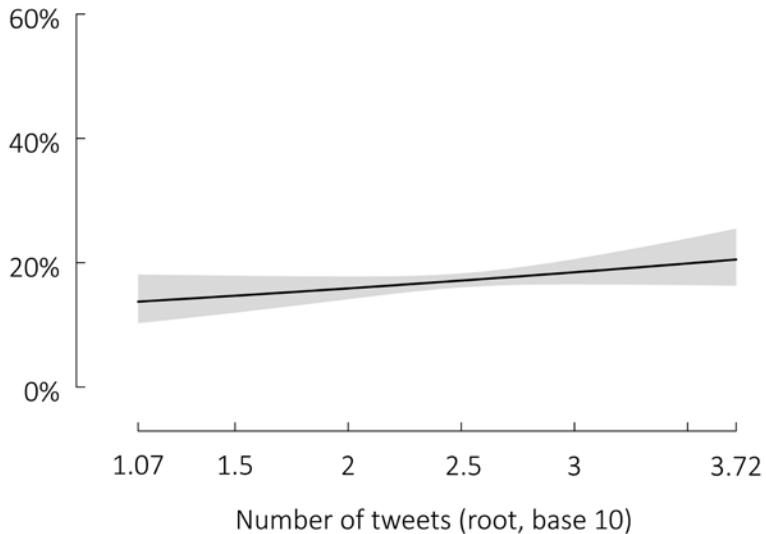

Coefficient = 0.11

P-value < 0.0001

Not confirmed

Onomasiological competition: *pimpen* vs. *opleuken*

- Activity tweeter: *pimp* is hip en cool
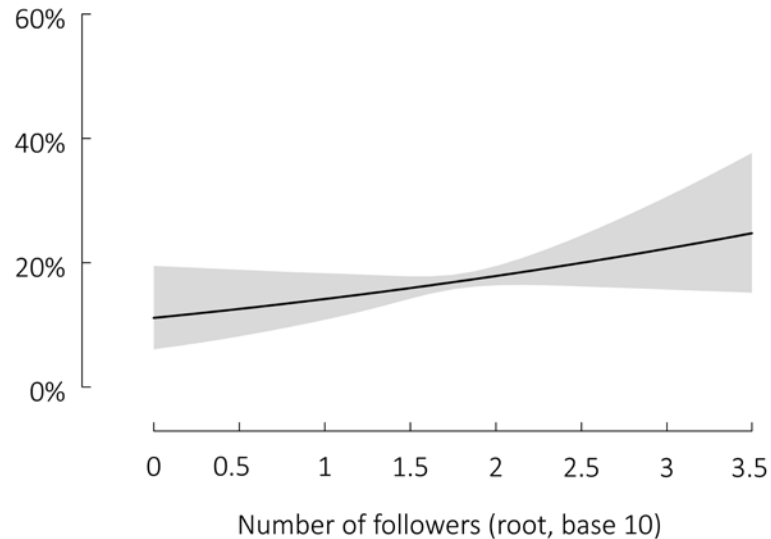
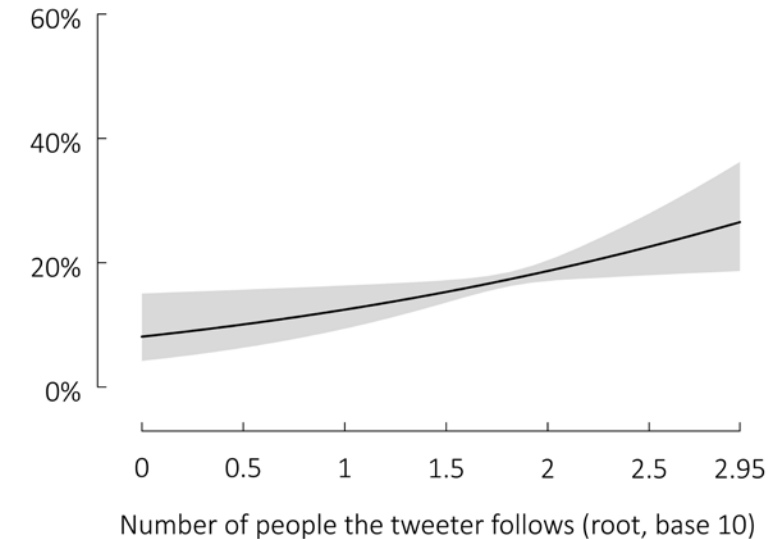 → Kabouter Kwebbel *pimps*, Kabouter Lui *leukt op*   Not confirmed

