KU LEUVEN

Core vocabulary, borrowability, and entrenchment

A usage-based onomasiological approach

Eline Zenner

FWO Flanders/University of Leuven
RU Quantitative Lexicology and Variational Linguistics

Universalities of Borrowing

Universalities of Borrowing

1. Nouns are more borrowable than verbs (POS-clines)

2. Core vocabulary is highly resistant to borrowing

Universalities of Borrowing

1. Nouns are more borrowable than verbs (POS-clines)

2. Core vocabulary is highly resistant to borrowing

Universalities of Borrowing

1. Nouns are more borrowable than verbs (POS-clines)

2. Core vocabulary is highly resistant to borrowing

Analytic proposition	Synthetic proposition
resistance to borrowing is the defining feature of coreness	coreness is independent from resistance to borrowing
« bachelors are unmarried »	« bachelors are unhappy »

Universalities of Borrowing

1. Nouns are more borrowable than verbs (POS-clines)

2. Core vocabulary is highly resistant to borrowing

empirical testing is needed ←

Synthetic proposition
coreness is independent from resistance to borrowing
« bachelors are unhappy »

Universalities of Borrowing

1. Nouns are more borrowable than verbs (POS-clines)

2. Core vocabulary is highly resistant to borrowing

A usage-based approach ←

Synthetic proposition
coreness is independent from resistance to borrowing
« bachelors are unhappy »

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary
2. define resistance to borrowing
3. establish the link between both

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary
2. define resistance to borrowing
3. establish the link between both

Core Vocabulary

mainly applied

Glottochronology: rates of change in vocabulary

wrong estimations of time depths

Lexicostatistics: genealogical relatedness between languages

- false positives
- false negatives

Applied Linguistics

- dictionaries: lemma writing
- L2: vocabulary lists

Main Issues

- → theoretical unclarity: what does "coreness" mean?
- → dichotomous approach: *lists* of core items
- → terminological inconsistency

Main Issues

→ theoretical unclarity: what does "coreness" mean?

= resistance to borrowing (analytic proposition)

= stability / resistance to change in general

= universality

= semantically general

= highly frequent

= ...

→ dichotomous approach: *lists* of core items

→ terminological inconsistency

Main Issues

- → theoretical unclarity: what does "coreness" mean?
- → dichotomous approach: *lists* of core items
  - Swadesh 100
  - how long should the list be?
  - we cannot expect "each item of any finite list to be basic in every respect" (Hymes 1960: 11)
  - better to have a continuous measure (that can be assigned to *every* meaning/concept)
- → terminological inconsistency

Main Issues

- → theoretical unclarity: what does "coreness" mean?
- → dichotomous approach: *lists* of core items
- → terminological inconsistency

  core vocab – core meaning – core concepts

Proposed Solution

Proposed Solution

Link coreness to entrenchment

Proposed Solution

Link coreness to entrenchment

- captures the idea behind coreness quite well

  "the degree to which the formation and activation of a cognitive unit is routinized and automated" (Schmid 2010)

  "well-entrenched structures can inhibit or even block the adoption of novel structures" (ibid.)

Proposed Solution

Link coreness to entrenchment

- captures the idea behind coreness quite well

  "the degree to which the formation and activation of a cognitive unit is routinized and automated" (Schmid 2010)

  "well-entrenched structures can inhibit or even block the adoption of novel structures" (ibid.)

- explicitly conceptual

  it is not "real-world entities themselves that get entrenched but possible concepts of entities" (ibid.)

Proposed Solution

Link coreness to entrenchment

- captures the idea behind coreness quite well

  "the degree to which the formation and activation of a cognitive unit is routinized and automated" (Schmid 2010)

  "well-entrenched structures can inhibit or even block the adoption of novel structures" (ibid.)

- explicitly conceptual

  it is not "real-world entities themselves that get entrenched but possible concepts of entities" (ibid.)

- explicitly continuous

  there is a "continuous scale of entrenchment in cognitive organization" (Langacker 1987: 59)

Proposed Solution

Link coreness to entrenchment

- captures the idea behind coreness quite well

  "the degree to which the formation and activation of a cognitive unit is routinized and automated" (Schmid 2010)

  "well-entrenched structures can inhibit or even block the adoption of novel structures" (ibid.)

- explicitly conceptual

  it is not "real-world entities themselves that get entrenched but possible concepts of entities" (ibid.)

- explicitly continuous

  there is a "continuous scale of entrenchment in cognitive organization" (Langacker 1987: 59)

  → frequency of usage

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary
2. define resistance to borrowing
3. establish the link between both

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary → entrenchment
2. define resistance to borrowing
3. establish the link between both

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary → entrenchment
2. define resistance to borrowing
3. establish the link between both

What does resistance to borrowing mean?

What does resistance to borrowing mean?

→ resistance to replacement ~~native~~ borrowed

What does resistance to borrowing mean?

→	resistance to replacement	~~native~~	borrowed
→	resistance to co-existence	native	borrowed

What does resistance to borrowing mean?

→	resistance to replacement	~~native~~	borrowed
→	resistance to co-existence	native	borrowed
→	resistance to successful coexistence	native	borrowed

"the words of foreign origin might simply appear as minor alternatives (...), but never become the first choice of most native speakers" (Fischer 1961: 263)

What does resistance to borrowing mean?

→ resistance to replacement
→ resistance to co-existence

types

→ resistance to successful coexistence

tokens

Main Issues

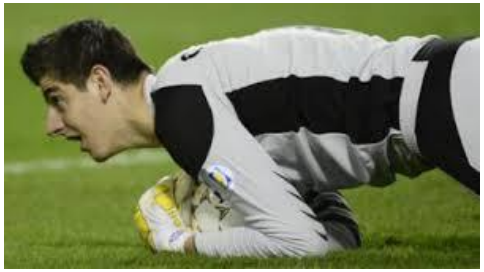
→ resistance to successful coexistence underdeveloped

→ methodological: how to measure coexistence?

Proposed Solution

profile-based method of onomasiological variation

Proposed Solution



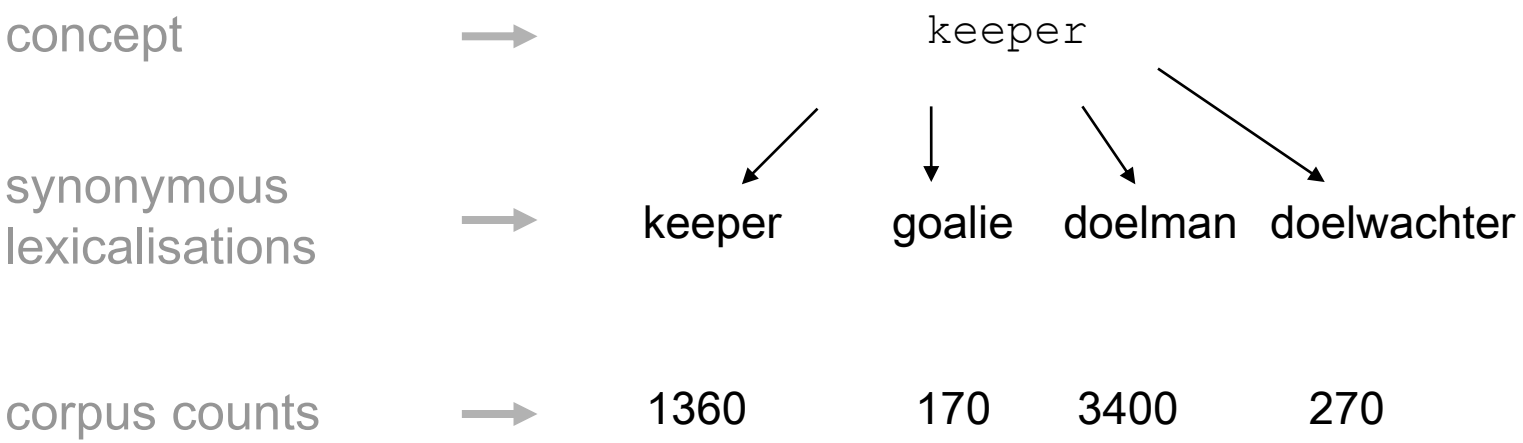
profile-based method of onomasiological variation

concept → `keeper`

synonymous lexicalisations → keeper, goalie, doelman, doelwachter

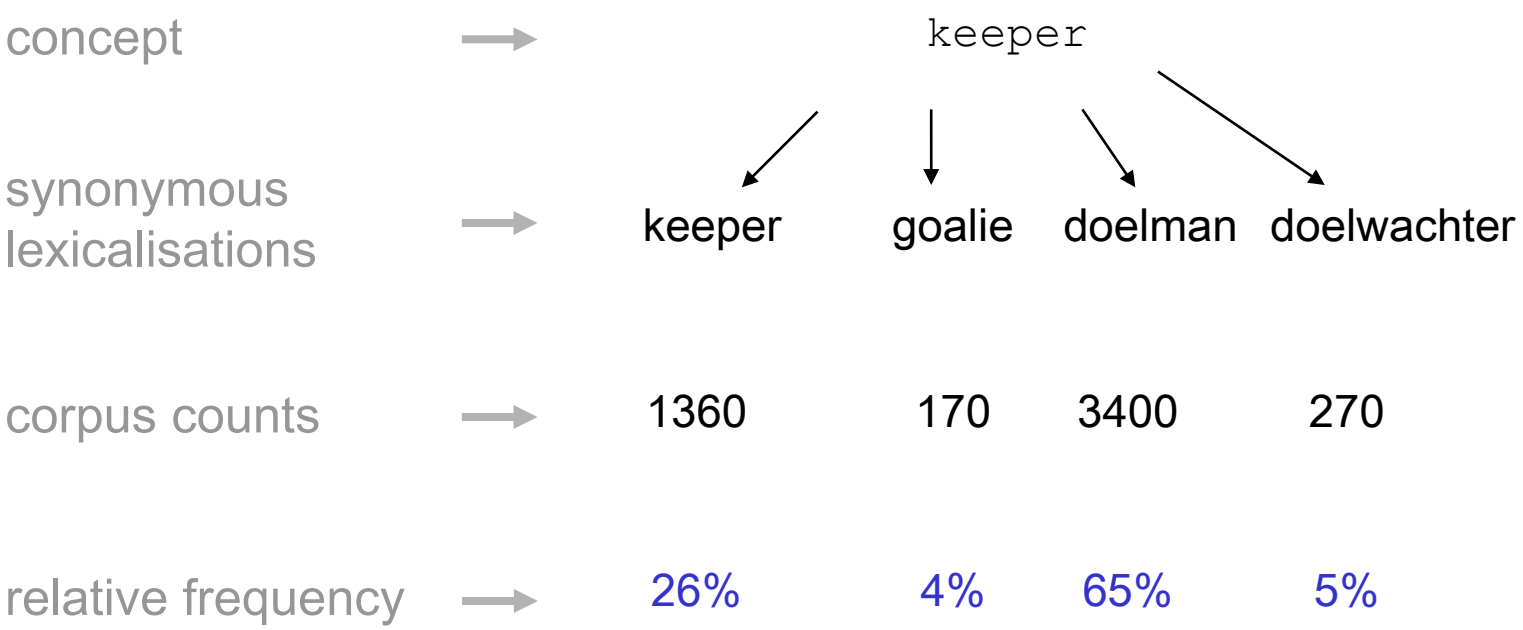
Proposed Solution

profile-based method of onomasiological variation



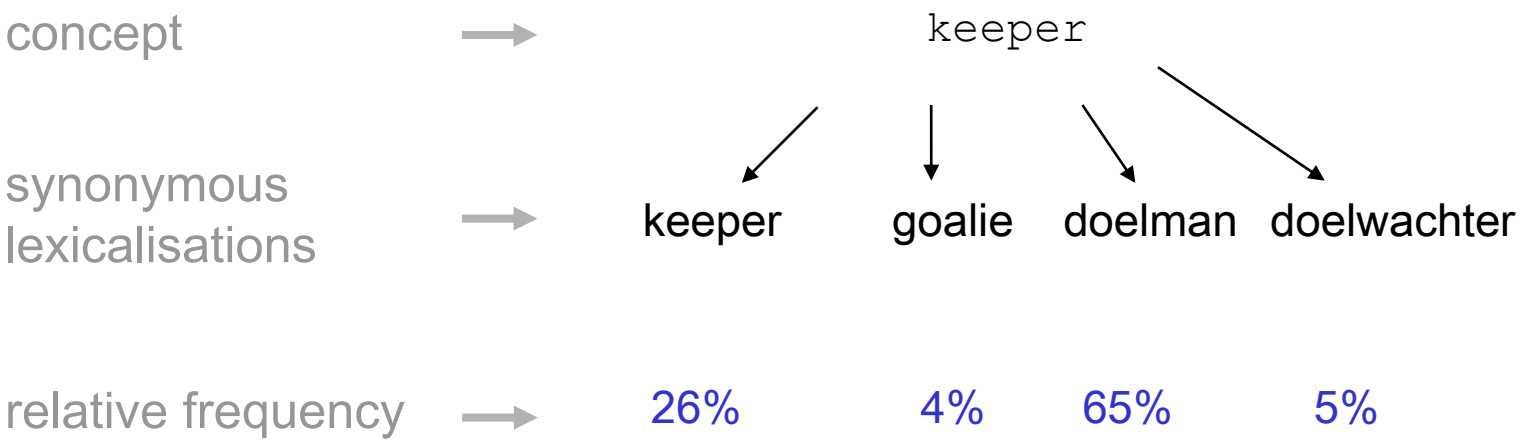
Proposed Solution

profile-based method of onomasiological variation



Proposed Solution

profile-based method of onomasiological variation



per concept	per loanword
resistance to borrowing 100 – (26+4) = 70%	success keeper: 26%
openness to borrowing 26+4 = 30%	success goalie: 4%

Proposed Solution

profile-based method of onomasiological variation

concept → keeper

synonymous lexicalisations → keeper, goalie, doelman, doelwachter

relative frequency → 26%, 4%, 65%, 5%

per loanword
success keeper: 26%
success goalie: 4%

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

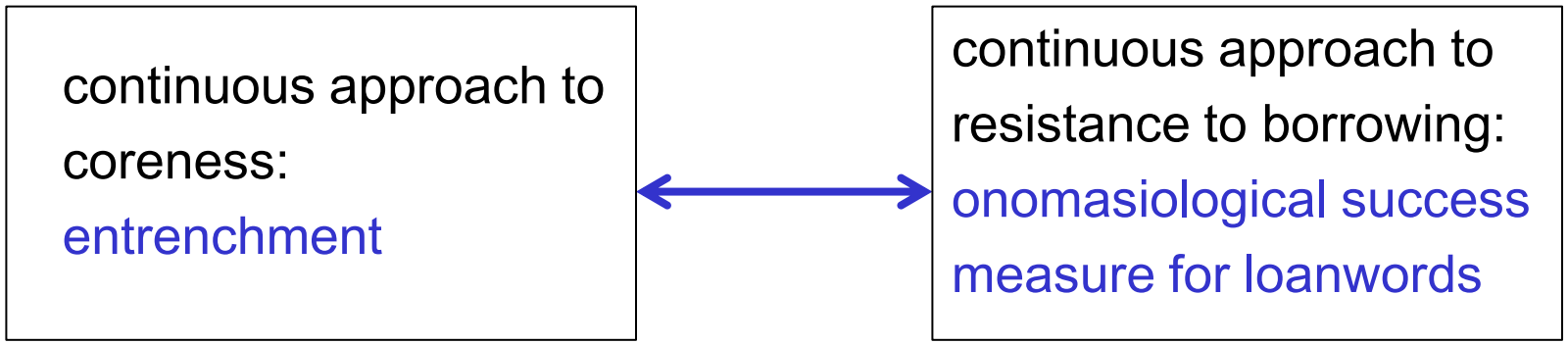
1. define core vocabulary → entrenchment
2. define resistance to borrowing → profile-based
3. establish the link between both

