Measuring and comparing the success of loanwords

An onomasiological approach to the use of anglicisms in Dutch

Eline Zenner (FWO Flanders),
Dirk Geeraerts & Dirk Speelman

University of Leuven
RU Quantitative Lexicology and Variational Linguistics

An onomasiological perspective on borrowing

Background

Case Study: English Person Reference Nouns in Dutch

- method
  - building profiles
  - explaining the variation: possibly influential features
  - multivariate approach
- results

Conclusions and Prospects

An onomasiological perspective on borrowing

Background

Case Study: English Person Reference Nouns in Dutch

- method
  - building profiles
  - explaining the variation: possibly influential features
  - multivariate approach
- results

Conclusions and Prospects

Lexical borrowing

attitudes/
normativity

adaptation/
categorization

borrowability/
success

Borrowability

Developed in typology and in research on weak contact situations:

- Likelihood of borrowing
- Success of loanwords

Borrowability

Developed in typology and in research on weak contact situations:

- Likelihood of borrowing

  chance item A is transferred from language 1 to language 2?

  type-based

  1. clines based on POS
  2. core-vocabulary
  3. macro-sociolx: scales of receptivity of languages

  big advances in typology: LTP (Haspelmath & Tadmor)

- Success of loanwords

  once borrowed, how successful will a loanword be?

How successful will a loanword be?

Mainly dealt with in traditional loanword research (e.g. anglicisms):

- success of loanword = nr. of attestations (token counts)
- small corpora (manual extraction)
- variation in success linked to structural features

computer: 23 516

bestseller: 2280

baseball: 113

How successful will a loanword be?

Mainly dealt with in traditional loanword research (e.g. anglicisms):

- success of loanword = nr. of attestations (token counts)
- small corpora (manual extraction)
- variation in success linked to structural features

problematic:

topic specificity (Speelman *et al.* 2003)

→ set-external proof (Van Hout & Muysken 1994)

→concept-based: native alternatives
(Rhode et al. 1999, Winter-Froemel 2010, Gzrega 2003)

How successful will a loanword be?

Mainly dealt with in traditional loanword research (e.g. anglicisms):

- success of loanword = nr. of attestations (token counts)
- small corpora (manual extraction)
- variation in success linked to structural features

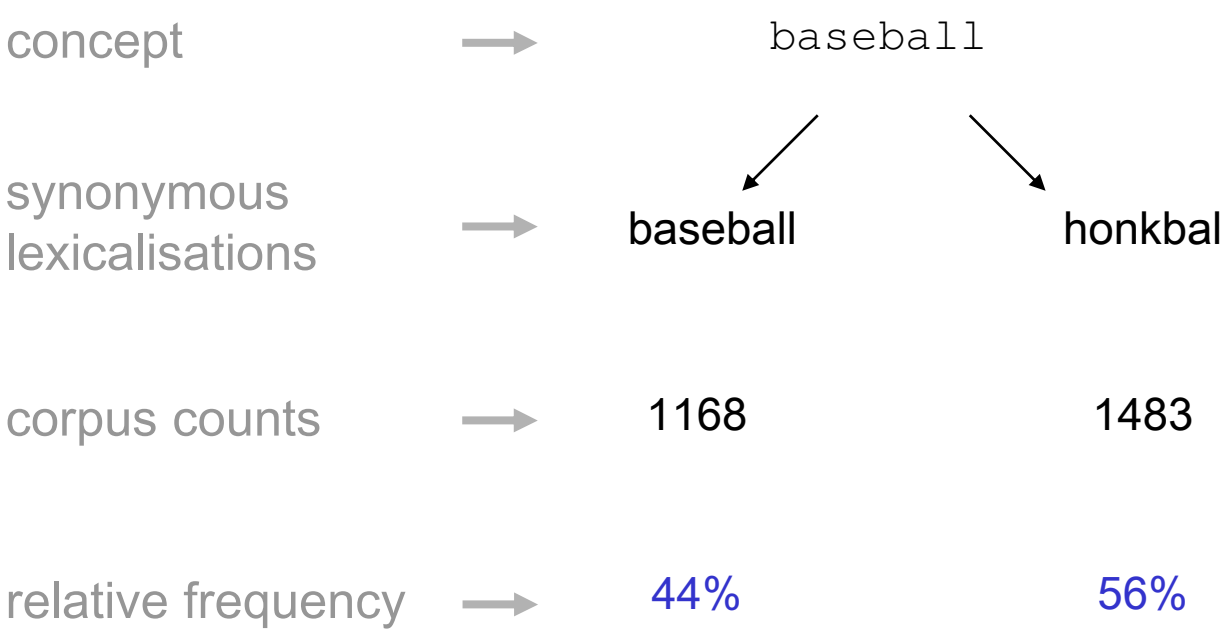
needed:

attention for synonymy in loanword research

- onomasiological approach
- adaptation of profile-based method (Geeraerts *et al.* 1999)