Coefficient = 0.18; p-value = 0.1081   Coefficient = 0.28; p-value = 0.1323   Coefficient = 0.48; p-value = 0.015

Onomasiological competition: *pimpen* vs. *opleuken*

Hypotheses

- Sex: Women typically lead a language change

    → Women *pimp*, men *leuken op*. Confirmed

- Country: English influences both Belgian and Netherlandic Dutch, but *leuk* is a Netherlandic adjective

    → Belgians *pimp*, Dutchmen *leuken op*. Confirmed

- Tweet length: longer tweets tend to be more formal

    → *pimp* short, in longer sentences however we would rather use *opleuken*. Confirmed

- Year: Pimp is young

    → We used to *opleuken*, now we *pimp*. Not Confirmed

- Activity tweeter: pimpen is hip en cool

    → Kabouter Kwebbel *pimps*, Kabouter Lui *leukt op*. Not confirmed
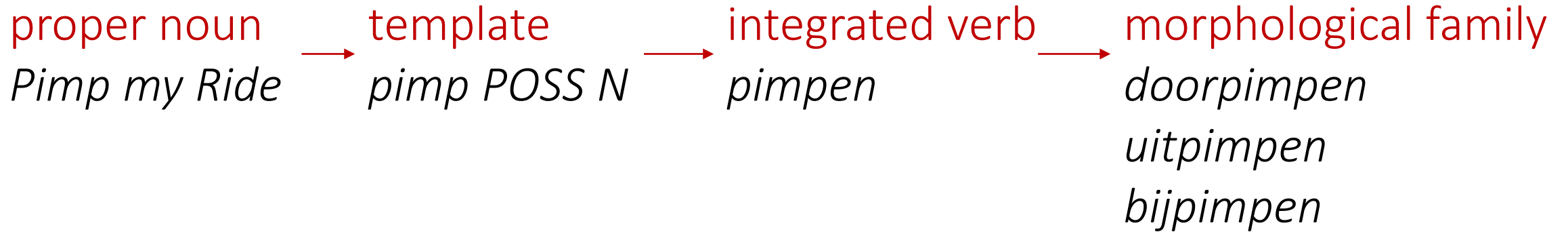
Onomasiological competition: *pimpen* vs. *opleuken*

Interpretation

- Year: *Pimp* is young

    → We used to *opleuken*, now we *pimp*. Not Confirmed

    → A rising tide lifts all boats?

- Activity tweeter: *pimp* is hip en cool

    → Kabouter Kwebbel *pimps*, Kabouter Lui leukt op. Not confirmed

   → Official accounts of news agencies, firms, institutions, etc. seem to be more active than regular tweeters

Morphological productivity: *pimpen* and its derivations

---

proper noun $\longrightarrow$ template $\longrightarrow$ integrated verb $\longrightarrow$ morphological family

Pimp my Ride   *pimp POSS N*   *pimpen*   *doorpimpen*
                                                uitpimpen
                                                bijpimpen
                                                ...

- morphological families and big data

    — enough data for the long tail of infrequent observations

    — how to calculate morphological diversity and productivity?