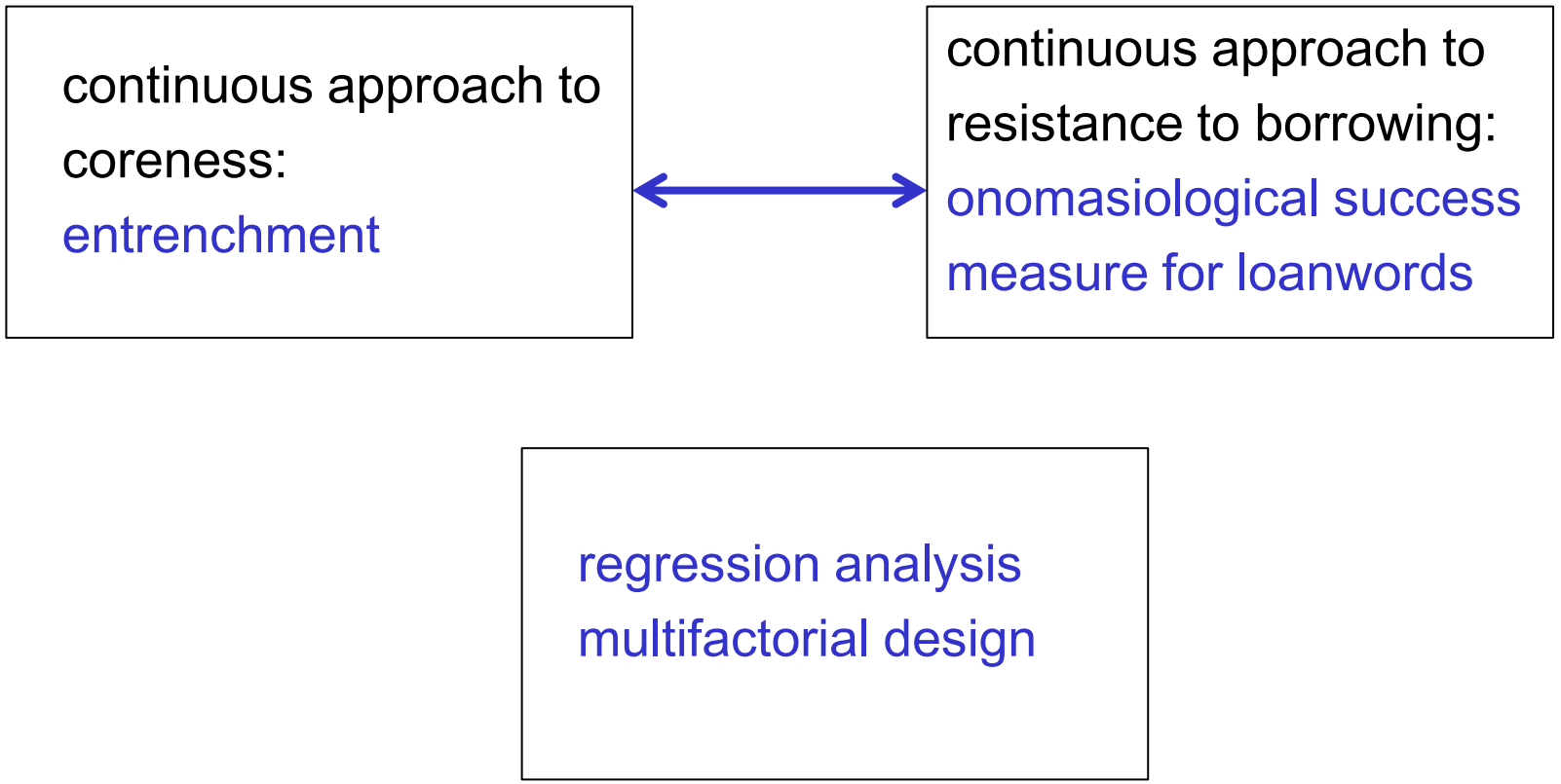
Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary → entrenchment
2. define resistance to borrowing → profile-based
3. establish the link between both


continuous approach to coreness:
entrenchment
continuous approach to resistance to borrowing:
onomasiological success measure for loanwords


continuous approach to coreness:
entrenchment
continuous approach to resistance to borrowing:
onomasiological success measure for loanwords
regression analysis
multifactorial design

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. define core vocabulary → entrenchment
2. define resistance to borrowing → profile-based
3. establish the link between both → statistics

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. resistance to borrowing: success measures for 150 English loanwords
2. defining coreness: entrenchment-level
3. competing predictors of success
4. regression analysis
5. results

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

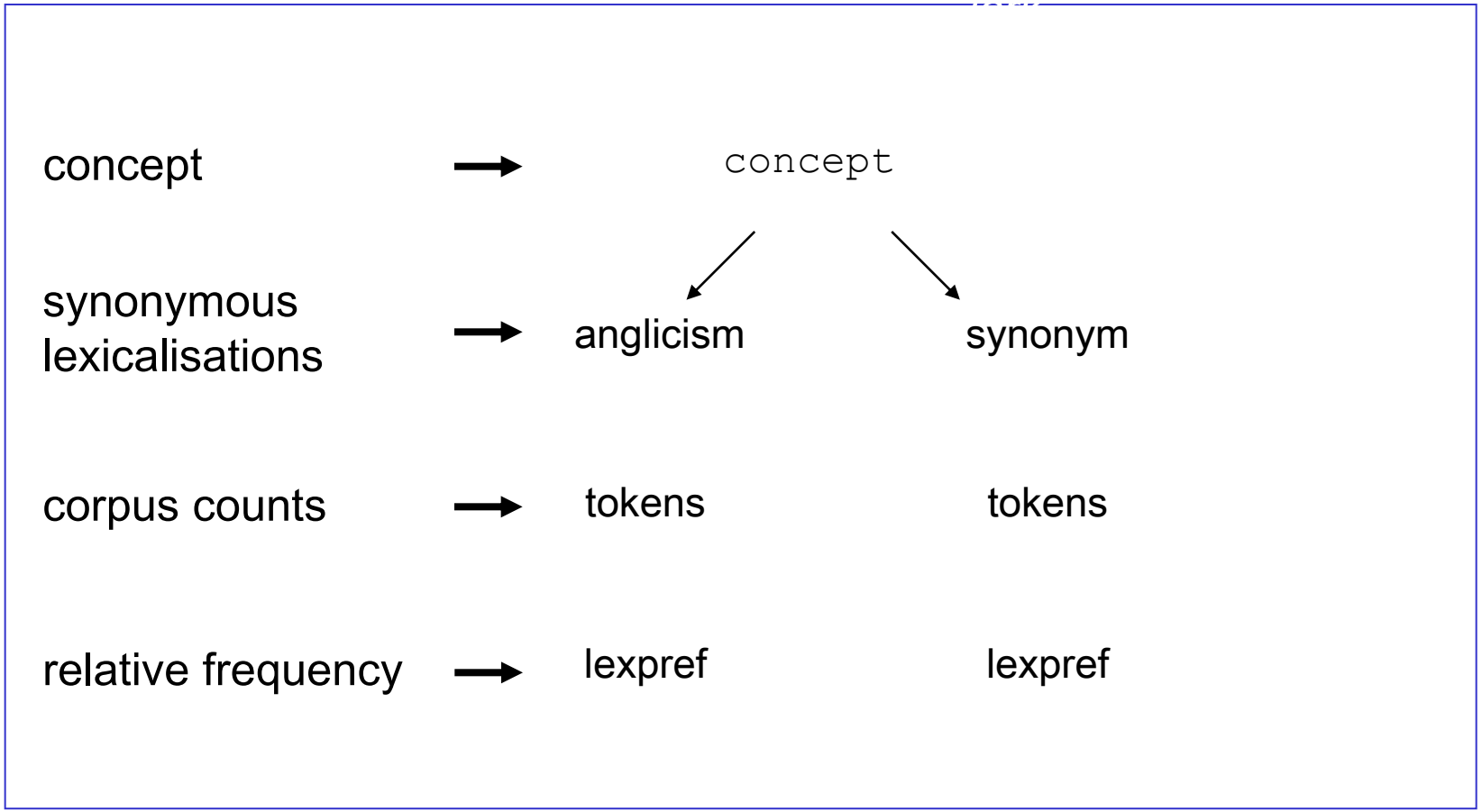
1. resistance to borrowing: success measures for 150 English loanwords
2. defining coreness: entrenchment-level
3. competing predictors of success
4. regression analysis
5. results

A. Corpus

Two Dutch newspaper corpora (parsed, lemmatised)

-	TwNC	Netherlandic Dutch	1999-2002	300 million words
-	LeNC	Belgian Dutch	1999-2005	1.3 billion words

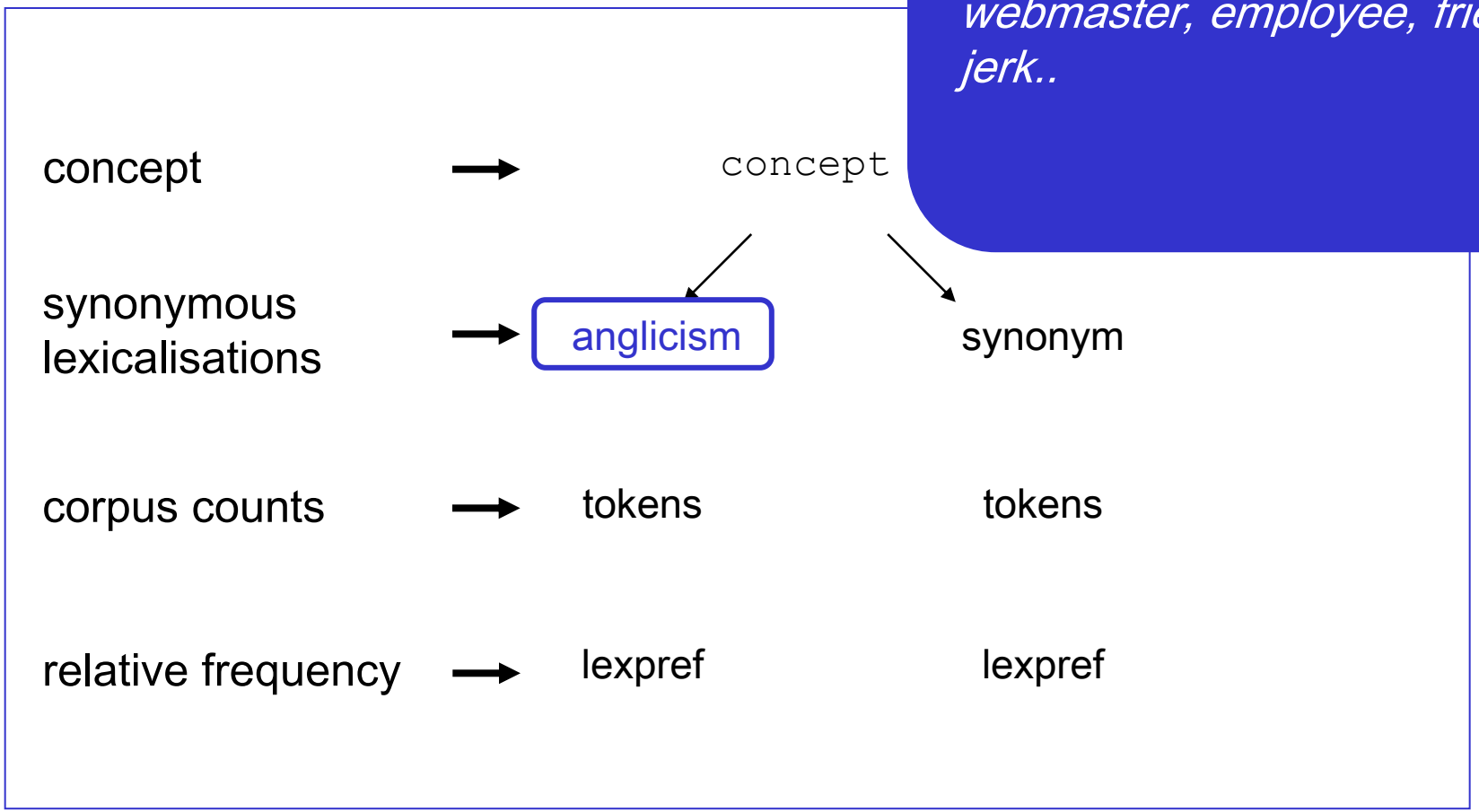
B. Profile-Based Method



B. Profile-Based Method

a. English person reference nouns:

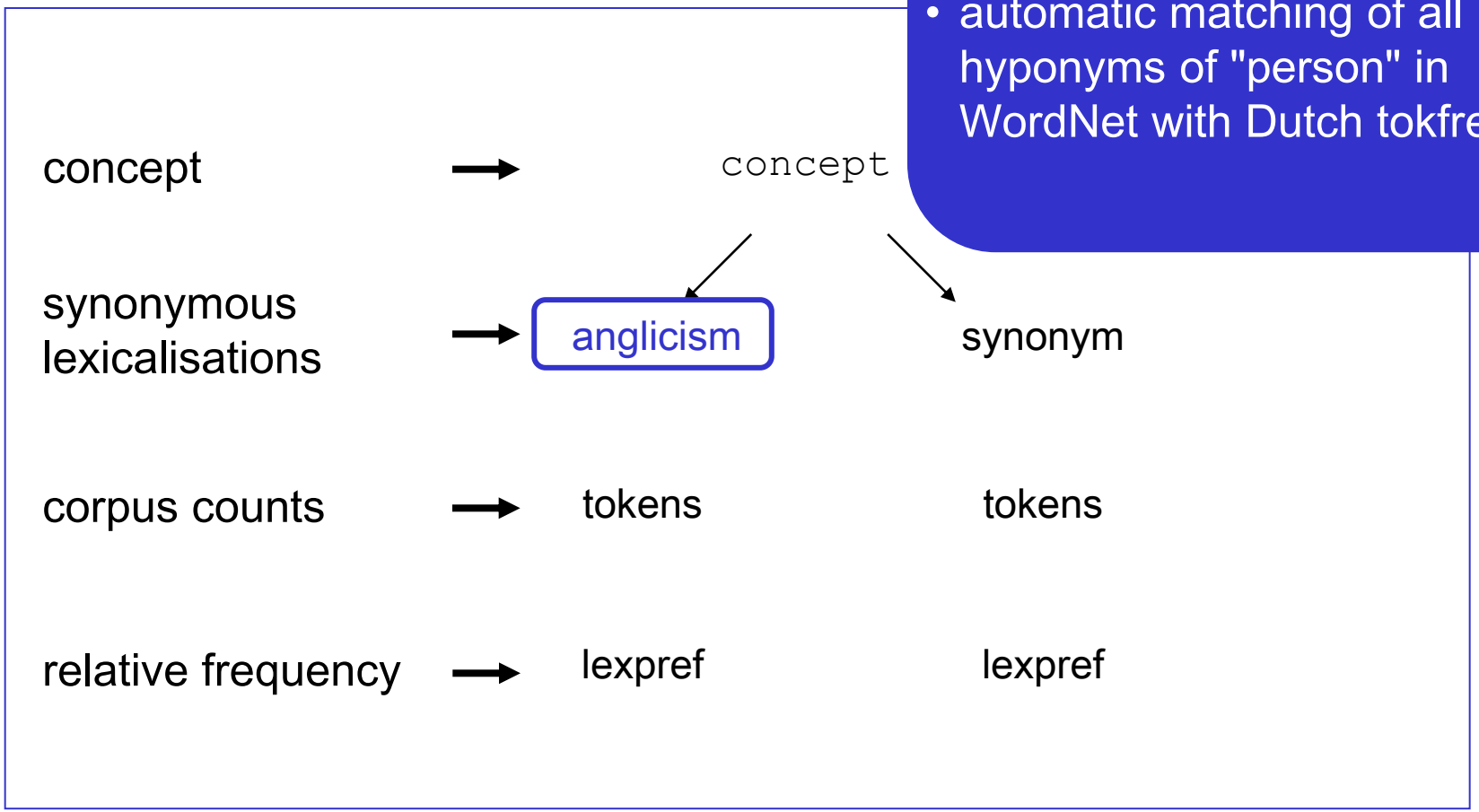
manager, babysitter, hooligan, webmaster, employee, friend, jerk..