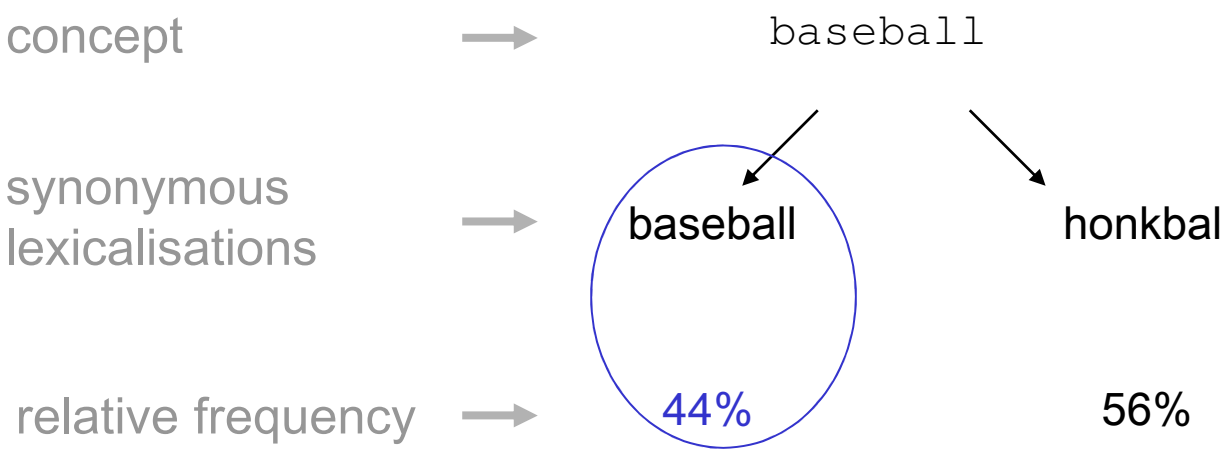
Profile-based method:

Corpus-based comparison of success of lexicalizations



Profile-based method:

Corpus-based comparison of success of lexicalizations



zoom in on the anglicism: **define the success** of an anglicism as the relative preference for the anglicism vis-à-vis existing **synonymous expressions**

Profile-based method

Compare success rates:

→ what features explain variation in the success-rate of a set of anglicisms?

An onomasiological perspective on borrowing

Background

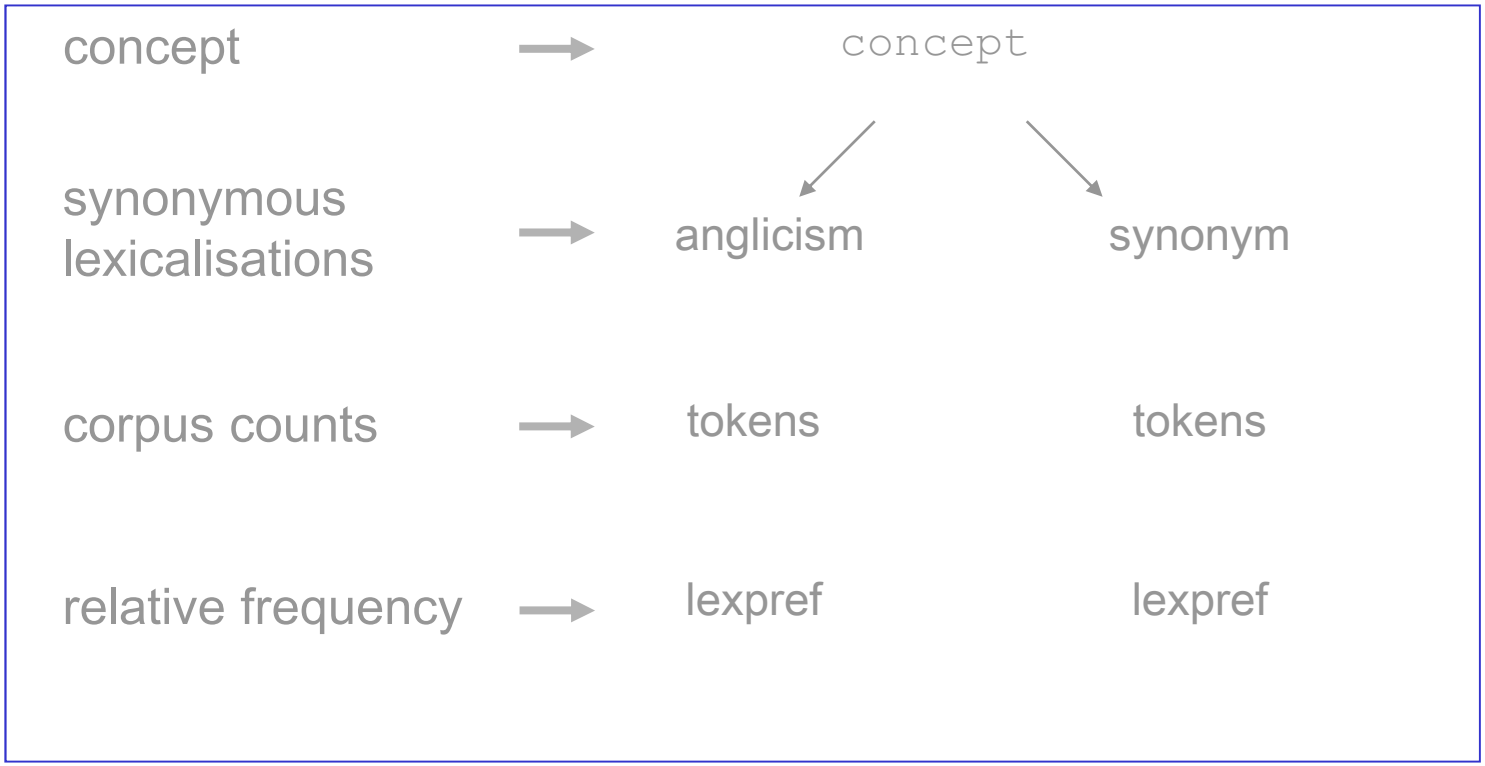
Case Study: English Person Reference Nouns in Dutch