Morphological productivity: *pimpen* and its derivations
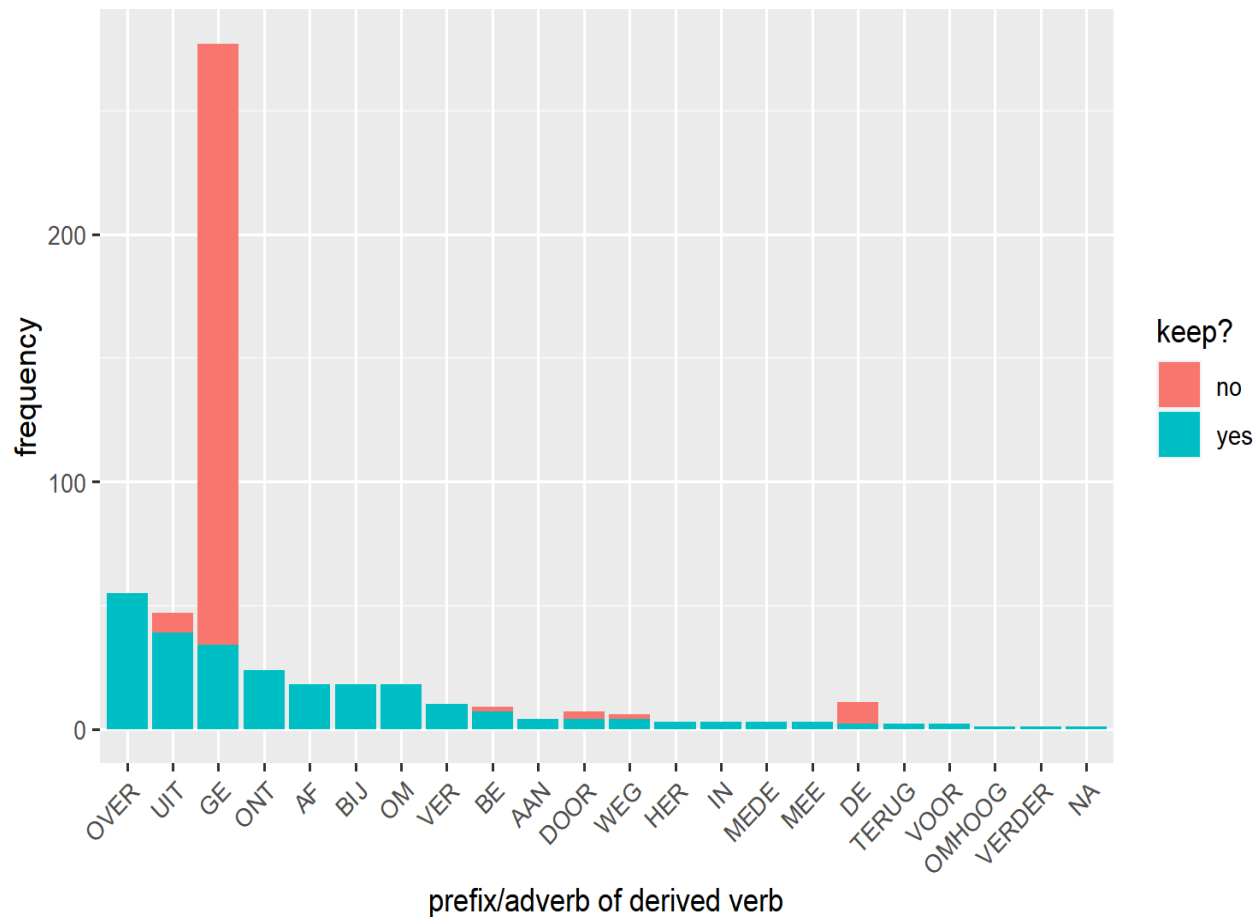
overpimpen ('unnecessary fancify')
Van die boeren met hun overgepimpte auto's en loeiharde,marginale muziek

*Ga die overpimpte dopingwedstrijd geen 3 weken mn TL laten verpleuren.*

uitpimpen ('to prostitute someone')
Poor girl werd gewoon gecaffished en daarna uitgepimpd

uitpimpen ('to finish fancifying something')
Als je uitgepimpt bent is er in Gouda nog een ongepimpte versie waar je je mag uitleven ;-)