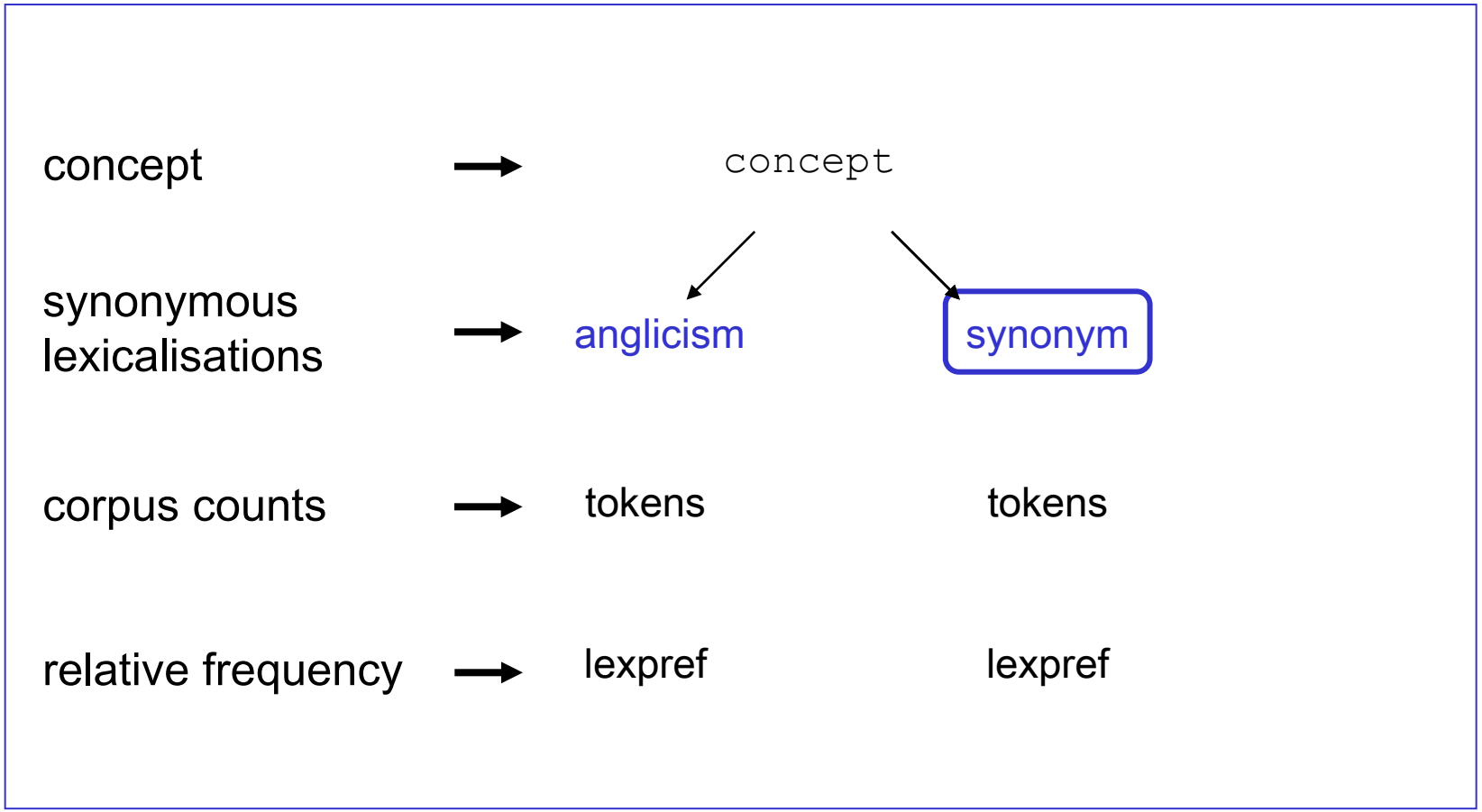
B. Profile-Based Method

selection of 150 English PRN occurring in Dutch:

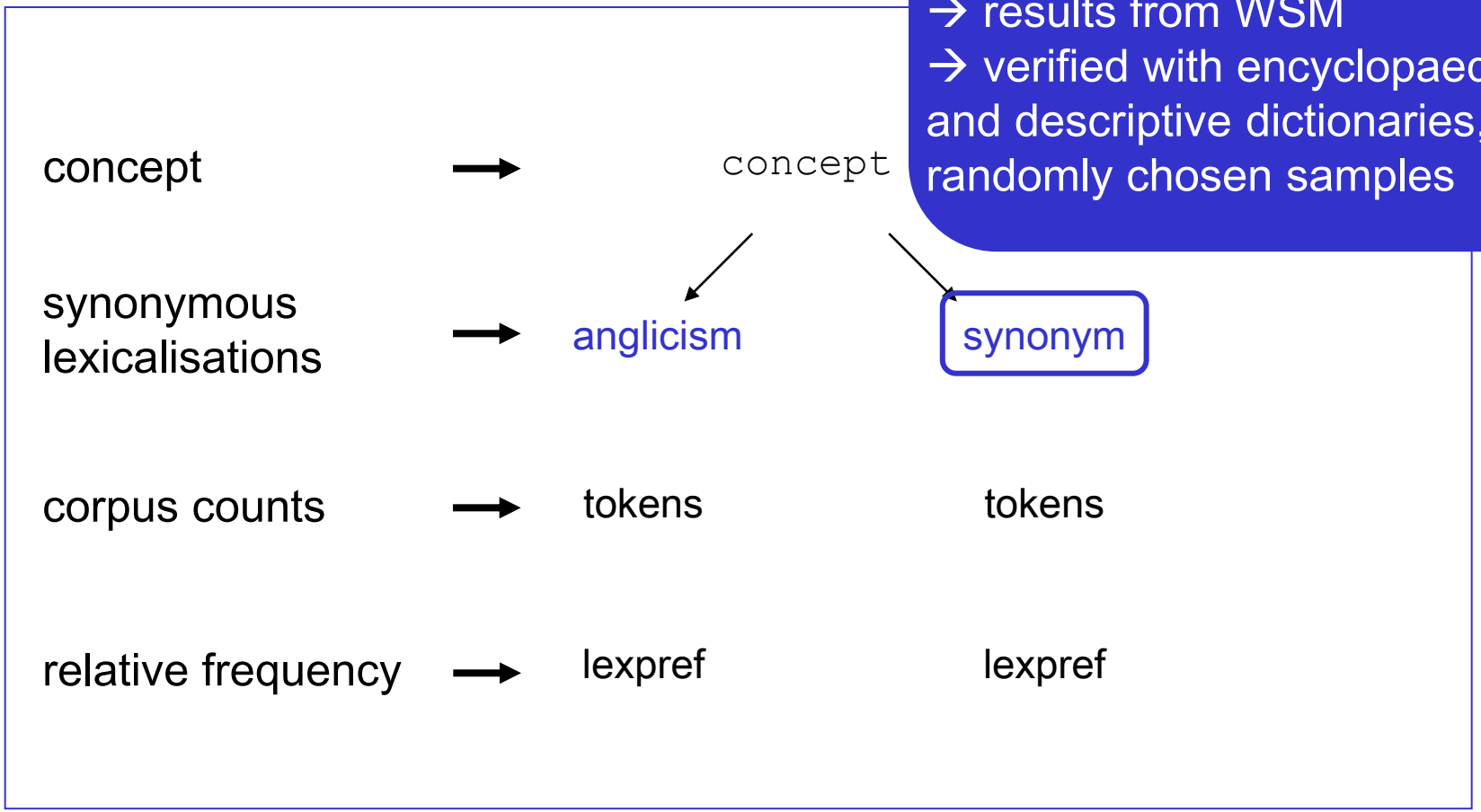
- lexicographical sources
- automatic matching of all hyponyms of "person" in WordNet with Dutch tokfreqlist



B. Profile-Based Method



B. Profile-Based Method



looking for synonyms
no blind trust in lexicography

→ 10 different lex.sources
→ results from WSM
→ verified with encyclopaedia's and descriptive dictionaries, 200 randomly chosen samples

B. Profile-Bas

concept

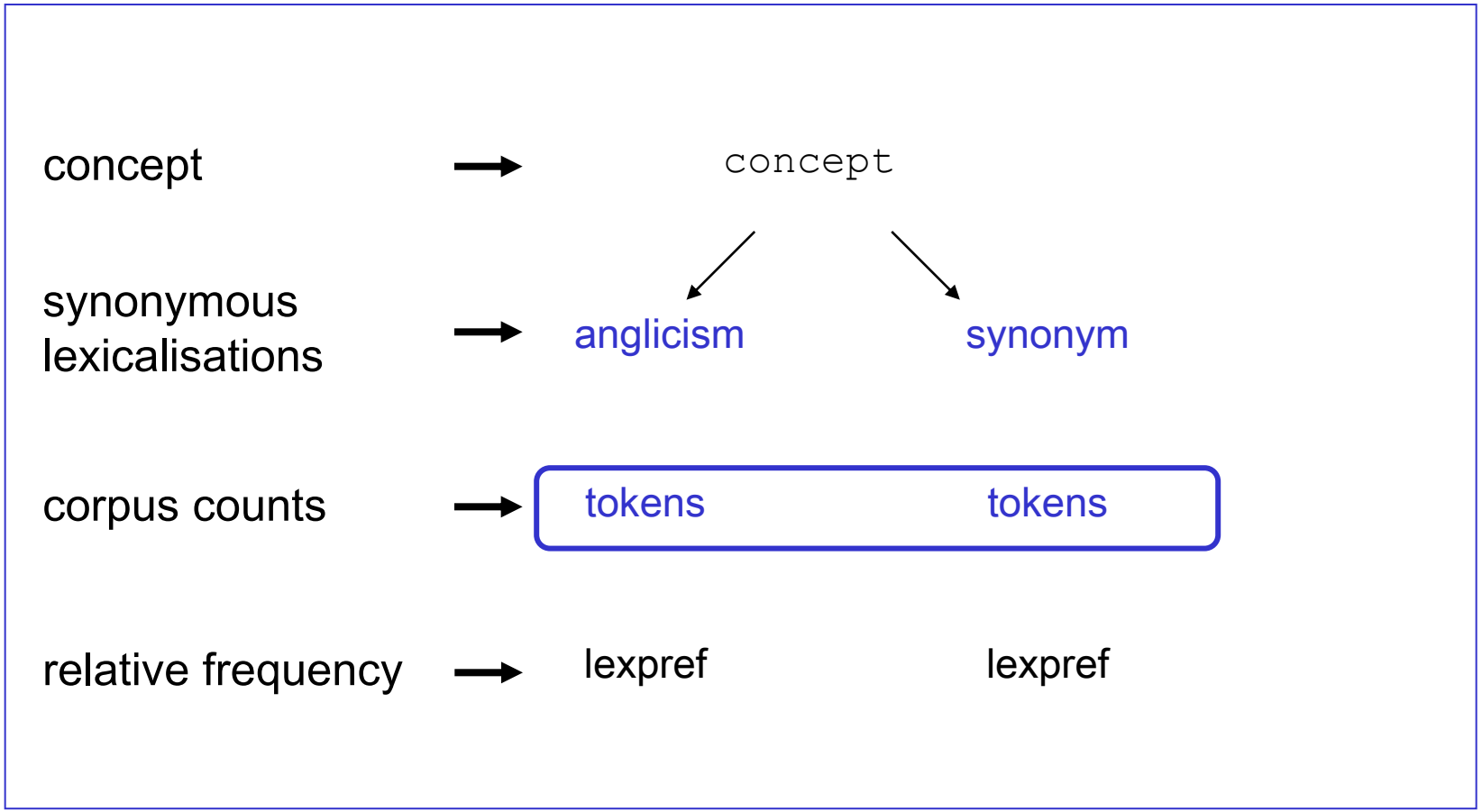
synonymous lexicalisations

corpus counts

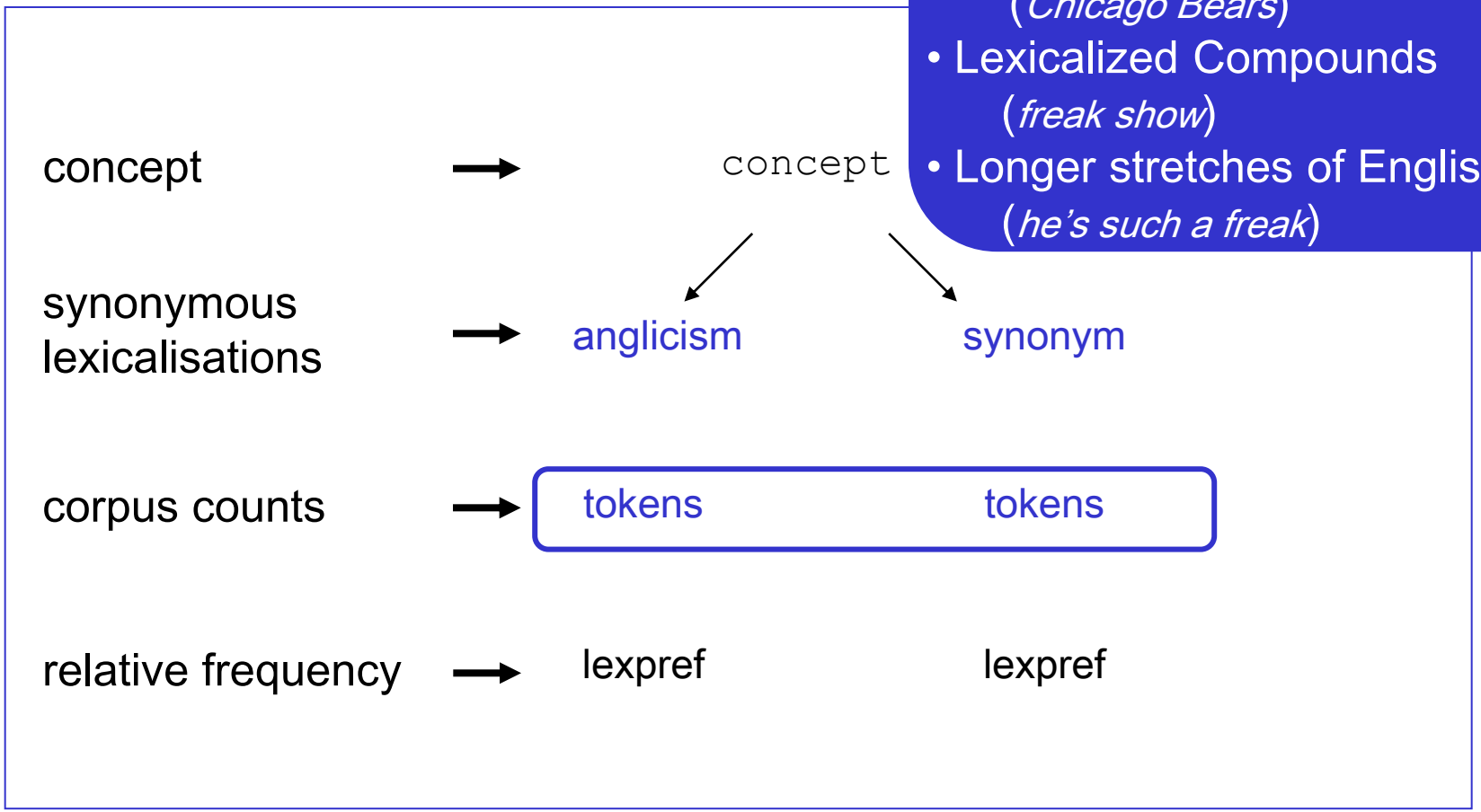
relative frequency

<u>profiles: examples</u>
babyboomer – boomer – geboortegolver
babysitter – babysit – kinderoppas
backpacker – rugzakker – rugzaktoerist
bitch – cunt – teef – feeks – kreng – kutwijf – secreet
copycat – na-aper – nabootser
foodie – culi
freak[fan] – fanatiekeling – fanaticus – fanaat
freak[weird] – weirdo – zonderling – excentriekeling
goalgetter – goaltjesdief – doelpuntenmachine
hacker – computerkraker
jobhopper
jogger
merchandiser – verkoopadviseur – verkoopstrateeg
trader – beurshandelaar
workaholic – werkverslaafde - arbeidsmaniak

B. Profile-Based Method



B. Profile-Based Method



Automatic extraction

Noise (automatically excluded)

- Proper names (*Chicago Bears*)
- Lexicalized Compounds (*freak show*)
- Longer stretches of English (*he's such a freak*)