- method
  - building profiles & measuring success
  - explaining the variation: possibly influential features
  - multivariate approach
- results

Conclusions and Prospects

Step 1:

Inherent corpus-based approach: select corpus



presenting the corpus

When tracking down variation: need of stable corpus with reliable amount of tokens:

Two Dutch newspaper corpora (parsed, lemmatised)

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

presenting the corpus

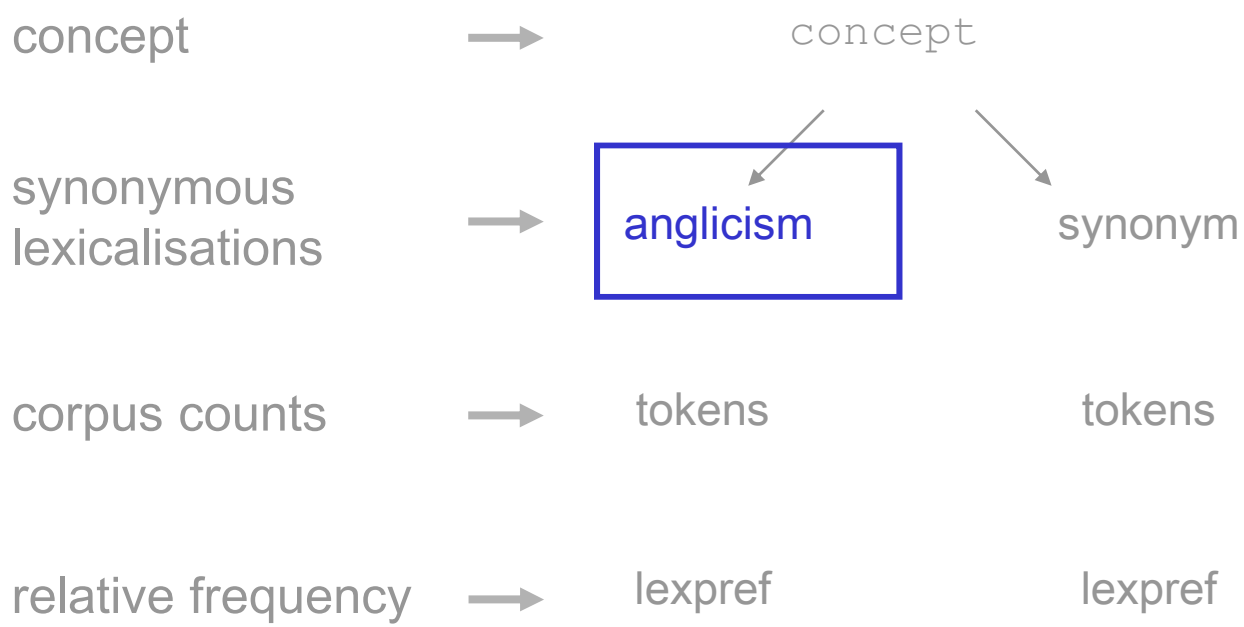
When tracking down variation: need of stable corpus with reliable amount of tokens:

Two Dutch ne sed)

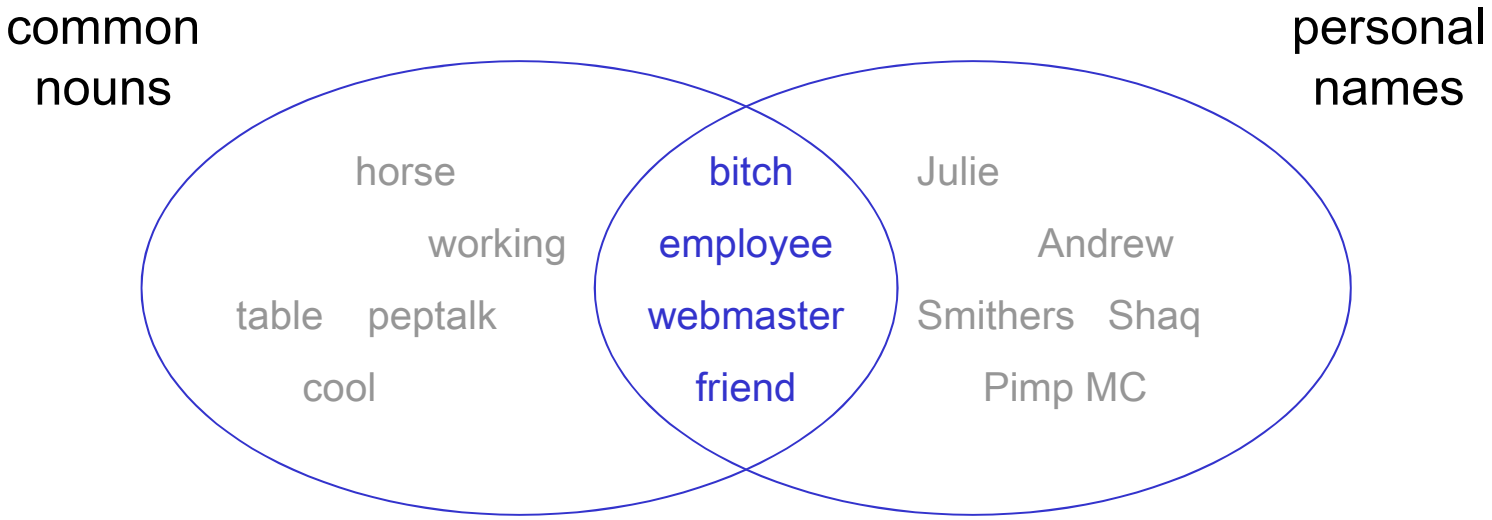
- TwNC 300 million words
- LeNC 1.3 billion words