# Morphological productivity: *pimpen* and its derivations
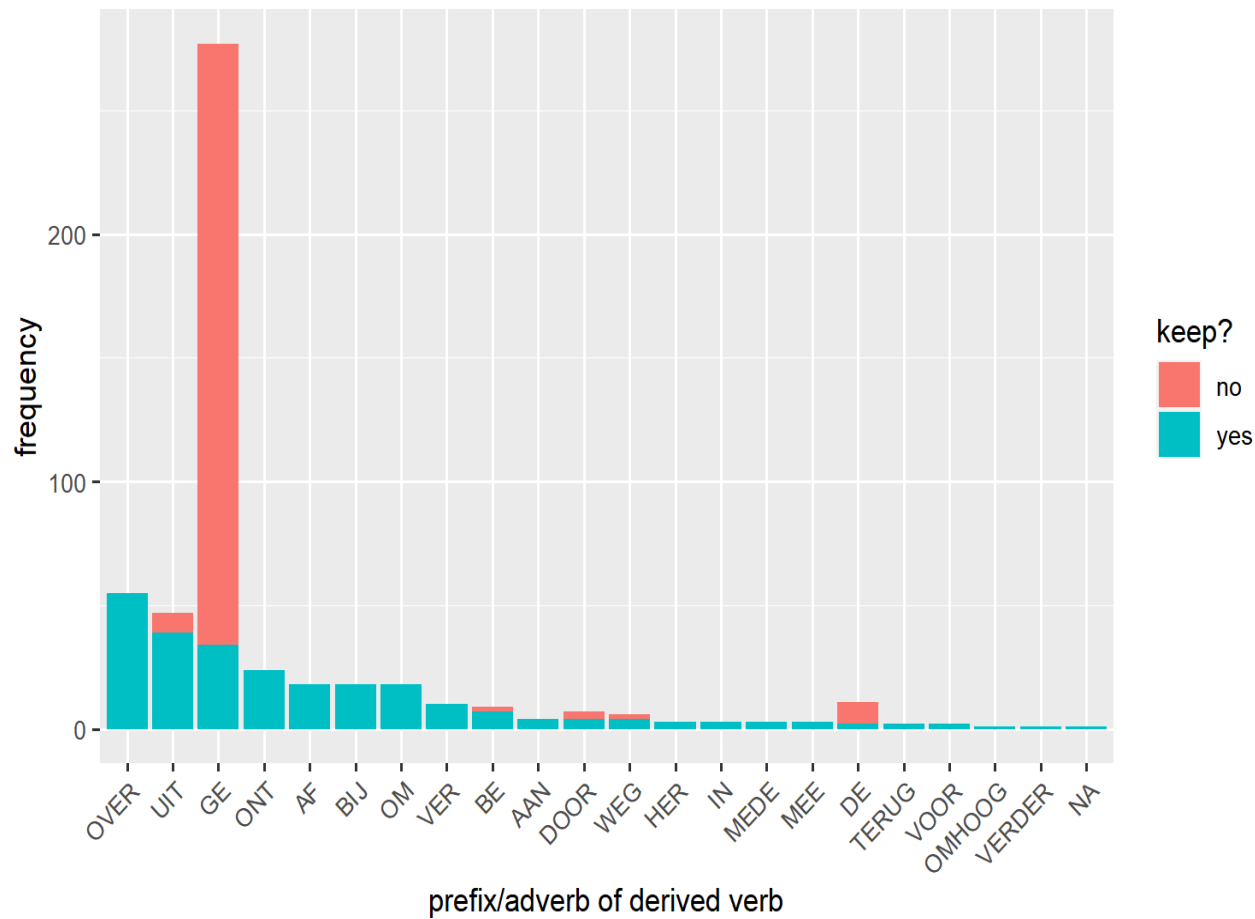
overpimpen ('unnecessary fancification')
Van die boeren met hun **overgepimpte**
auto's en loeiharde,marginale muziek

formal variation

Ga die **overpimpte** *dopingwedstrijd geen 3 weken mn TL laten verpleuren.*

uitpimpen ('to prostitute someone')
Poor girl werd gewoon gecaffished en daarna uitgepimpd

semantic variation

uitpimpen ('to finish fancifying something')
Als je uitgepimpt bent is er in Gouda nog een ongepimpte versie waar je je mag uitleven ;-)