B. Profile-Based Method

concept → concept

synonymous lexicalisations → anglicism

corpus counts → tokens

relative frequency → lexpref

Polysemy

• **manually**: polysemous items with reasonable frequency (*chicken*)

chicken

B. Profile-Based Method

concept → `concept`

synonymous lexicalisations → anglicism

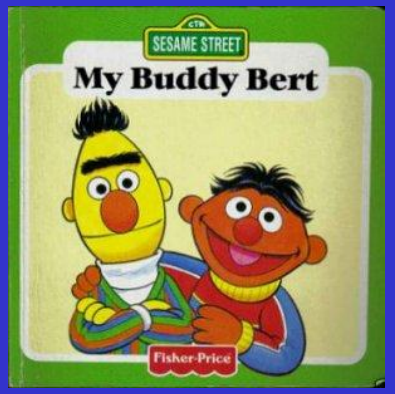
corpus counts → tokens

relative frequency → lexpref

Polysemy

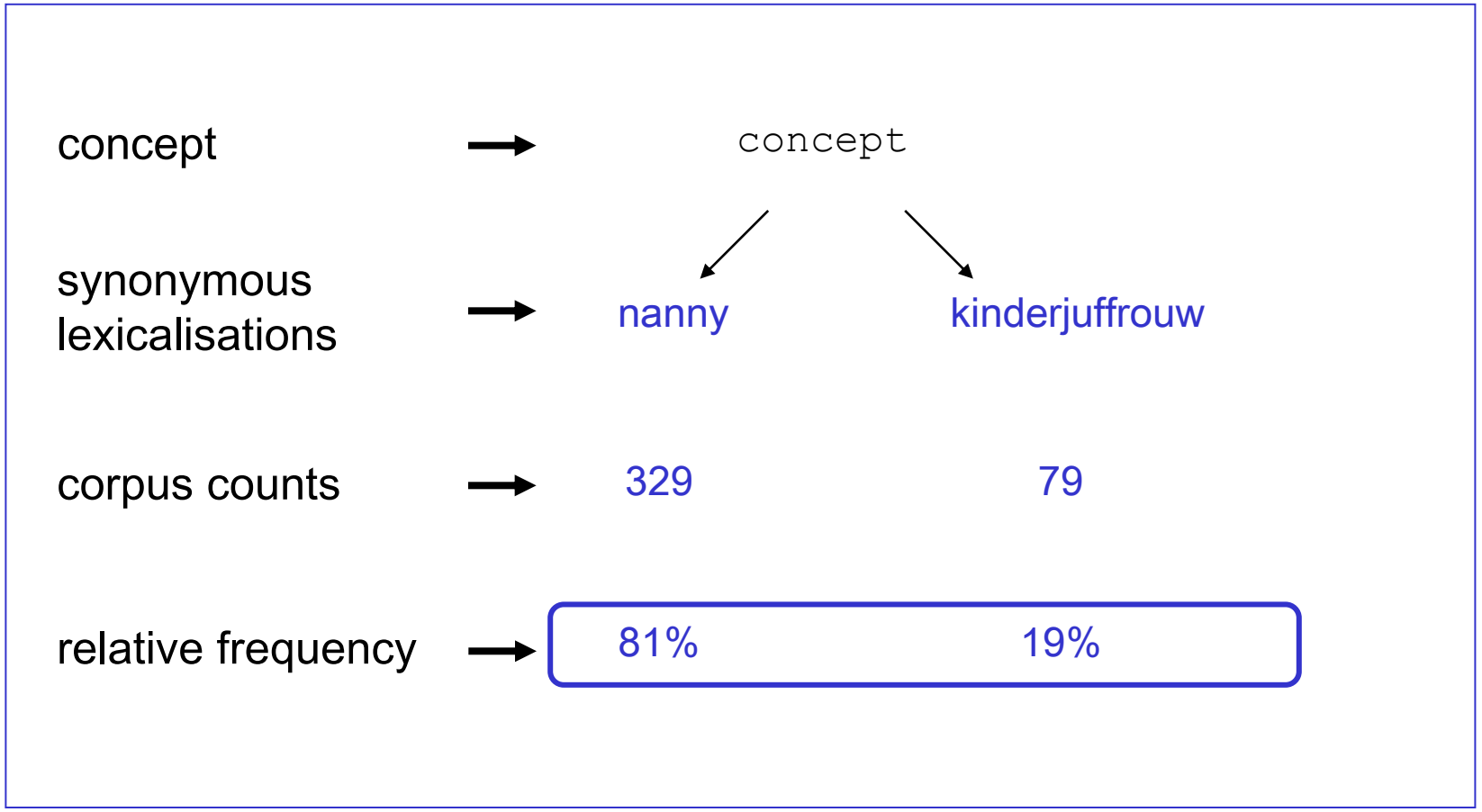
- **manually**: polysemous items with reasonable frequency (*chicken*)
- **semi-automatically or excluced**: concepts with high-frequent polysemous lexicalisations (*maatje*)

maatje



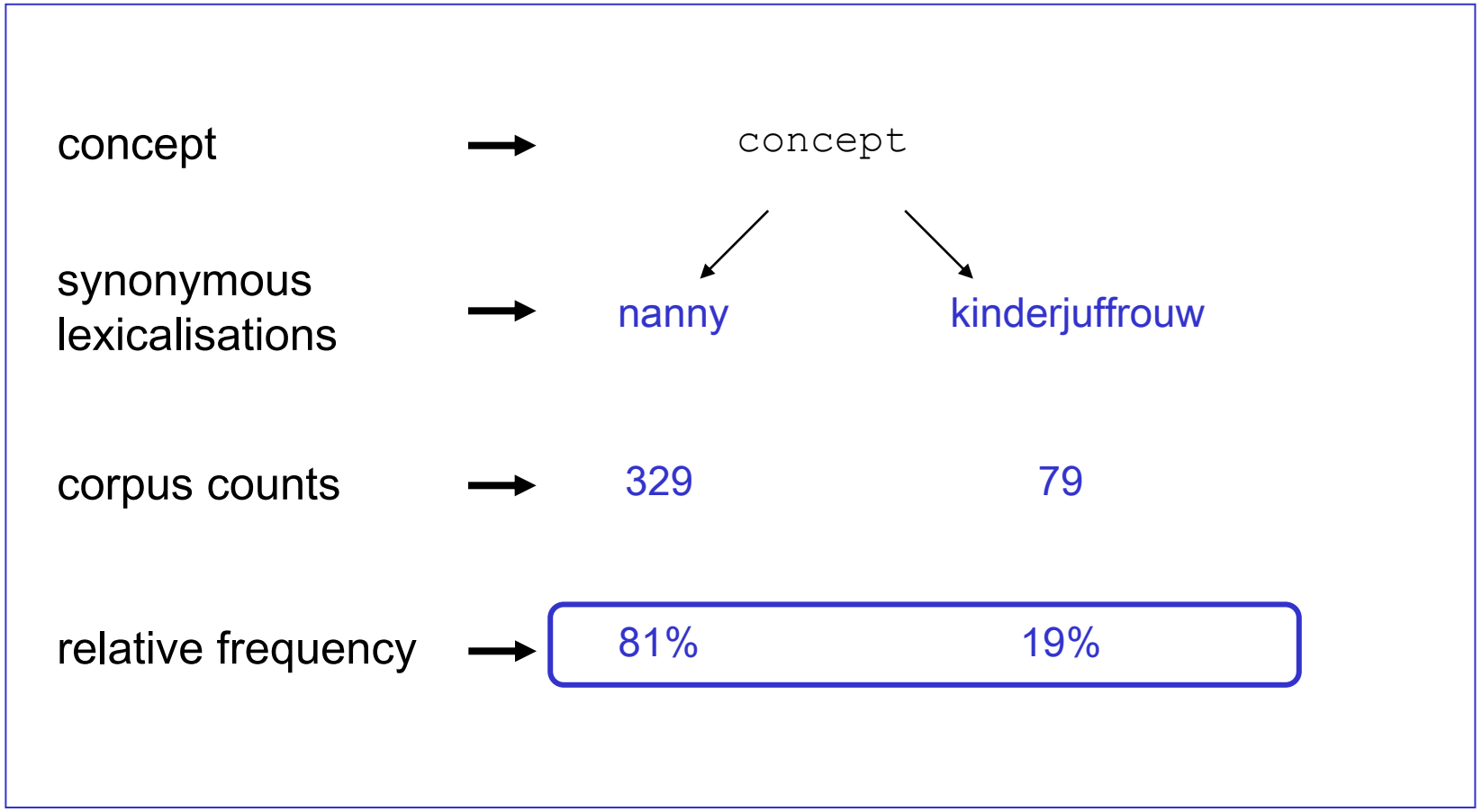
d. calculate success rates

B. Profile-Based Method

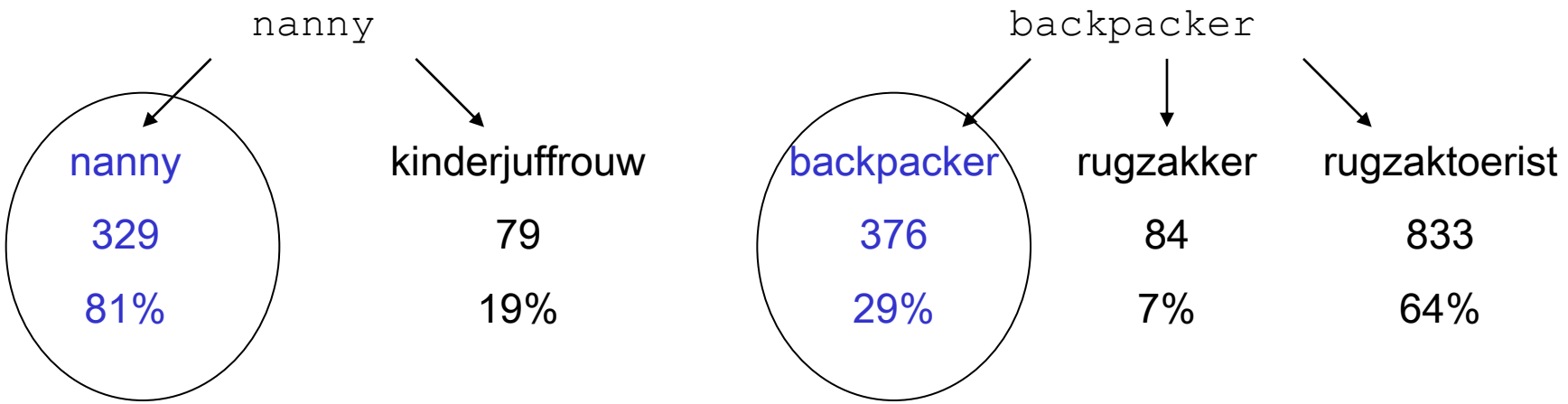


d. calculate success rates

B. Profile-Based Method



C. comparing success of all English PRN



explaining the variation

entrenchment-based vs. other predictors

Core vocabulary is resistant to borrowing

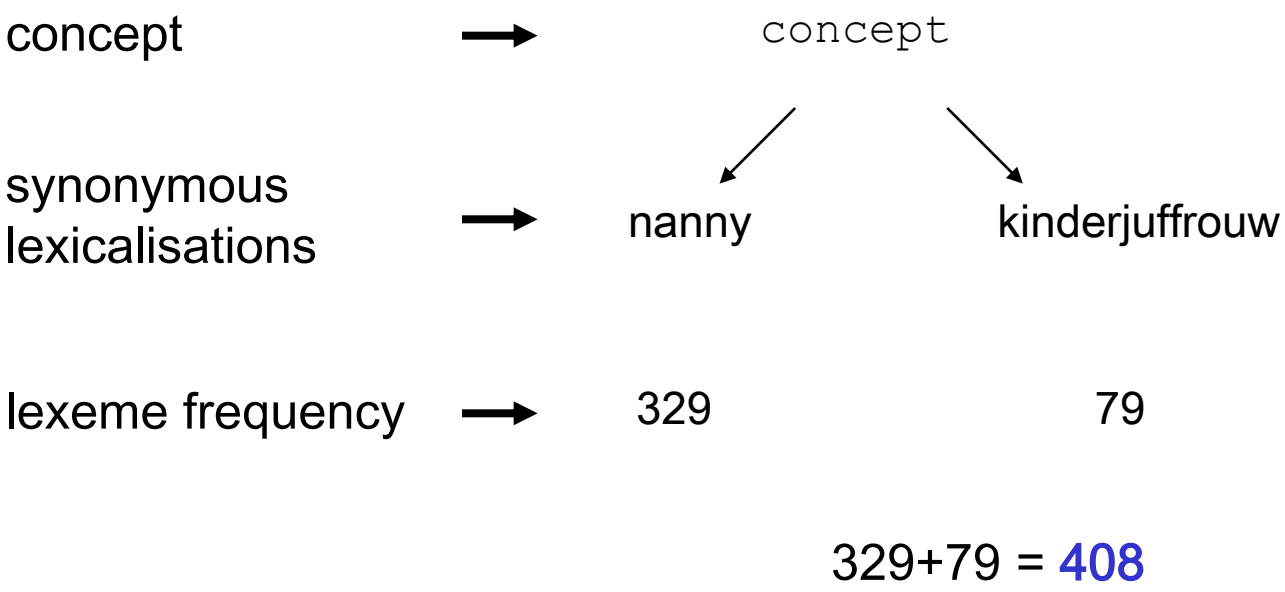
Empirical testing:

- Necessary ingredients
- Case study

1. resistance to borrowing: success measures for 150 English loanwords
2. defining coreness: entrenchment-level
3. competing predictors of success
4. regression analysis
5. results

Entrenchment of the Concept Expressed

→ Corpus frequency of the concept expressed



329+79 = **408**

Entrenchment of the Concept Expressed

→ Corpus frequency of the concept expressed

concept

synonymous lexicalisations

lexeme frequenc

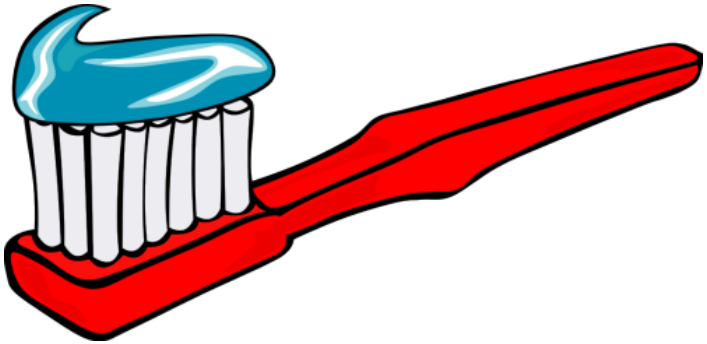
More frequent concepts
→ more frequently activated
→ higher entrenched/core
→ more resistance to borrowing
→ less success for the anglicism

high frequent concepts → low success loanword
low frequent concepts → high success loanword

329+79 = **408**

Entrenchment of the Concept Expressed

BUT:



Entrenchment of the Concept Expressed

Additional measure of entrenchment:

age of the concept at the time the loanword was introduced

older concepts
→ longer activation
→ higher entrenched/core
→ more resistance to borrowing
→ less success for the anglicism

old concepts → low success loanword
young concept → high success loanword

Entrenchment of the Concept Expressed

BUT: careful

old concepts are not necessarily very entrenched (YEOMAN)

More straightforward: concept novelty

Entrenchment of the Concept Expressed

BUT: careful

old concepts are not necessarily very entrenched (YEOMAN)

More straightforward: concept novelty

webmaster: introduced for a new concept

Entrenchment of the Concept Expressed

BUT: careful

old concepts are not necessarily very entrenched (YEOMAN)

More straightforward: concept novelty

webmaster: introduced for a new concept

bull: introduced for an already lexicalized concept

Entrenchment of the Concept Expressed

BUT: careful

old concepts are not necessarily very entrenched (YEOMAN)

More straightforward: concept novelty

webmaster: introduced for a new concept

bull: introduced for an already lexicalized concept

lexicalisations for BULL	introduced in Dutch in
haussier	1864
bull	**1914**
stier	1976

Entrenchment of the Concept Expressed

BUT: careful

old concepts are not necessarily very entrenched (YEOMAN)

More straightforward: concept novelty

webmaster: introduced for a new concept — NECESSARY

bull: introduced for an already lexicalized concept — LUXURY

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

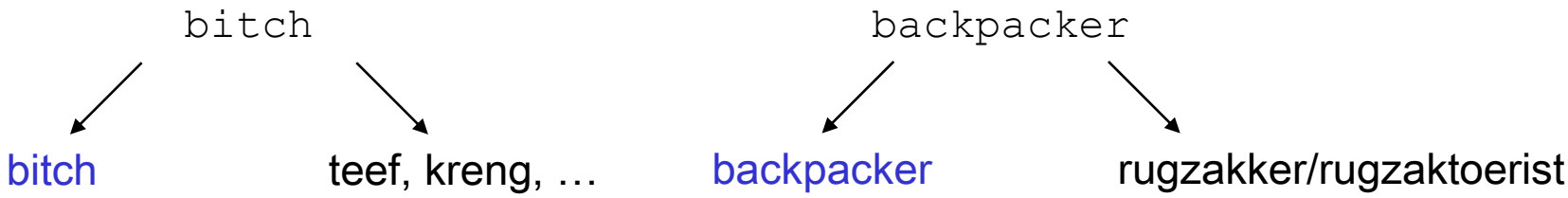
1. resistance to borrowing: success measures for 150 English loanwords
2. defining coreness: entrenchment-level
3. competing predictors of success
4. regression analysis
5. results

Competing Features

- speech economy: shortest yes/no

Competing Features

- speech economy: shortest yes/no
- concept neutrality: yes/no



Competing Features

- speech economy: shortest yes/no
- concept neutrality: yes/no
- age of the loanword: <5, 5-25, >25

Competing Features

- speech economy: shortest yes/no
- concept neutrality: yes/no
- age of the loanword: <5, 5-25, >25
- region, register, diachronic period: BD/ND, QUAL/POP, year

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. resistance to borrowing: success measures for 150 English loanwords
2. defining coreness: entrenchment-level
3. competing predictors of success
4. regression analysis
5. results

Summarizing

Resistance to borrowing

- success of borrowed forms (PRN)

Entrenchment/coreness:

- concept frequency
- concept novelty (new/old)

Other predictors:

- speech economy
- concept neutrality
- age loanword
- region/register/diachronic period

Regression Analyses

Dependent variable: success of the anglicism

- problem with %: heavy tails due to cap at 0 and 1

  → transform to log(odds) (without 0/1-cases)

- in order to include lectal variation: 6 measuring points

One MP per subcorpus: split out for (1) region; (2) register; (3) year

measuring point	freq. *hacker*	conc.freq	angl.perc
hacker BD POP 9902	1000	1099	91%
hacker BD QUAL 9902	1343	1421	95%
hacker BD POP 0305	335	365	92%
hacker BD QUAL 0305	619	646	96%
hacker ND POP 9902	767	833	92%
hacker ND QUAL 9902	578	620	93%

Regression Analyses

Dependent variable: success of the anglicism

- problem with %: heavy tails due to cap at 0 and 1

  → transform to log(odds) (without 0/1-cases)

- in order to include lectal variation: 6 measuring points

One MP per subcorpus: split out for (1) region; (2) register; (3) year

measuring point	freq. *hacker*	conc.freq	angl.perc
hacker BD POP 9902	1000	1099	91%
hacker BD QUAL 9902	1343	1421	95%
hacker BD POP 0305	335	365	92%
hacker BD QUAL 0305	619	646	96%
hacker ND POP 9902	767	833	92%
hacker ND QUAL 9902	578	620	93%

Regression Analyses

Mixed effect model; random variable "lexeme"
needed to take into account multiple measuring points

MODEL FOR ENTIRE DATASET

fixed only	R^2: 34.4%
mixed	reduction Std.Dev random variable: 21.6%

Core vocabulary is resistant to borrowing

Empirical testing:

- Necessary ingredients
- Case study

1. resistance to borrowing: success measures for 150 English loanwords
2. defining coreness: entrenchment-level
3. competing predictors of success
4. regression analysis
5. results

	Estim	Std.Err	Z-val	P	
(Intercept)	6.101	1.089	5.604	0.000	***
concnovelty.existing	-2.976	0.536	-5.555	0.000	***
log(concept frequency)	-0.740	0.146	-5.062	0.000	***
speechecon.shortest	-5.529	1.802	-3.069	0.002	**
log(concfreq) : speechecon.shortest	0.765	0.255	2.998	0.003	**
concnovelty.existing : speechecon.shortest	1.519	0.862	1.763	0.078	.