Step 2:

Look for a set of anglicisms



selecting English PRN used in Dutch

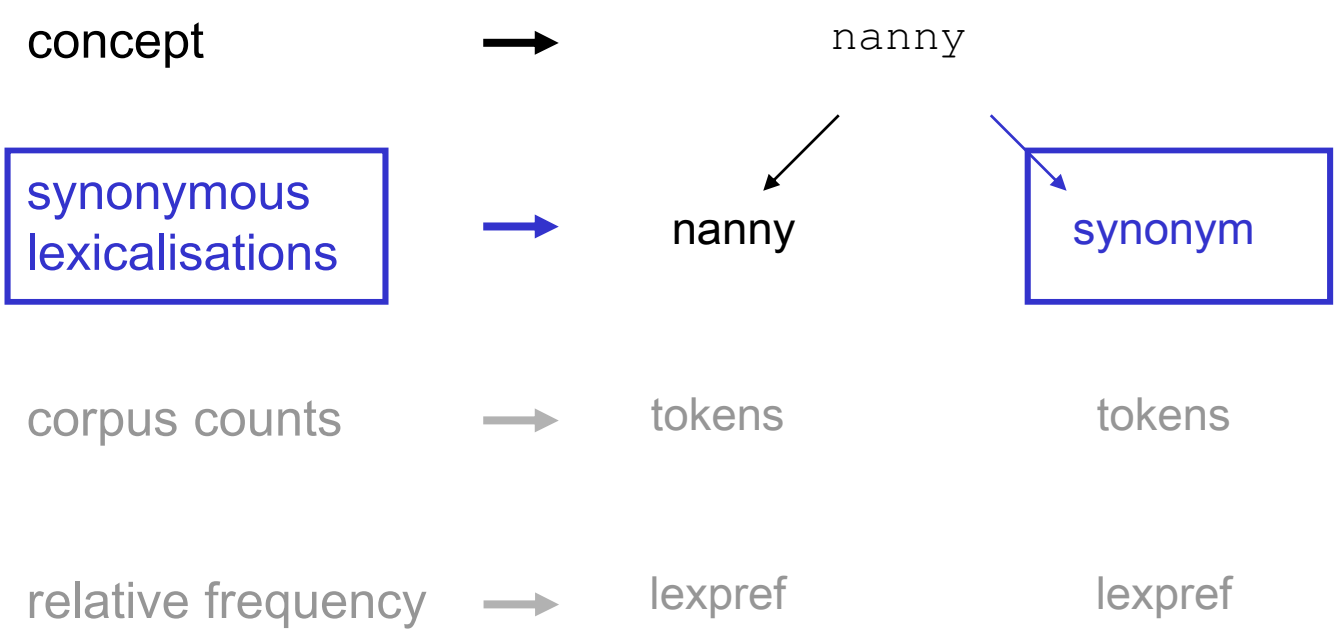


Selection of approximately 200 English PRN in Dutch:

- Lexicographical sources
- Automatic matching of hyponyms of “person” in WordNet with Dutch tokfreqlist

Step 3:

Look for synonymous expressions



defining synonymy

Pre-supposition: synonymy exists

Coarse-grained (Edmonds & Hirst 2002)

denotational synonyms, little attention for fine-grained stylistic differences or connotational nuances

Restrictive approach (Geeraerts 2010)

focus on items with maximal equivalence and interchangeability; no comprehensive study in which all tokens are scrutinized

→ no fuzzy concepts with vague items

→ stable concepts

identifying synonyms

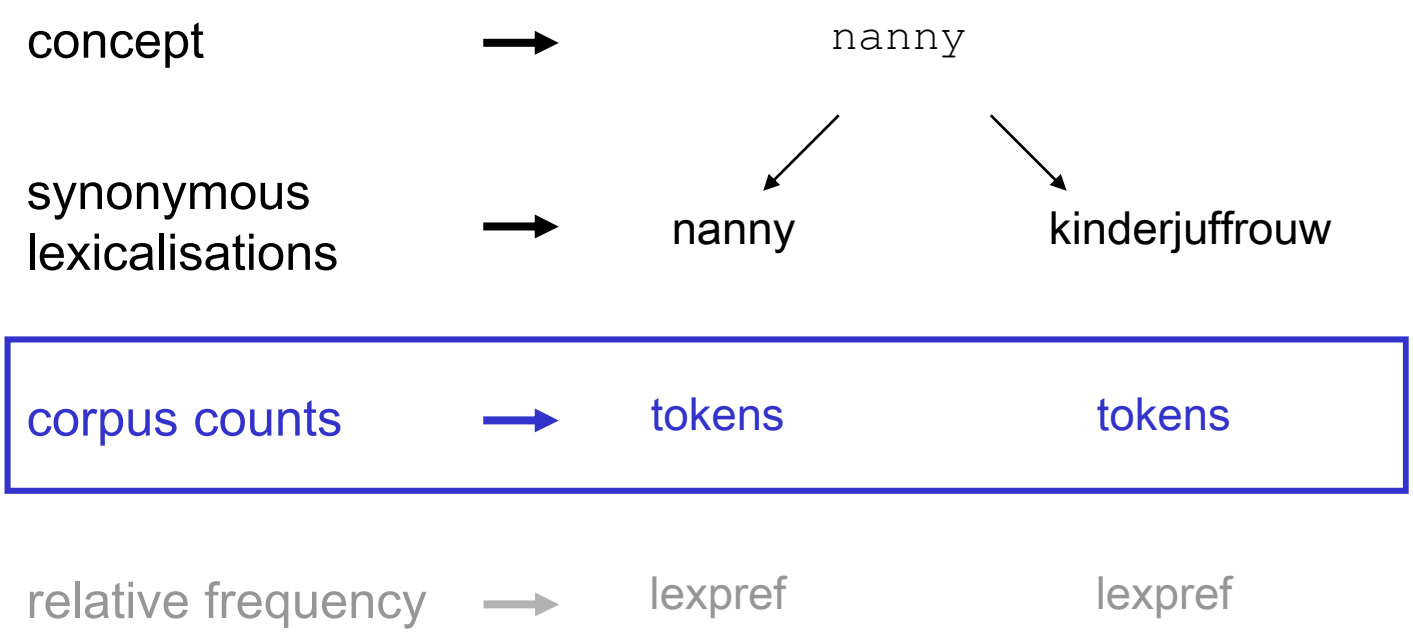
Looking for synonymys:

→ no blind trust in lexicography: combining sources

- start of with 10 lexicographical sources
- complement with results from WSM
- checking:
  - descriptive dictionaries
  - encyclopaedic information
  - experts

Step 4:

Retrieval of tokens from corpus



automatic retrieval tokens

Excluded Tokens (noise)

- Proper names (*Philips Consumer Communications*)
- Lexicalized Compounds (*management consultant*)
- Longer stretches of English (*he's such a fool*)
- Profiles exclusively containing English items (*spammer)*

Polysemy: stable concepts?

- Checked manually: polysemous items with reasonable frequency (*freak, chicken*)
- Excluded: concepts containing high-frequent polysemous lexicalisations or vague items (*buddy – maatje*)

automatic retrieval tokens

Excluded Tokens (noise)

- Proper names (*Philips Consumer Communications*)
- Lexicalized Compounds (*management consultant*)
- Longer stretc
- Profiles exclusively containing English items (*spammer)*

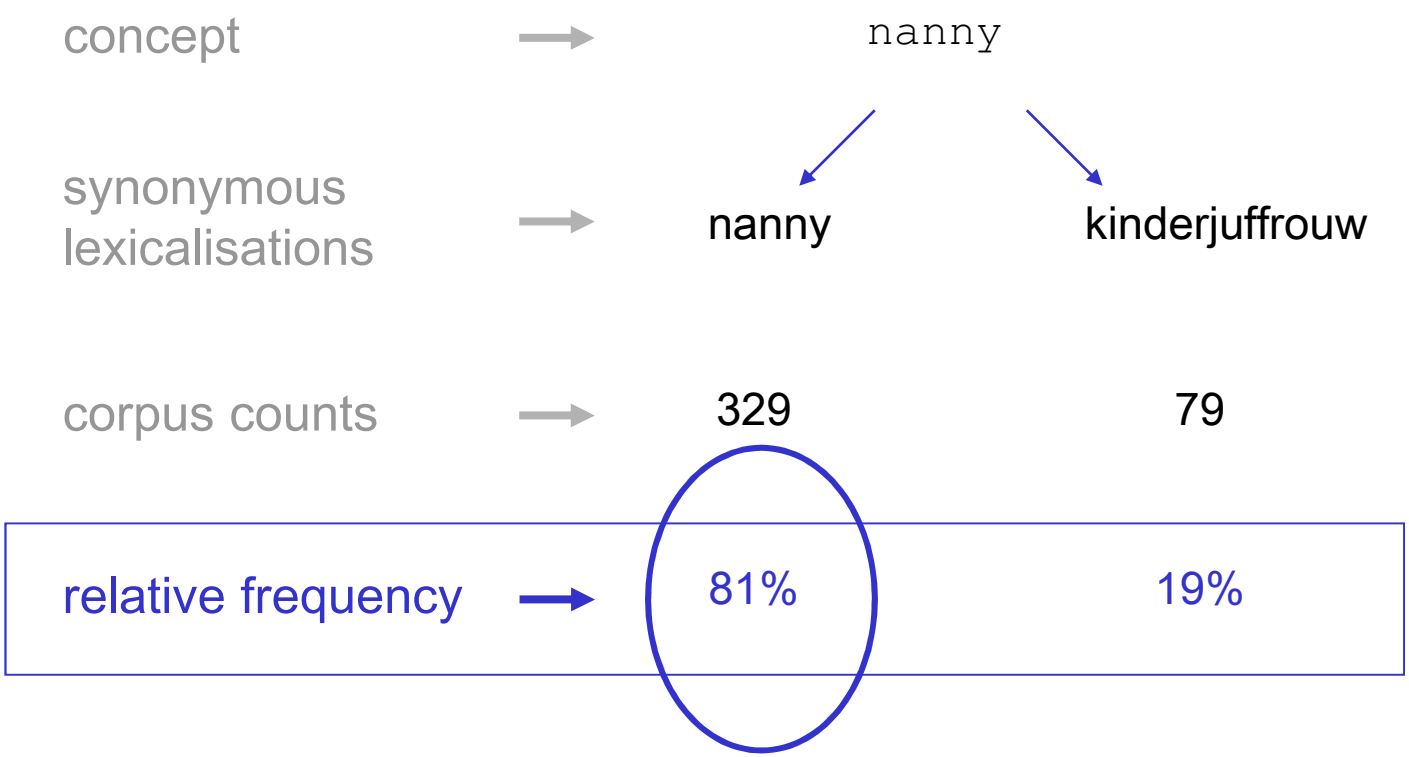
100 concepts remaining

Polysemy: stable concepts?