Morphological productivity: *pimpen* and its derivations
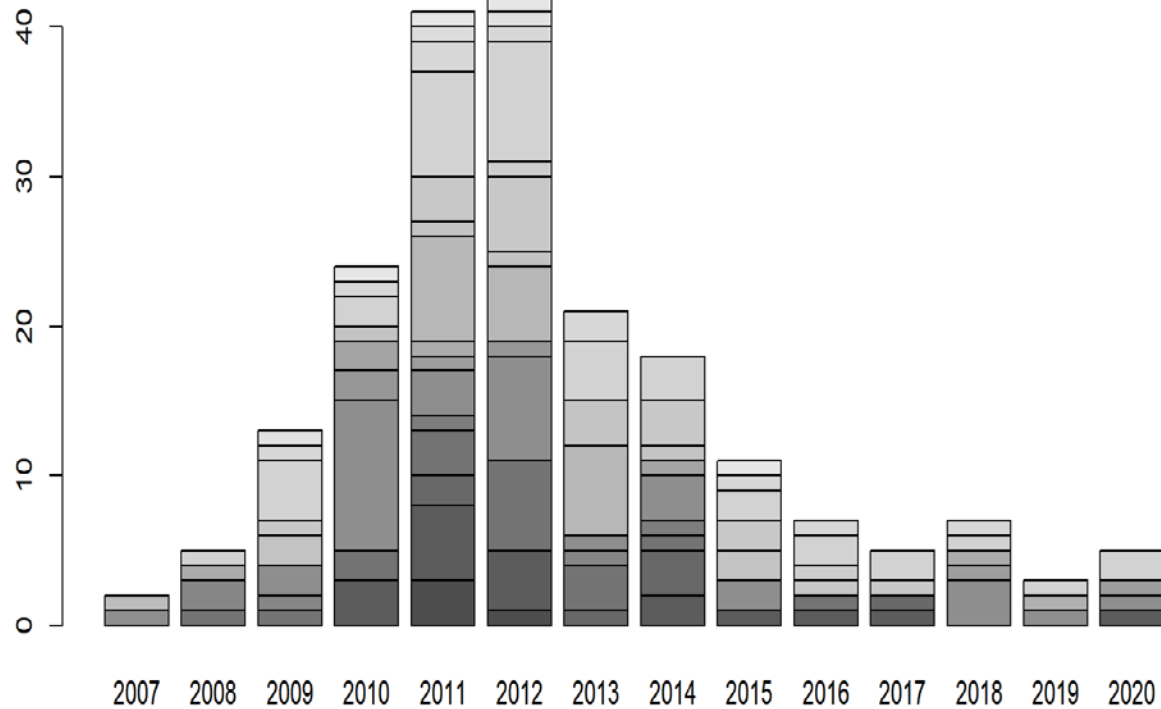
how to measure productivity in the long tail of infrequent events? [Baayen 2009]

- relative type frequency? type/token ratio?

  — too dependent on sample size (which is unknown)

  — frequency distribution of *pimpen* derivations is strongly dependent on popularity of Twitter!