→ Three predictors are significant
→ Both entrenchment-related predictors

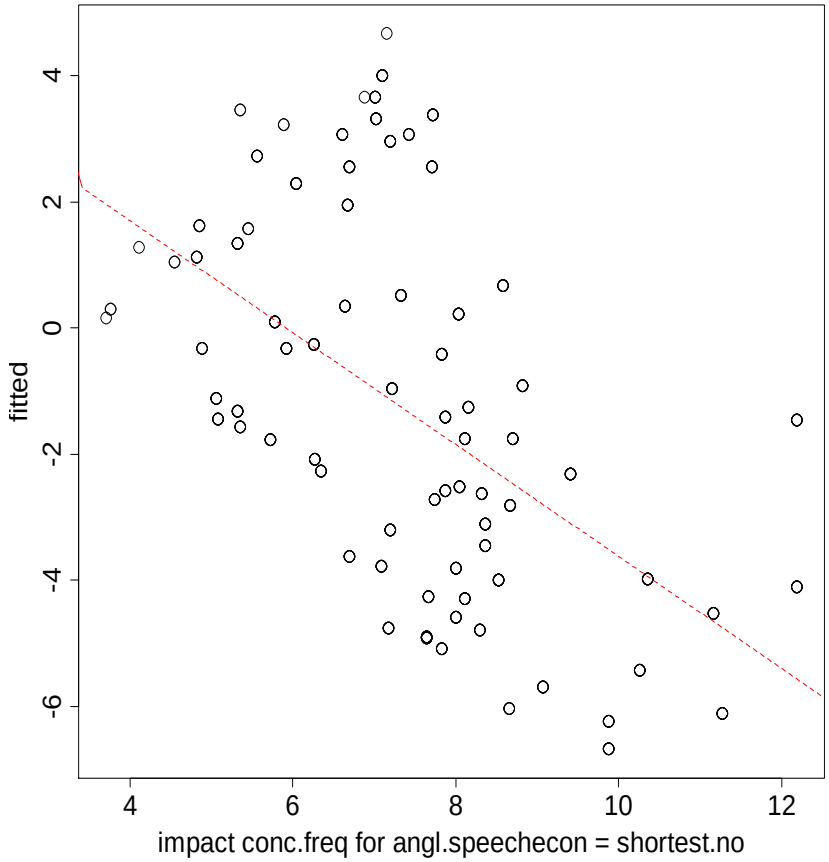
(Intercept)	6.101	1.089	5.604	0.000	***
concnovelty.existing	-2.976	0.536	-5.555	0.000	***
log(concept frequency)	-0.740	0.146	-5.062	0.000	***
speechecon.shortest	-5.529	1.802	-3.069	0.002	**
log(concfreq) : speechecon.shortest	0.765	0.255	2.998	0.003	**
concnovelty.existing : speechecon.shortest	1.519	0.862	1.763	0.078	.

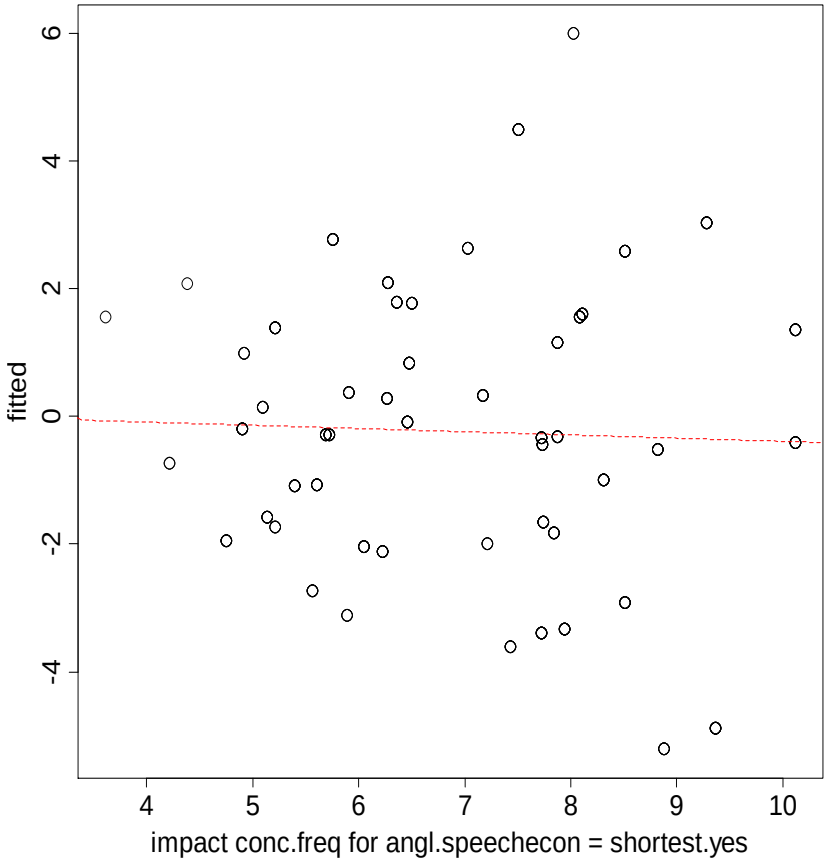
	Estim	Std.Err	Z-val	P	
(Intercept)	6.101	1.089	5.604	0.000	***
concnovelty.existing	-2.976	0.536	-5.555	0.000	***
log(concept frequency)	-0.740	0.146	-5.062	0.000	***
speechecon.shortest	-5.529	1.802	-3.069	0.002	**
log(concfreq) : speechecon.shortest	0.765	0.255	2.998	0.003	**
concnovelty.existing : speechecon.shortest	1.519	0.862	1.763	0.078	.

interactions between all three selected predictors

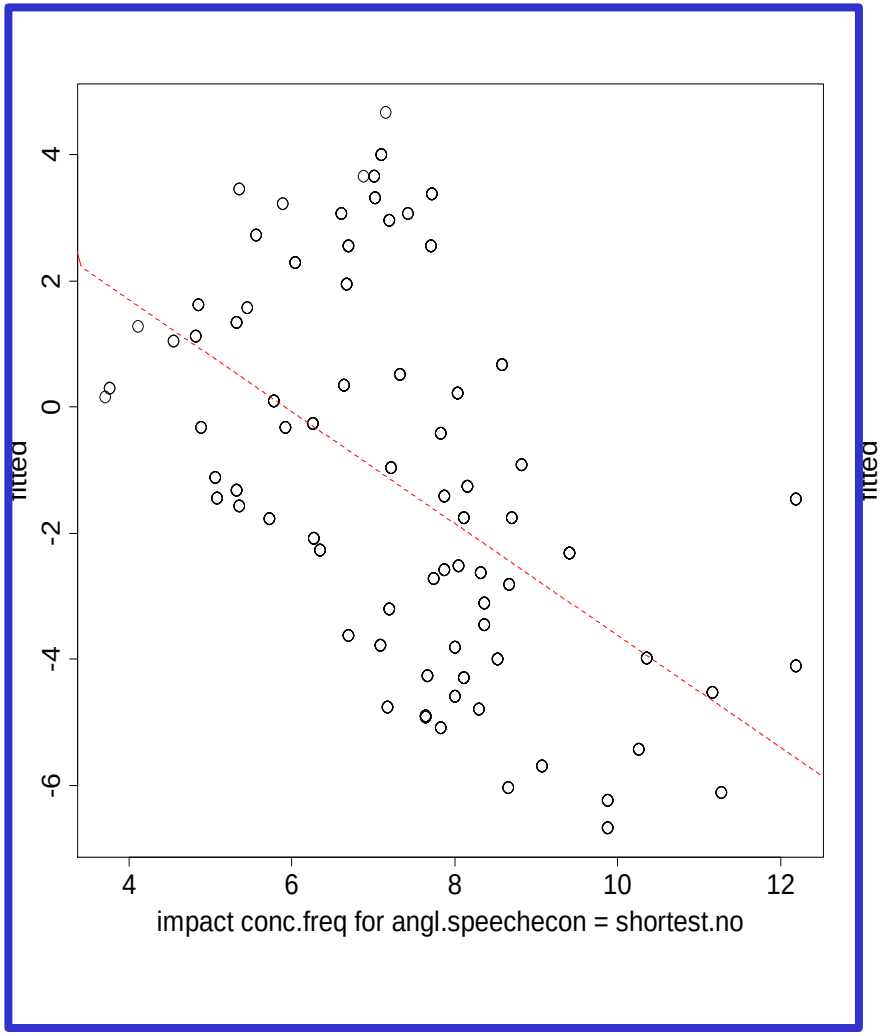
→ interaction plots

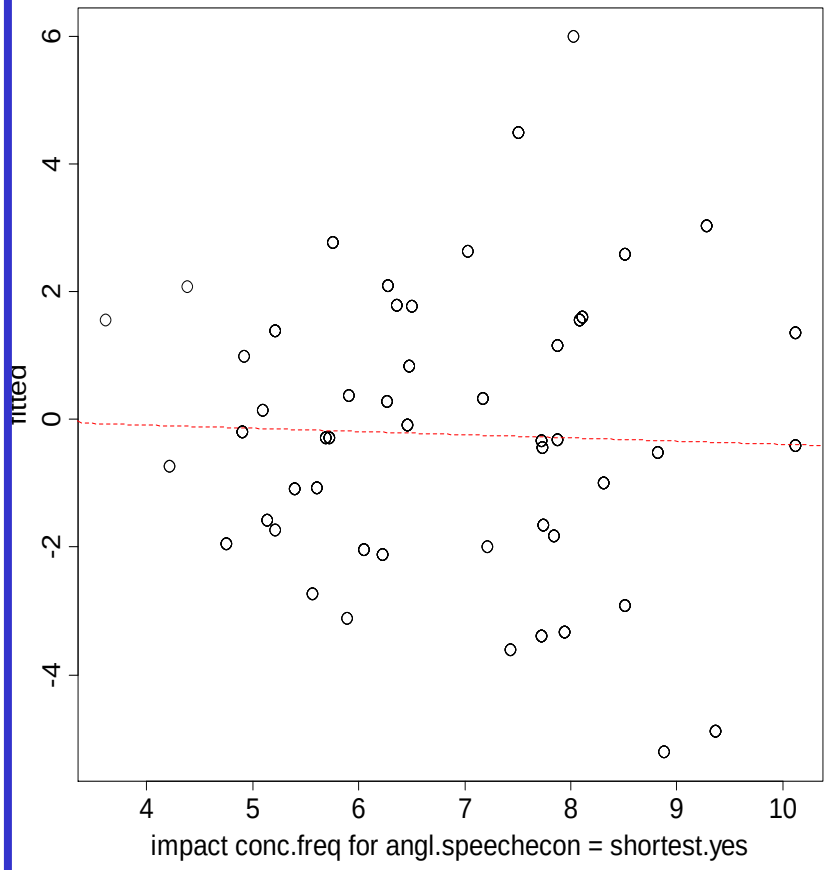
	Estim	Std.Err	Z-val	P	
(Intercept)	6.101	1.089	5.604	0.000	***
concnovelty.existing	-2.976	0.536	-5.555	0.000	***
log(concept frequency)	-0.740	0.146	-5.062	0.000	***
speechecon.shortest	-5.529	1.802	-3.069	0.002	**
log(concfreq) : speechecon.shortest	0.765	0.255	2.998	0.003	**
concnovelty.existing : speechecon.shortest	1.519	0.862	1.763	0.078	.


fitted
4
2
0
-2
-4
-6
4
6
8
10
12
impact conc.freq for angl.speechecon = shortest.no

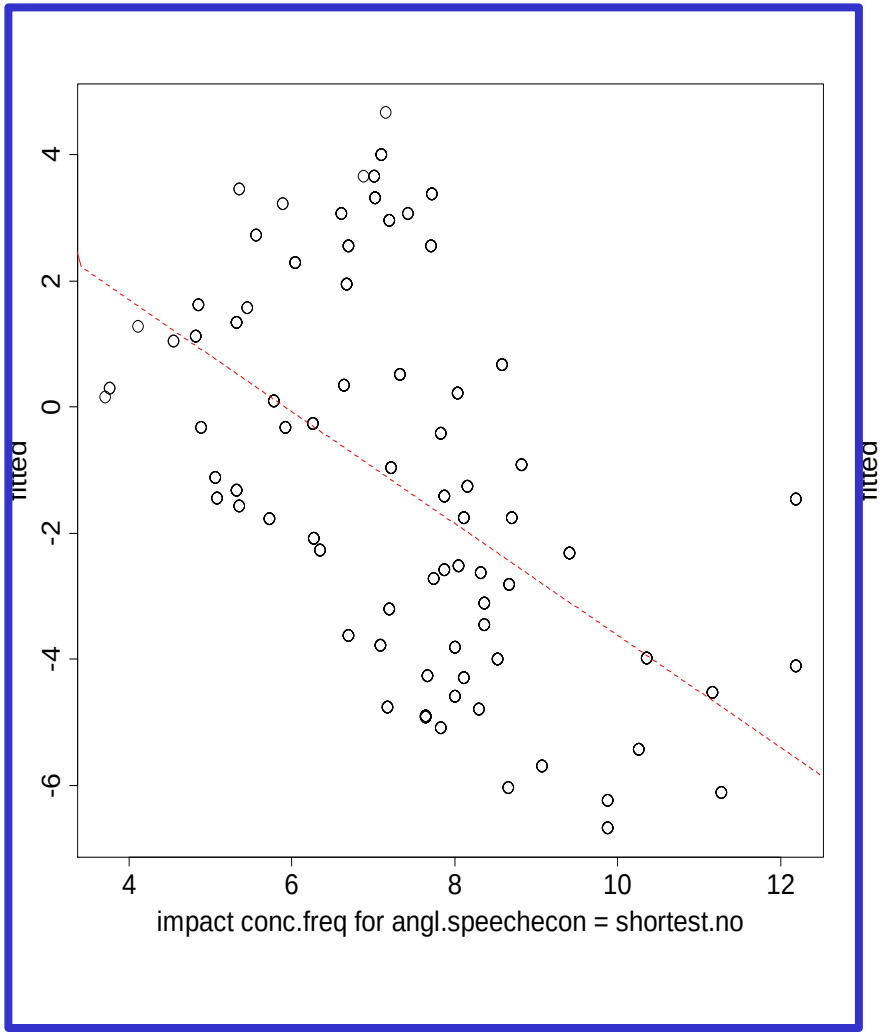

fitted
6
4
2
0
-2
-4
4
5
6
7
8
9
10
impact conc.freq for angl.speechecon = shortest.yes