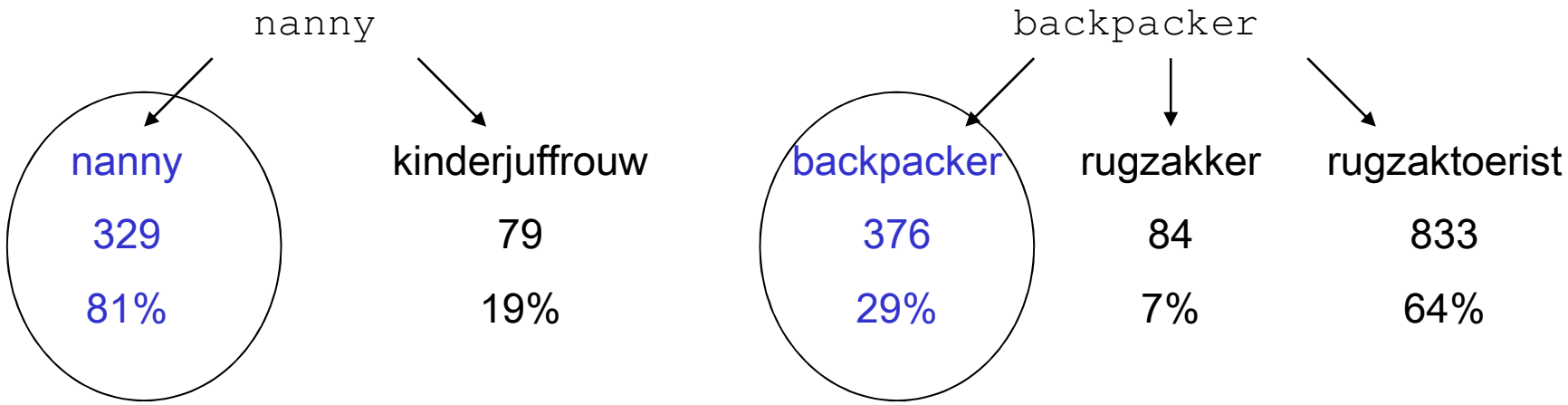
- Checked manually: polysemous items with reasonable frequency (*freak, chicken*)
- Excluded: concepts containing high-frequent polysemous lexicalisations or vague items (*buddy – maatje*)

Step 5:

focus on success rates of anglicisms



comparing success of all English PRN



How to explain variation in these success rates?

Comment: Response Variable

- Success rate for English lexemes: 4 measuring points
  → four subcorpora: split out for (1) region; (2) register

	freq. *hacker*	conc.freq	success rate
hacker_BelgDutch_POP	1001	1100	.910
hacker_BelgDutch_QUAL	1346	1424	.945
hacker_NethDutch_POP	440	478	.921
hacker_NethDutch_QUAL	915	987	.927

An onomasiological perspective on borrowing

Background

Case Study: English Person Reference Nouns in Dutch

- method
  - building profiles
  - explaining the variation: possibly influential features
  - multivariate approach
- results

Conclusions and Prospects

Existing research:

focus on formal features, ignoring lectal and conceptual features

Current approach: multivariate

Influential Features

Lectal Features

Word-Related Features

Conceptual Features

Influential Features

Lectal Features

- Region Belgian Dutch vs. Netherlandic Dutch
- Register Qualitative Newspapers vs. Regional Newspapers

Word-Related Features

Conceptual Features

Influential Features

Lectal Features

Word-Related Features

- etymological: era borrowing
- formal: length of the word

Conceptual Features

Influential Features

Lectal Features

Word-Related Features

- etymological: era of borrowing
  (based on >70 lexicographical sources)

<1945	*dandy*
1945-1989	*babyboomer*
>1989	*creep*

Conceptual Features

Influential Features

Lectal Features

Word-Related Features

- etymological: era of borrowing

- formal: length of the word

  shortest equivalent — *babysit* (vs. *kinderoppas*)

  not the shortest — *ghostwriter* (vs. *nègre*)

Conceptual Features

Influential Features

Lectal Features

Word-Related Features

Conceptual Features

- “necessary loans” vs. “luxury loans”
- concept frequency in RL
- lexical field

Influential Features

Lectal Features

Word-Related Features

Conceptual Features

- "necessary loans" vs. "luxury loans"

necessary: PRN is the first lexicalization of a new concept

webmaster

luxury: the PRN is an extra lexicalization:

0-80 yrs: first lexicalization introduced less than 80 years prior to borrowing

consultant

>80 yrs: The first lexicalization was introduced more than 80 years prior to borrowing