  — underestimation of rare events

Morphological productivity: *pimpen* and its derivations
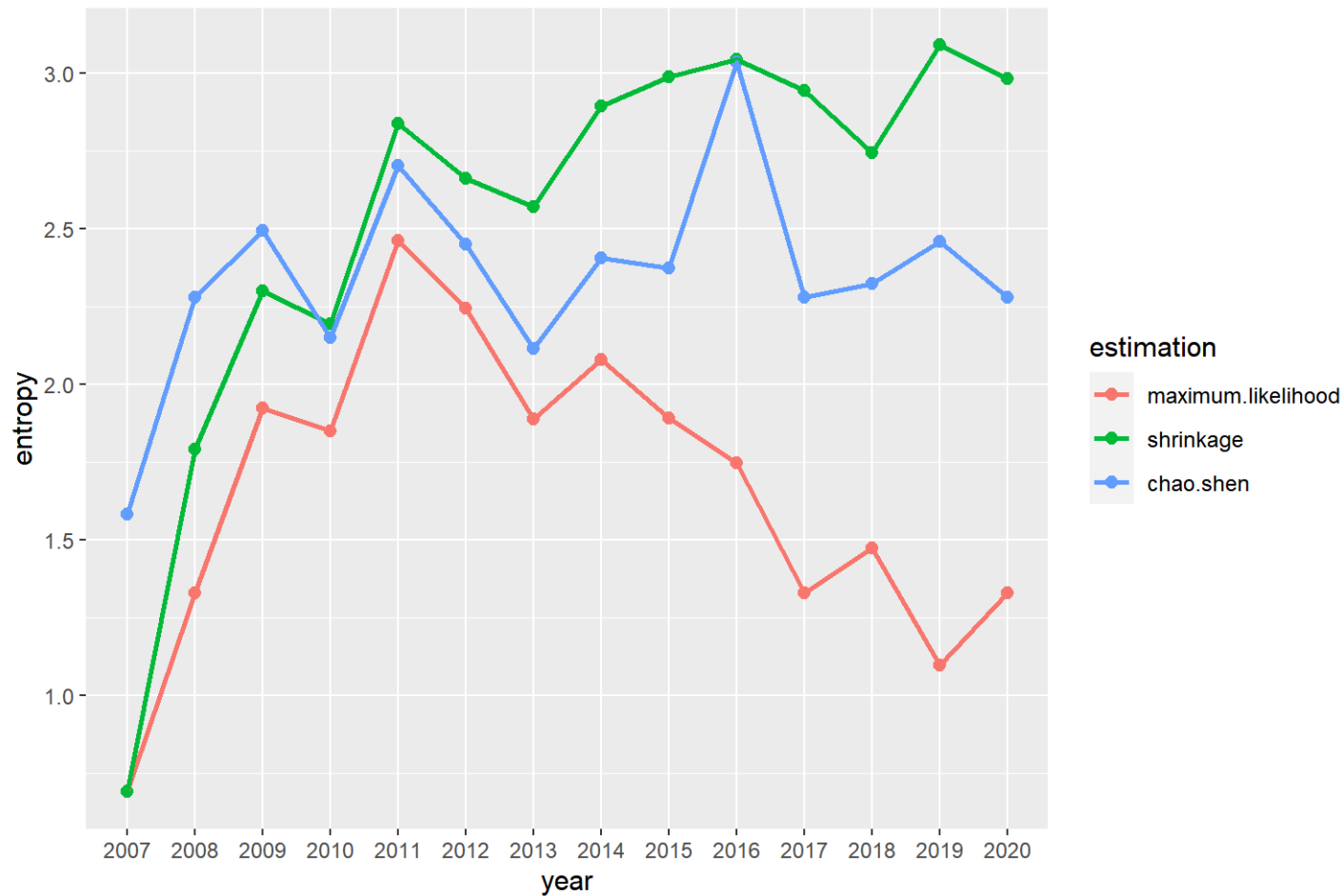
- solution? corrected Shannon entropy measures (Gotelli & Chao, 2013)
  - estimate the amount of diversity in a frequency distribution

- no correction = maximum likelihood
  - diversity of morphological family decreases over time

- correction for rare and unseen forms = shrinkage factor, Chao-Shen estimation
  - diversity of morphological family increases and stabilizes over time

Conslusions

- replications and extensions of 2010 study on *pimpen*

  — more data did not seem to confirm the neat trends that were visible in the small dataset: no increased deconstructionalization

  — perhaps constructional change already completed?

- focus on onomasiological competition:

  — predictions borne out for sex, country and 'formality'

  — what is Twitter really?!

- focus on morphological productivity:

  — sustained diversity over the years → moderate productivity of *pimp*

  — careful with calculations → opportunities by big data for the long tail

Bibliography

- Baayen, R. Harald. 2009. 41. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook. Volume 2*, 899–919. Berlin/New York: De Gruyter Mouton.

- Gotelli, Nicholas J and Anne Chao. 2013. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In Simon Asher Levin (ed.), *Encyclopedia of Biodiversity*, 195–211. 2nd ed. Waltham , MA: Academic Press.

- Pascale, Stefano De, Dirk Pijpops, Freek Van de Velde and Eline Zenner. under review. Reassembling the pimped ride: A quantitative take on the integration of a borrowed expression. *Frontiers in Communication*.

- Petré, Peter and Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94(4). 867–901.

- Velde, Freek Van de and Eline Zenner. 2010. Pimp my Lexis: het nut van corpusonderzoek in normatief taaladvies. In Els Hendrickx, Karl Hendrickx, Willy Martin, Hans Smessaert, William Van Belle & Joop van der Horst (eds.), *Liever meer of juist minder? Over normen en variatie in taal*, 51–68. Gent: Academia Press.