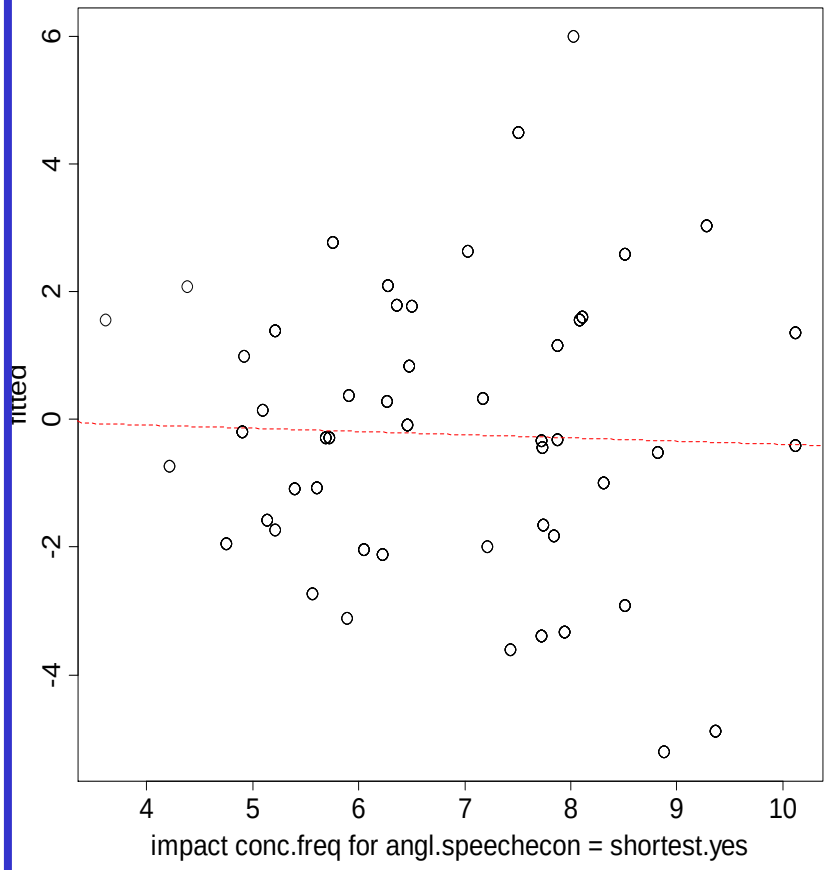
Anglicism is not the shortest equivalent (ghostwriter vs. negre)



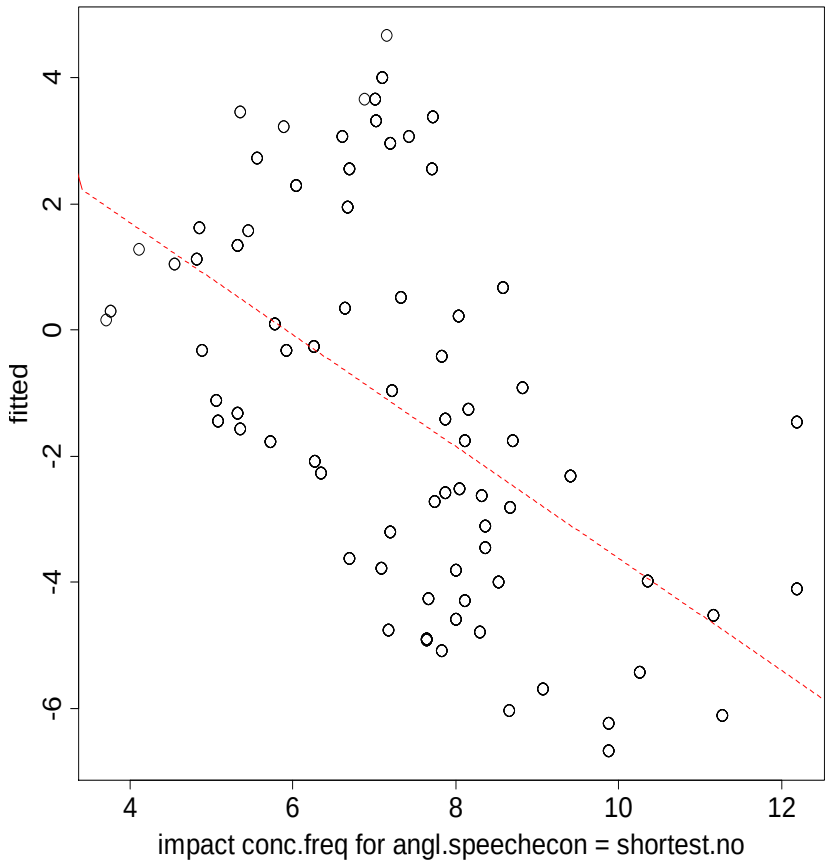


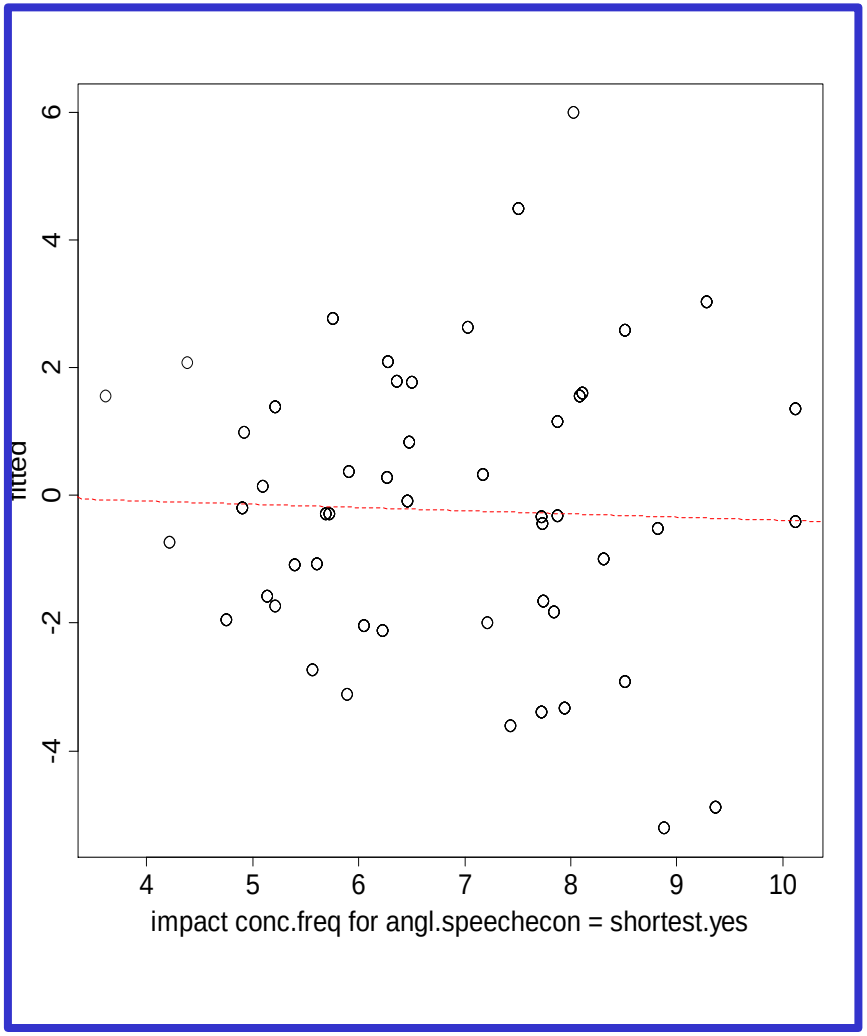
Higher concept frequency → lower success loanword



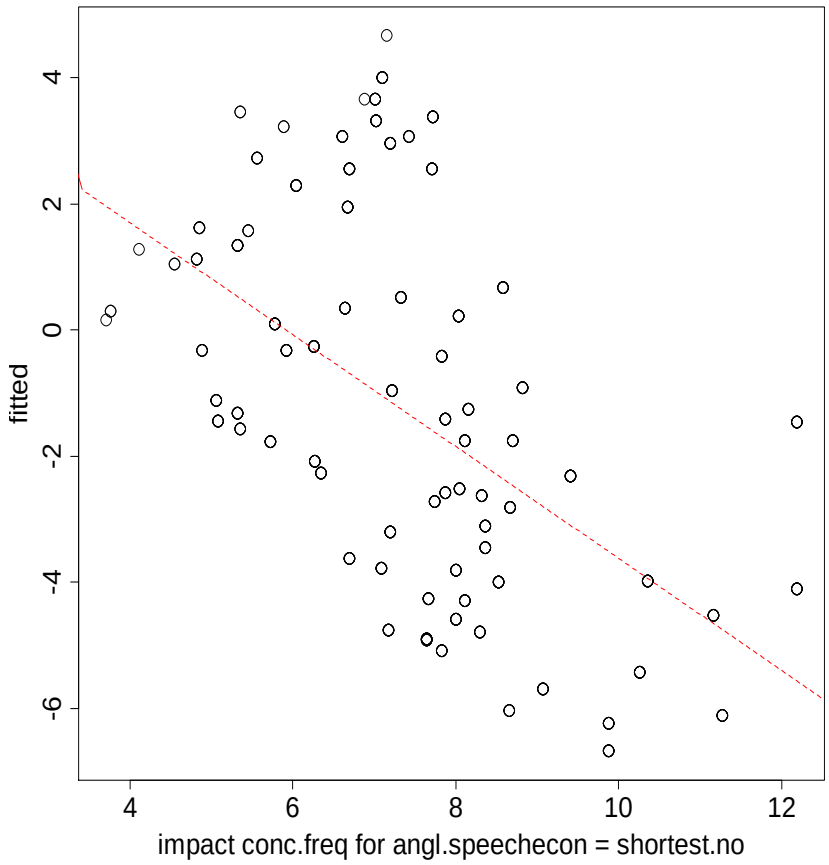


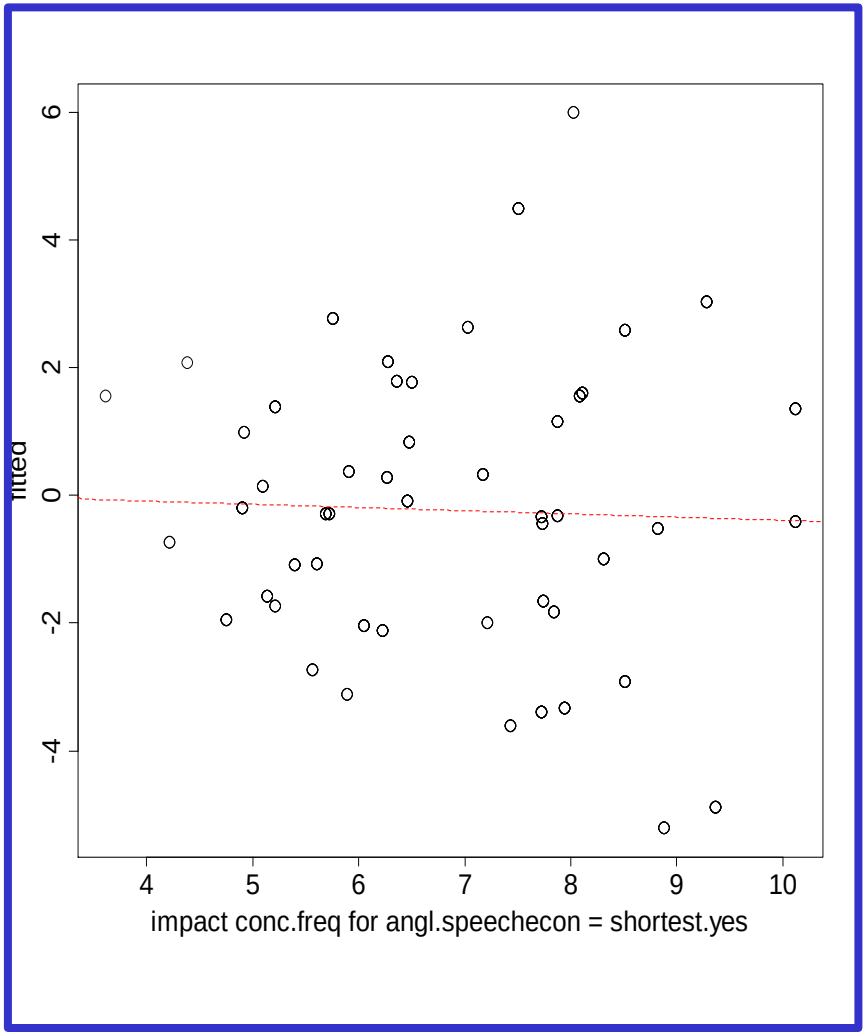
Anglicism is the shortest equivalent (bellboy vs. piccolo)