soulmate

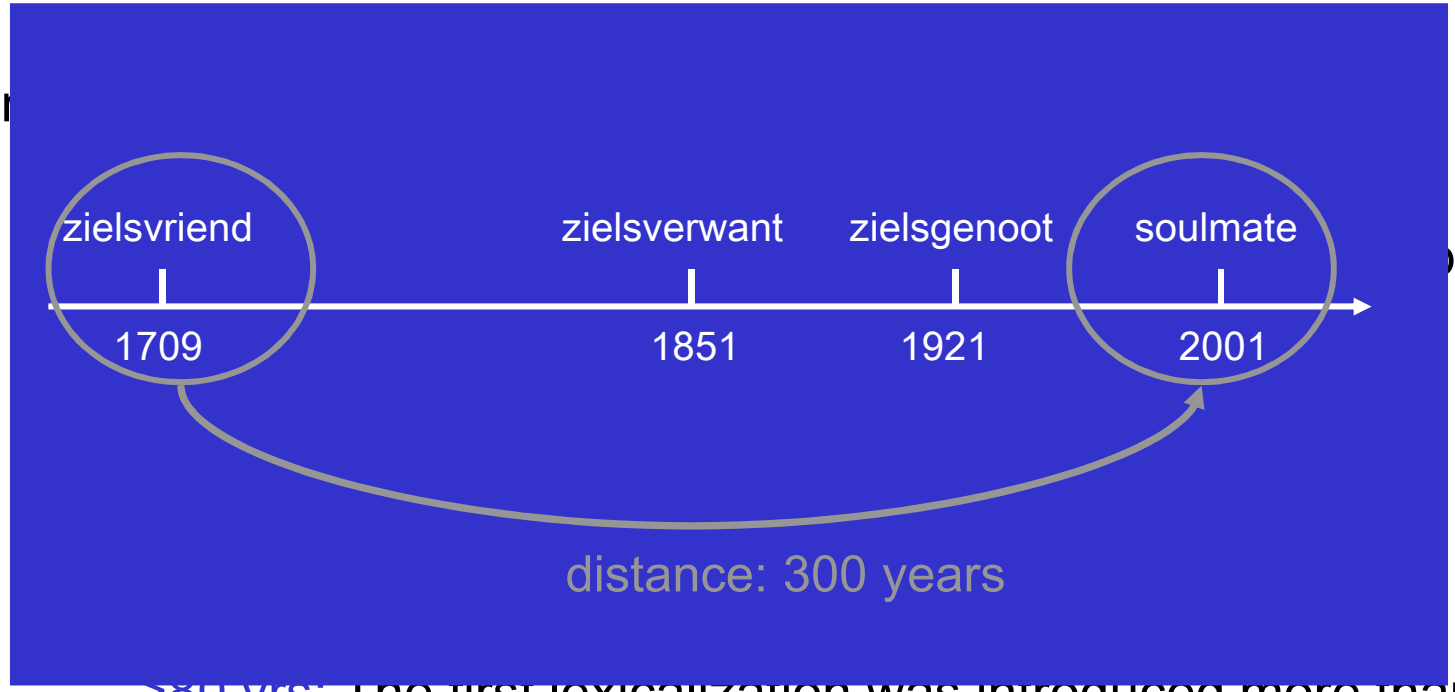
Influential Features

Lectal Features

Word-Related Features

Co

-



>80 yrs: The first lexicalization was introduced more than 80 years prior to borrowing

soulmate

Influential Features

Lectal Features

Word-Related Features

Conceptual Features

- age concept at time of introduction of the loanword
- concept frequency in RL

`keeper`: freq(keeper + goalkeeper + goalie + doelverdediger + doelman)

=

108 492

`sniper`: freq(sniper + scherpschutter + sluipschutter)

=

3430

Influential Features

Lectal Features

Word-Related Features

Conceptual Features

- age concept at time of introduction of the loanword
- concept frequency in RL
- lexical field:
  - Media & IT — `hacker`
  - Sports & Leisure — `golfer`
  - Making Money — `marketeer`
  - Social Life: Neutral — `teenager`
  - Social Life: Deviance — `motherfucker, hooker`

Summary: Predictors

- Lectal features
- Word-related features
  - etymological: era borrowing
  - formal: length of the word
- Conceptual features
  - age concept at time of introduction of the loanword
  - concept frequency in RL
  - lexical field

An onomasiological perspective on borrowing

Background

Case Study: English Person Reference Nouns in Dutch

- method
  - building profiles
  - explaining the variation: possibly influential features
  - multivariate approach
- results

Conclusions and Prospects

Analysis

Using the appropriate technique to:

(1) take the combined effect of features into account in determining which features are influential

(2) generalize over the 100 items (400 measuring points) under scrutiny

→ mixed-effect linear regression analysis

Main effects model

Interpretation -- extra: Interactions (based on more complex model)

An onomasiological perspective on borrowing

Background

Case Study: English Person Reference Nouns in Dutch

- method
  - building profiles
  - explaining the variation: possibly influential features
  - multivariate approach
- results

Conclusions and Prospects

Results →

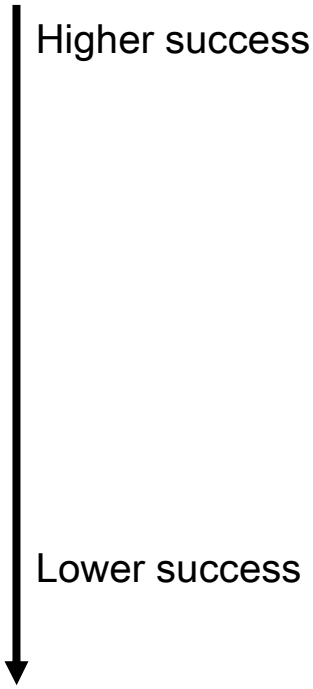
Only sign.predictors are shown

	Estim.	St.Dev.	z/t-value	pr(>|z)	Signif.
(Intercept)	5.38	1.20	4.48	0.000	***
necess_lux: cat_21	-0.74	0.30	-2.50	0.012	*
[illegible]	[illegible]	[illegible]	[illegible]	0.012	*
[illegible]	[illegible]	[illegible]	[illegible]	0.069	.
era_loanword: group_22	-0.32	0.15	-2.17	0.030	*
log(RL_concfreq)	-0.69	0.12	-5.56	0.000	***
lexfield: sports_leisure	-0.68	0.75	[illegible]	0.366	
lexfield: moneymaking	-1.04	0.72	-1.45	0.148	
lexfield: social_neutral	-0.42	0.71	-0.59	0.558	
lexfield: social_deviance	-1.22	0.75	-1.63	0.104	.