- Velde, Freek Van de and Eline Zenner. 2015. Pimp my Lexis: Constructional borrowing. Paper presented at the Taal & Tongval Colloquium, Gent, 27 Nov 2015.
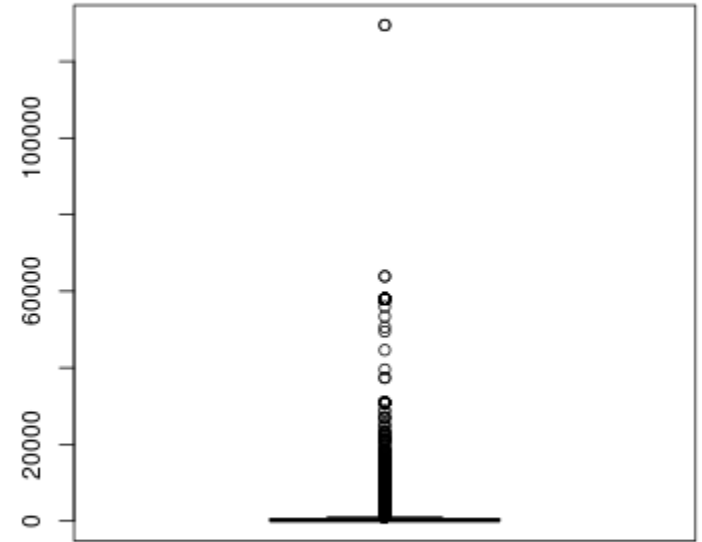
EXTRA SLIDES
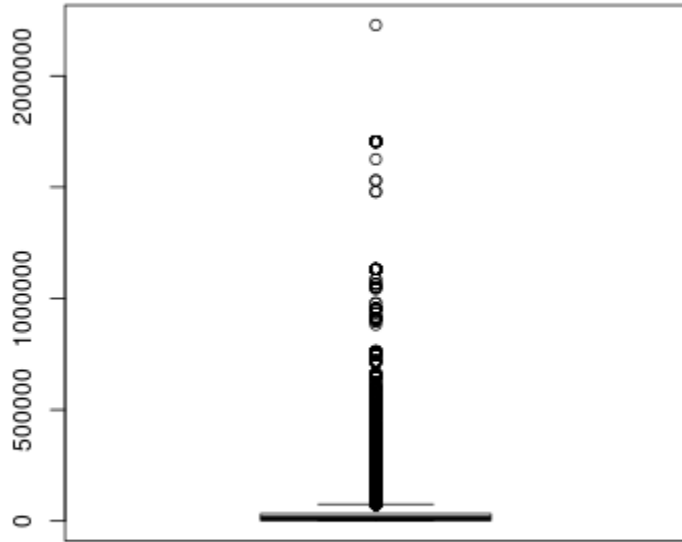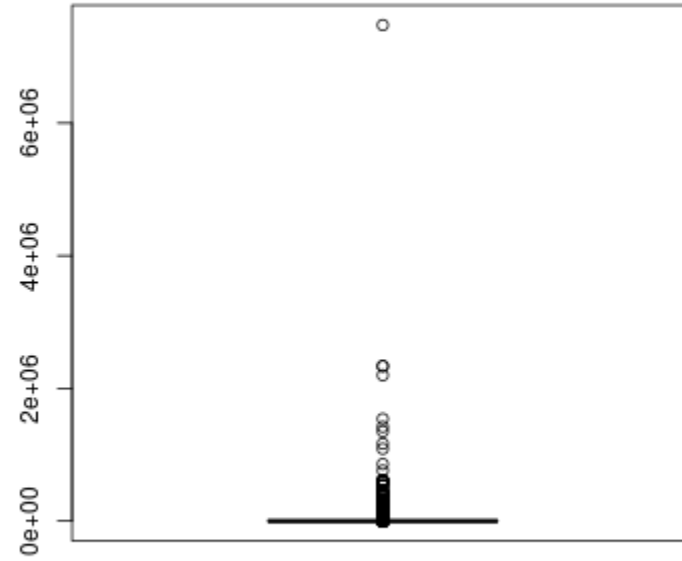
No transformation