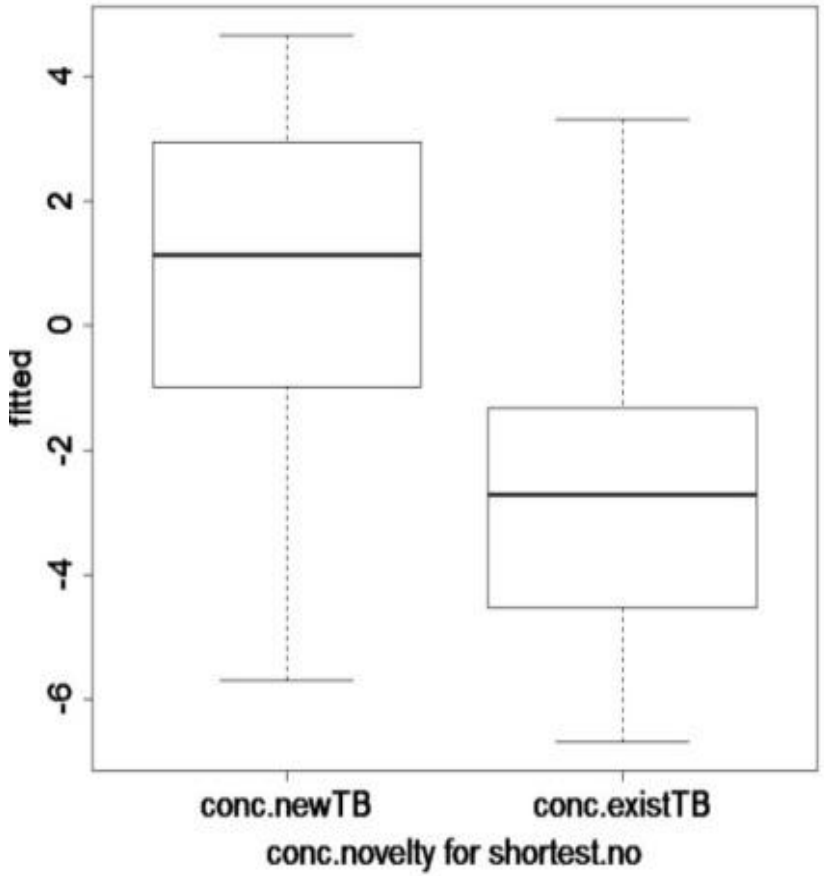


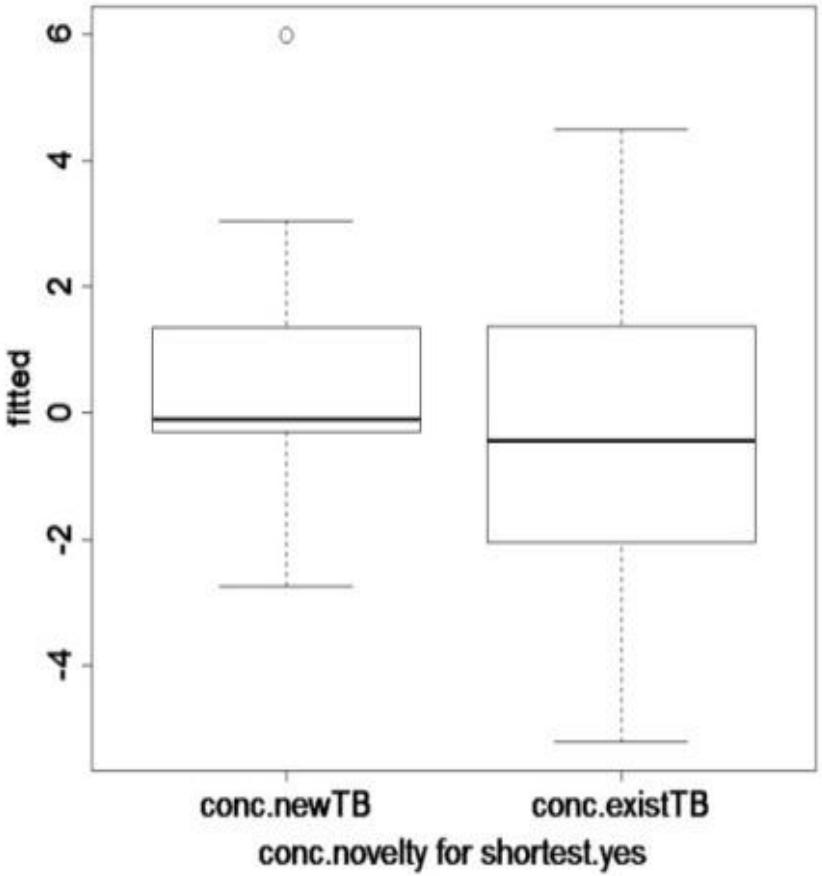
No effect for concept frequency



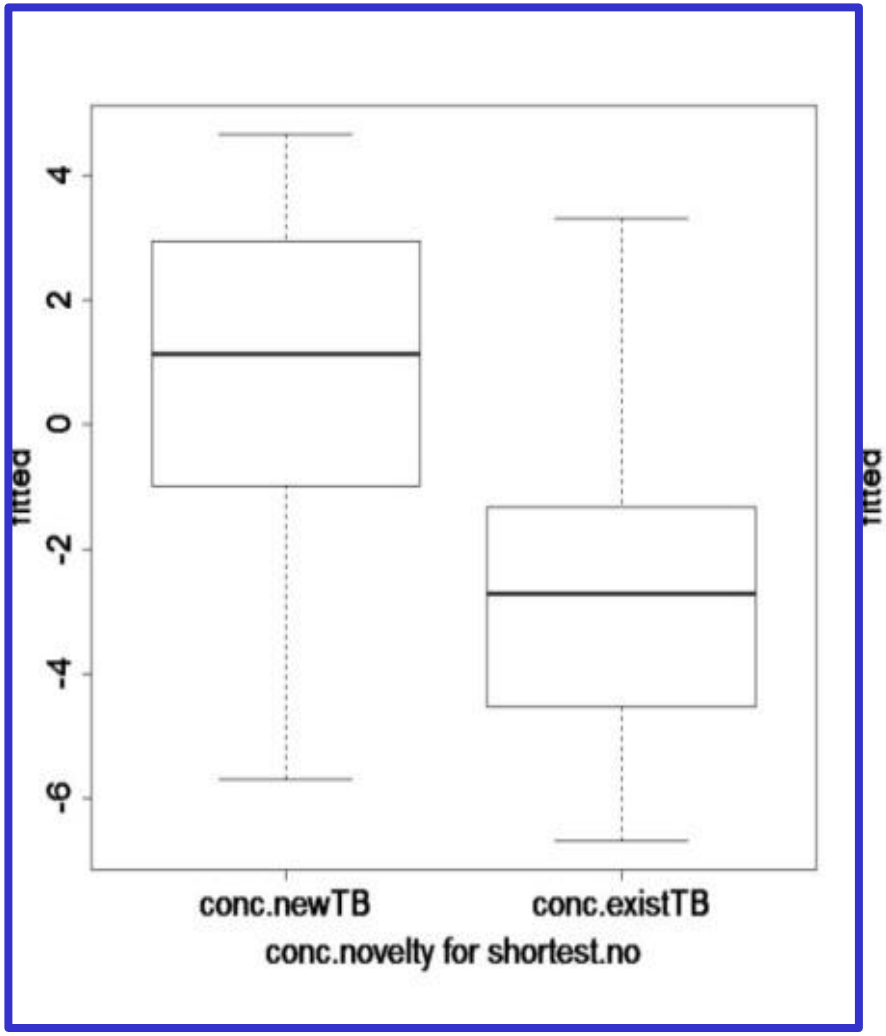


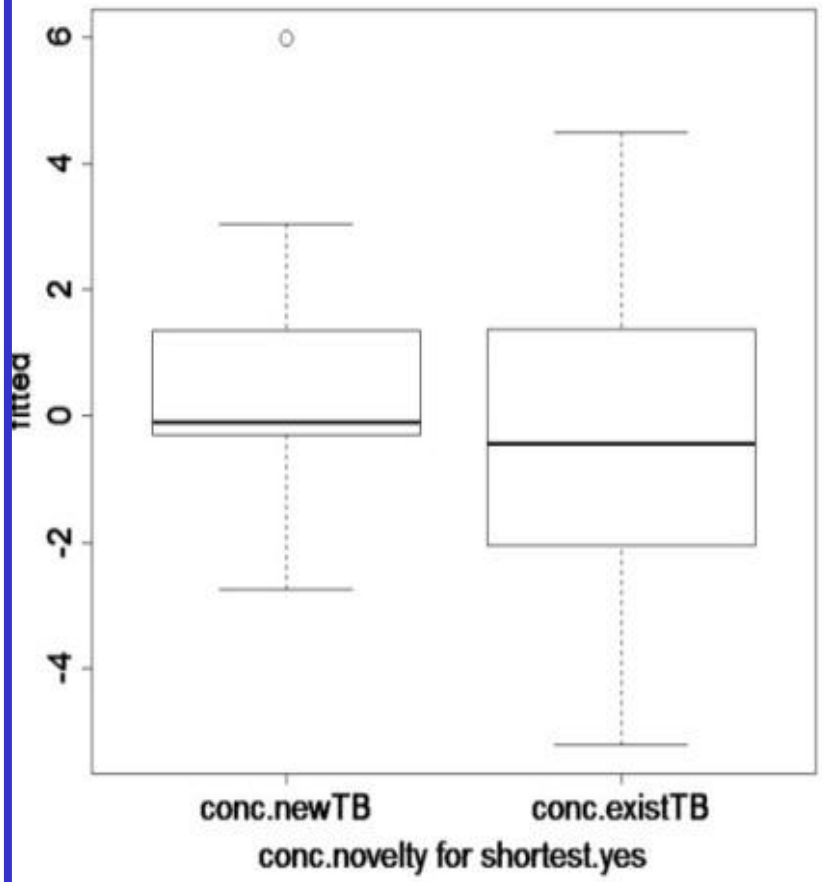
	Estim	Std.Err	Z-val	P	
(Intercept)	6.101	1.089	5.604	0.000	***
concnovelty.existing	-2.976	0.536	-5.555	0.000	***
log(concept frequency)	-0.740	0.146	-5.062	0.000	***
speechecon.shortest	-5.529	1.802	-3.069	0.002	**
log(concfreq) : speechecon.shortest	0.765	0.255	2.998	0.003	**
concnovelty.existing : speechecon.shortest	1.519	0.862	1.763	0.078	.


fitted
4
2
0
-2
-4
-6
conc.newTB
conc.existTB
conc.novelty for shortest.no

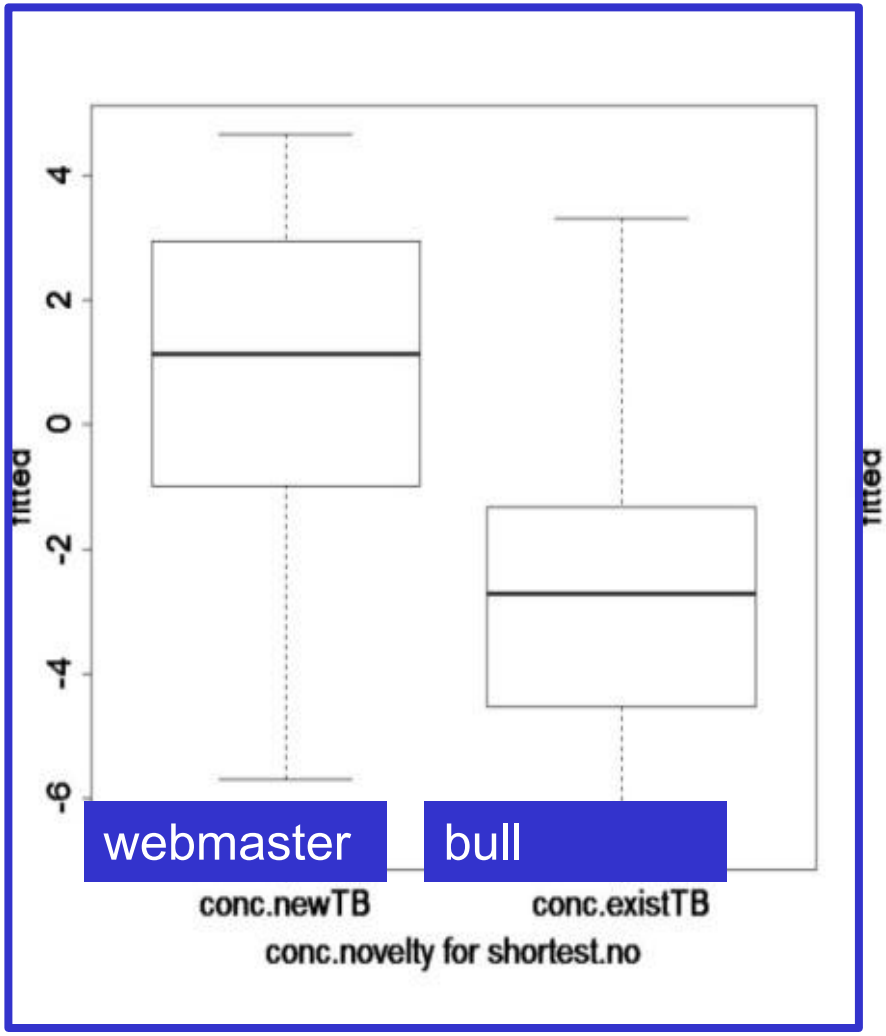

fitted
6
4
2
0
-2
-4
conc.newTB
conc.existTB
conc.novelty for shortest.yes

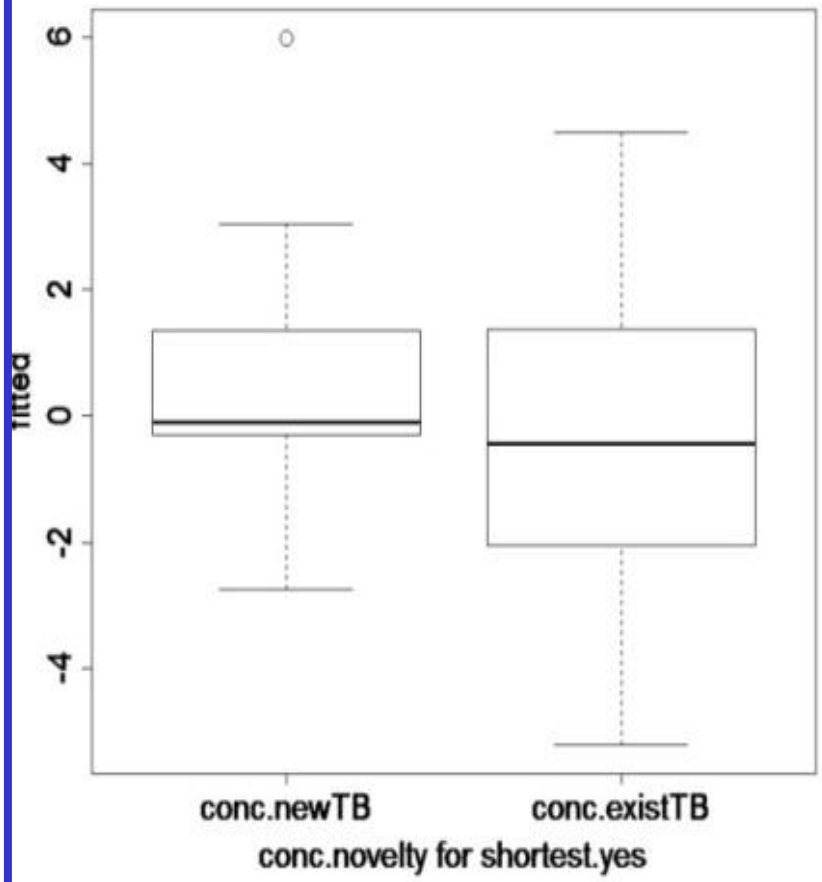
Anglicism is not the shortest equivalent (ghostwriter vs. negre)



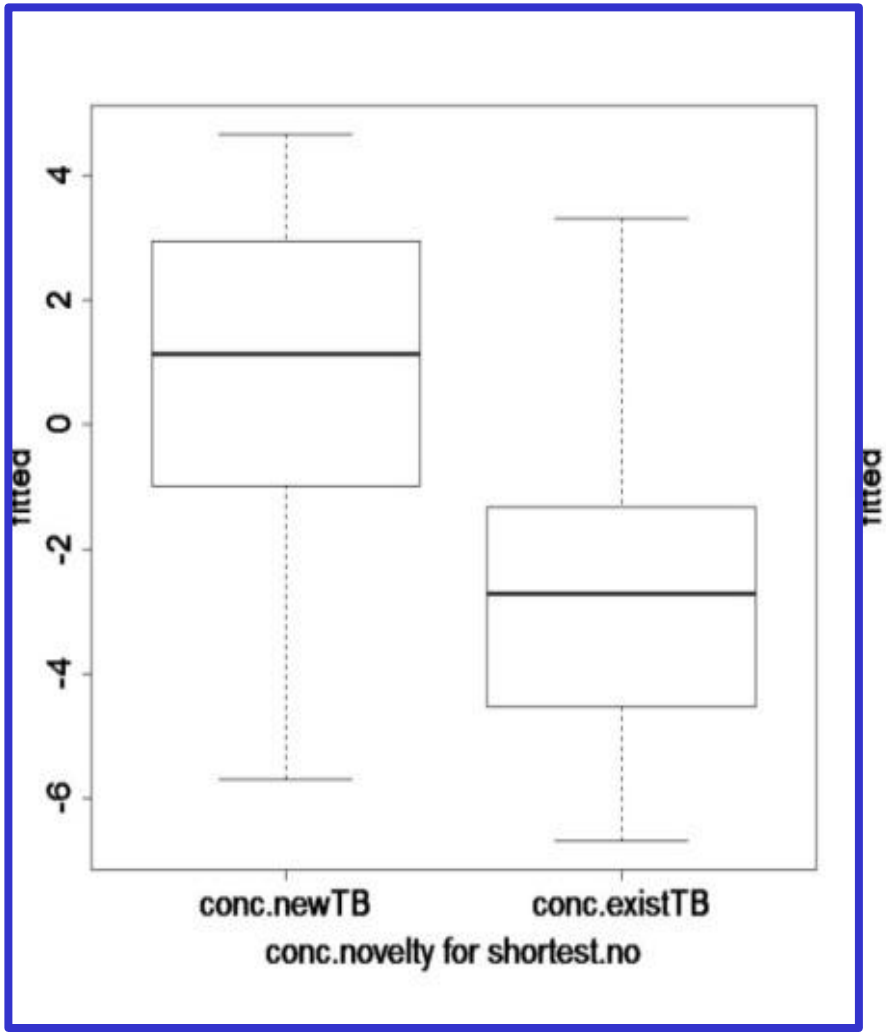


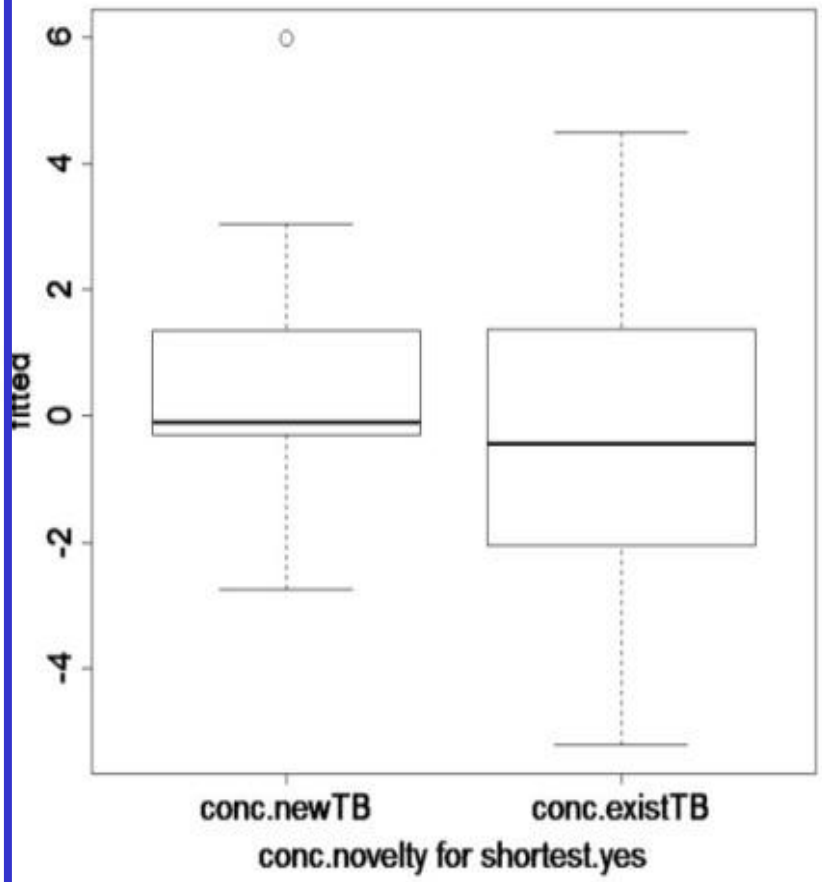
Anglicism is not the shortest equivalent (ghostwriter vs. negre)