Root 2 transformation

Root 10 transformation
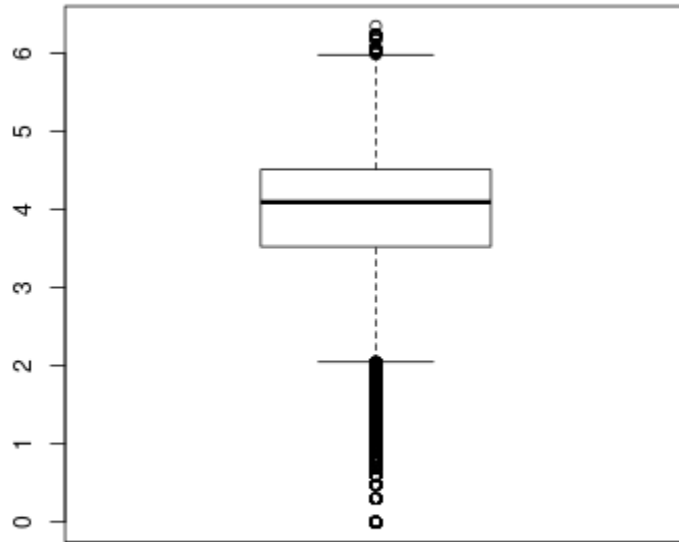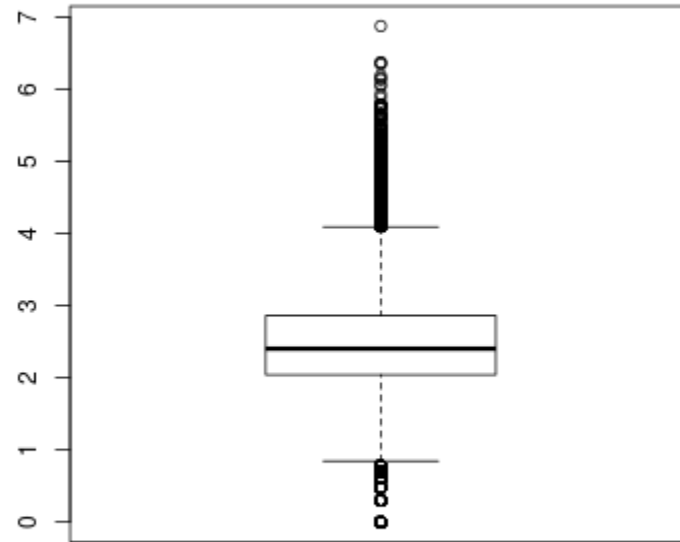
Logarithm 10 transformation

log10_user_statuses_clean01
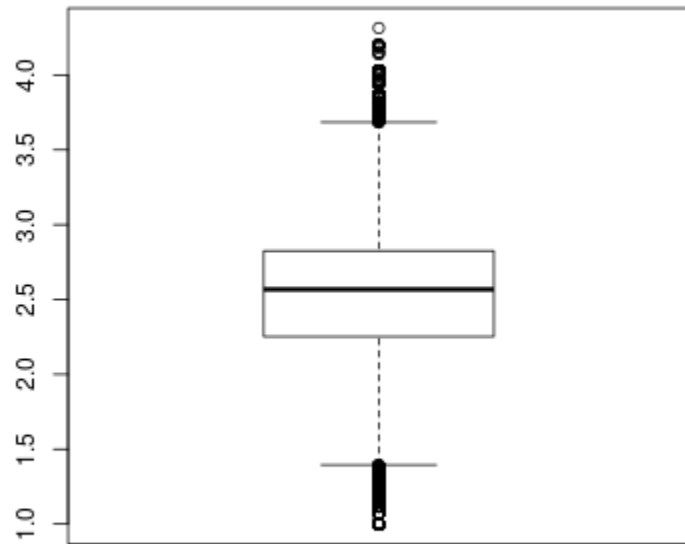
log10_user_followers_clean01

log10_user_friends_clean01