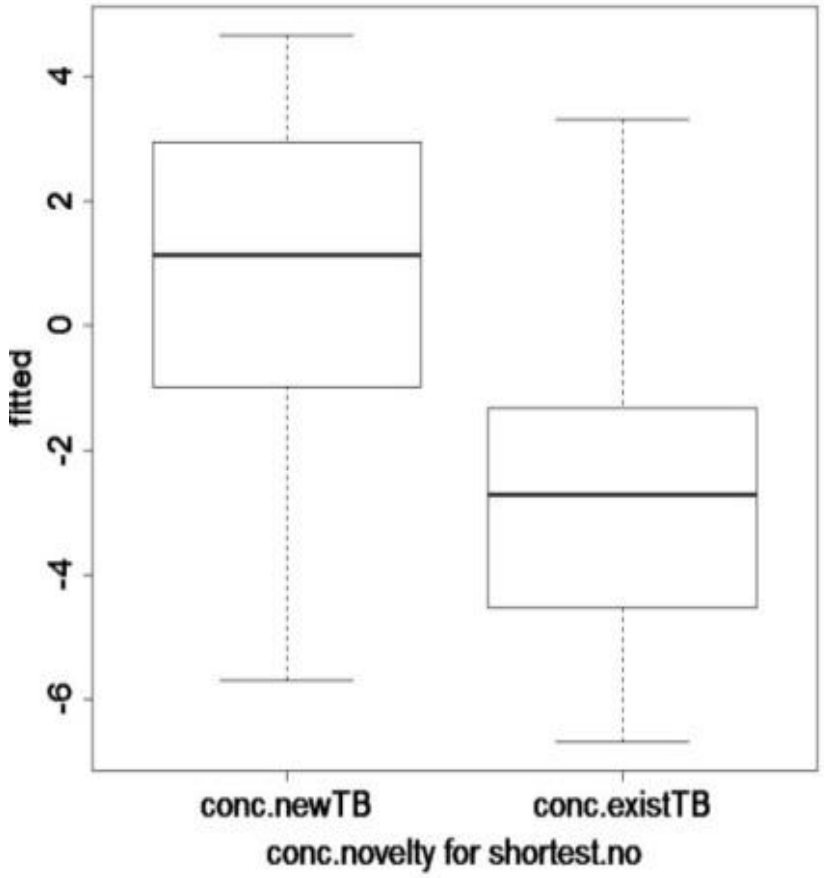


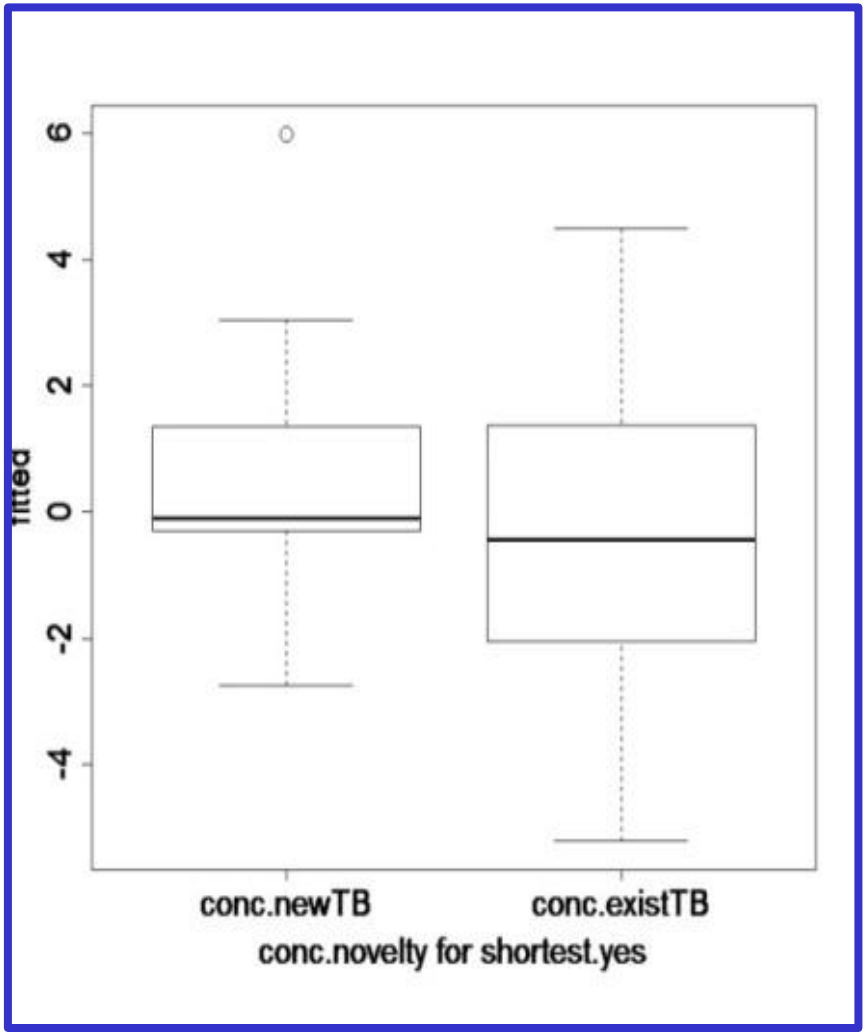
More success when filling lexical gap



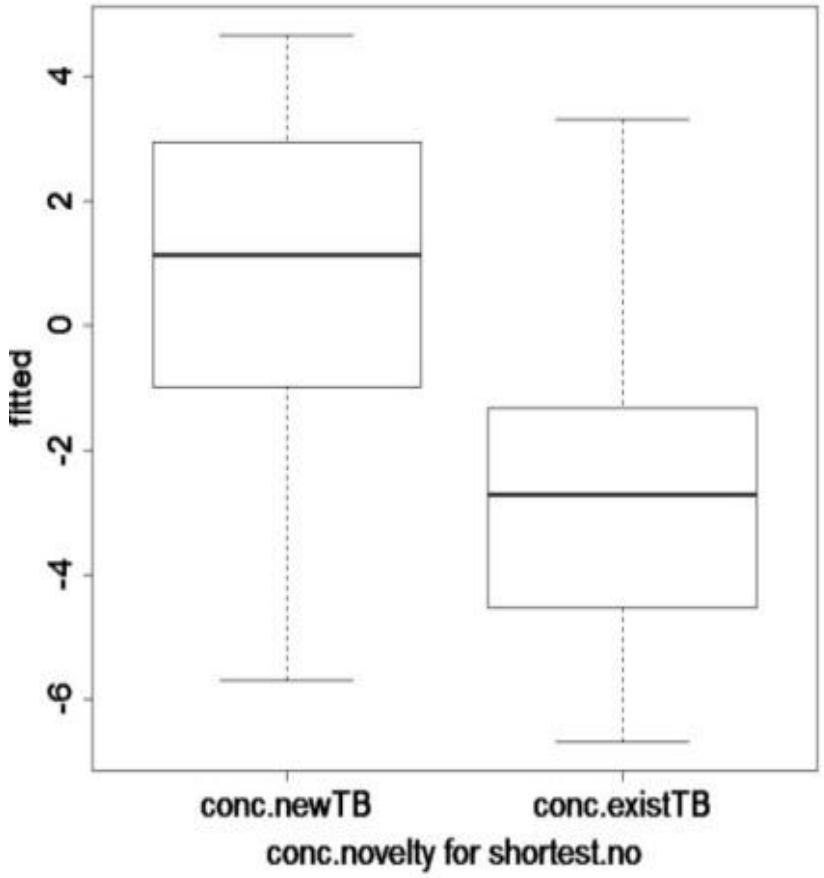


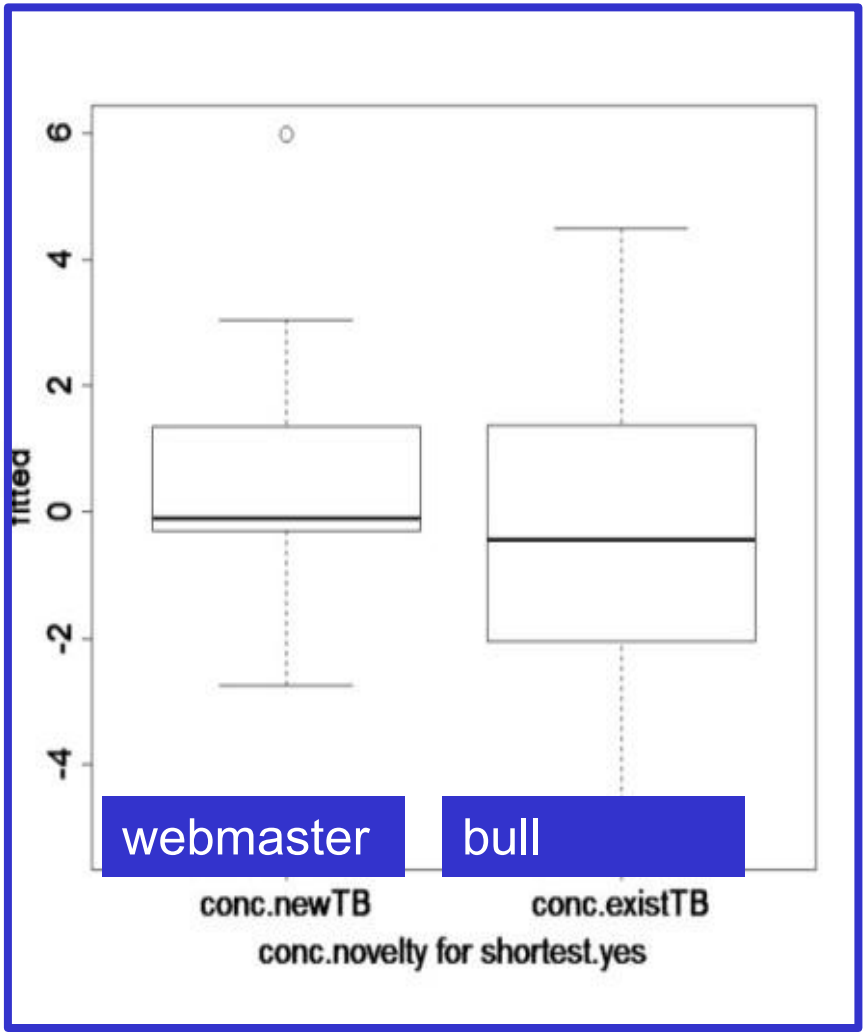
Anglicism is the shortest equivalent (bellboy vs. piccolo)



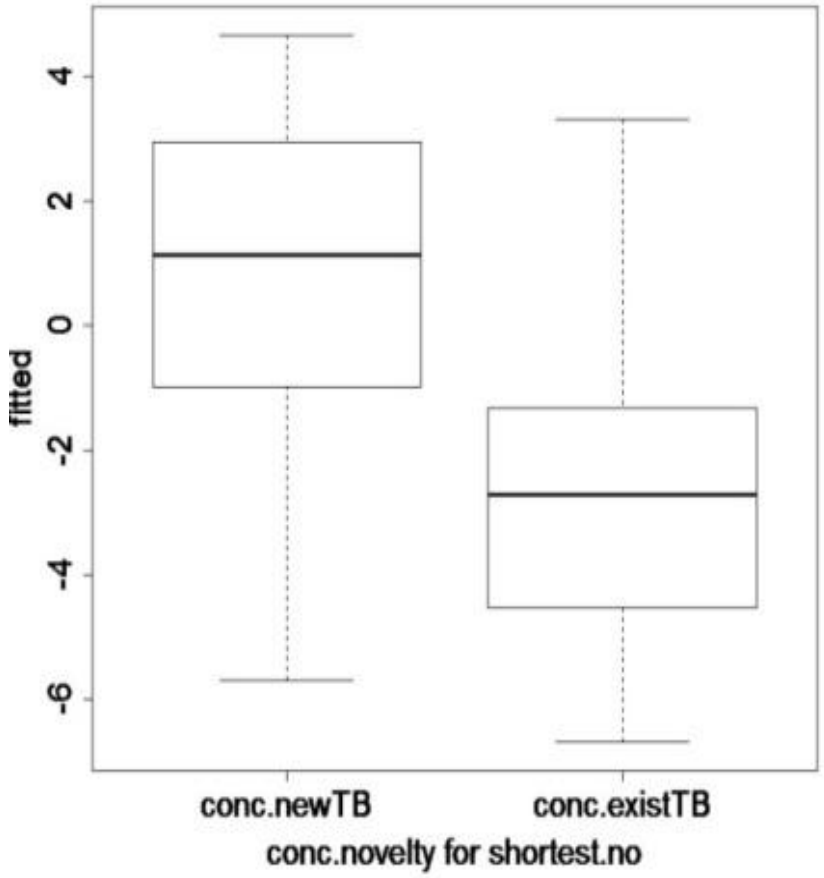


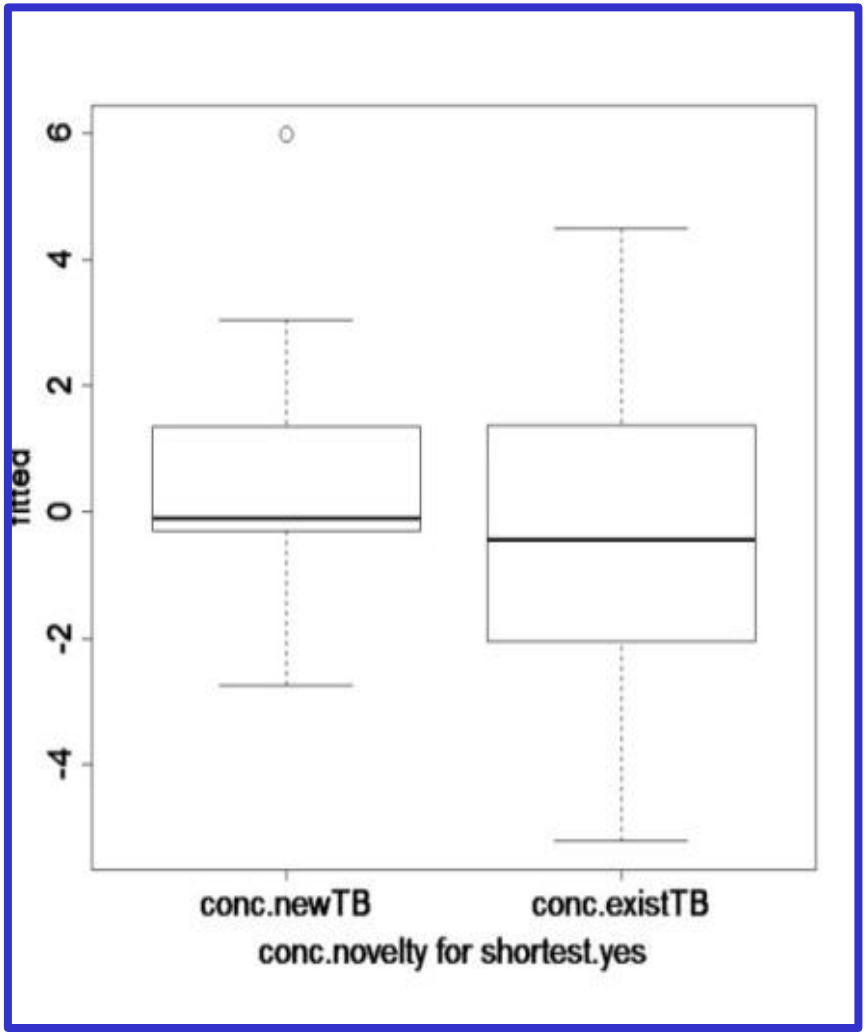
Anglicism is the shortest equivalent (bellboy vs. piccolo)





No effect for concept novelty





Conclusions

Conclusions

Results

- strongest effect for the entrenchment-based predictors
- neutralizing effect for speech economy

Conclusions

Results

- strongest effect for the entrenchment-based predictors
- neutralizing effect for speech economy

Methodology

- linking coreness to entrenchment
- providing an onomasiological measure for resistance to borrowing
- using inferential statistics to reveal the link between both

Conclusions

Results

- strongest effect for the entrenchment-based predictors
- neutralizing effect for speech economy

Methodology

- linking coreness to entrenchment
- providing an onomasiological measure for resistance to borrowing
- using inferential statistics to reveal the link between both

Attenuation

- proof of concept
- applicability to comparative historical linguistics?

For more information:

http://wwwling.arts.kuleuven.be/qlvl

eline.zenner@arts.kuleuven.be