Compared to reference value (Media & IT)

positive = English more success
negative = English less success

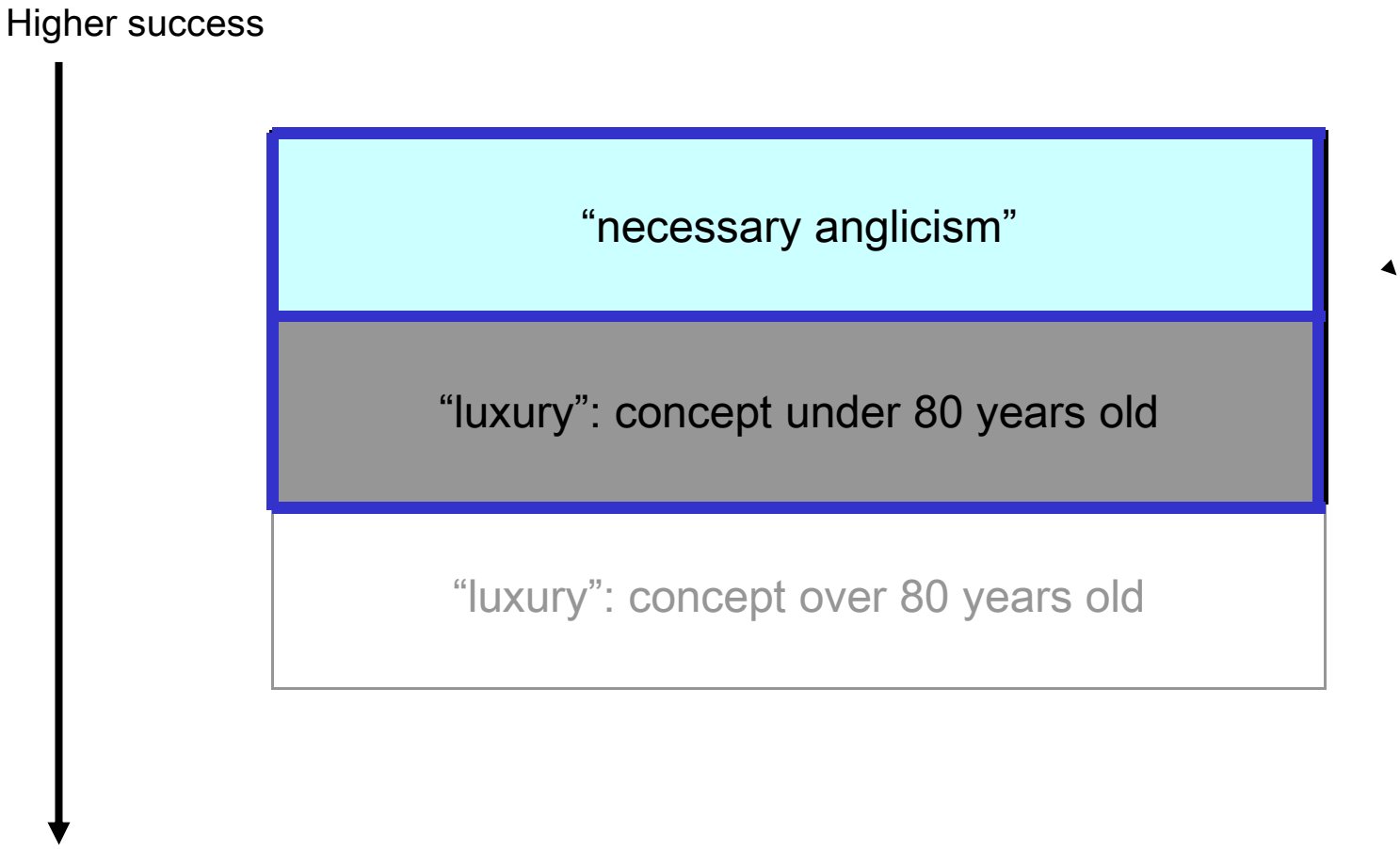
< 0.05 = significant


Higher success
Lower success

Results

	Estim.	St.Dev.	z/t-value	pr(>|z)	Signif.
(Intercept)	5.38	1.20	4.48	0.000	***
necess_lux: cat_21	-0.74	0.30	-2.50	0.012	*
necess_lux : cat_22	-0.37	0.15	-2.52	0.012	*
era_loanword: group_21	-0.46	0.25	-1.82	0.069	.
era_loanword: group_22	-0.32	0.15	-2.17	0.030	*
log(RL_concfreq)	-0.69	0.12	-5.56	0.000	***
lexfield: sports_leisure	-0.68	0.75	-0.90	0.366	
lexfield: moneymaking	-1.04	0.72	-1.45	0.148	
lexfield: social_neutral	-0.42	0.71	-0.59	0.558	
lexfield: social_deviance	-1.22	0.75	-1.63	0.104	.

Age Concept at Time of Borrowing



Age Concept at Time of Borrowing

Higher success

anglicisms filling lexical gaps have a higher success rate than anglicisms introduced as alternative lexicalization for an existing concept

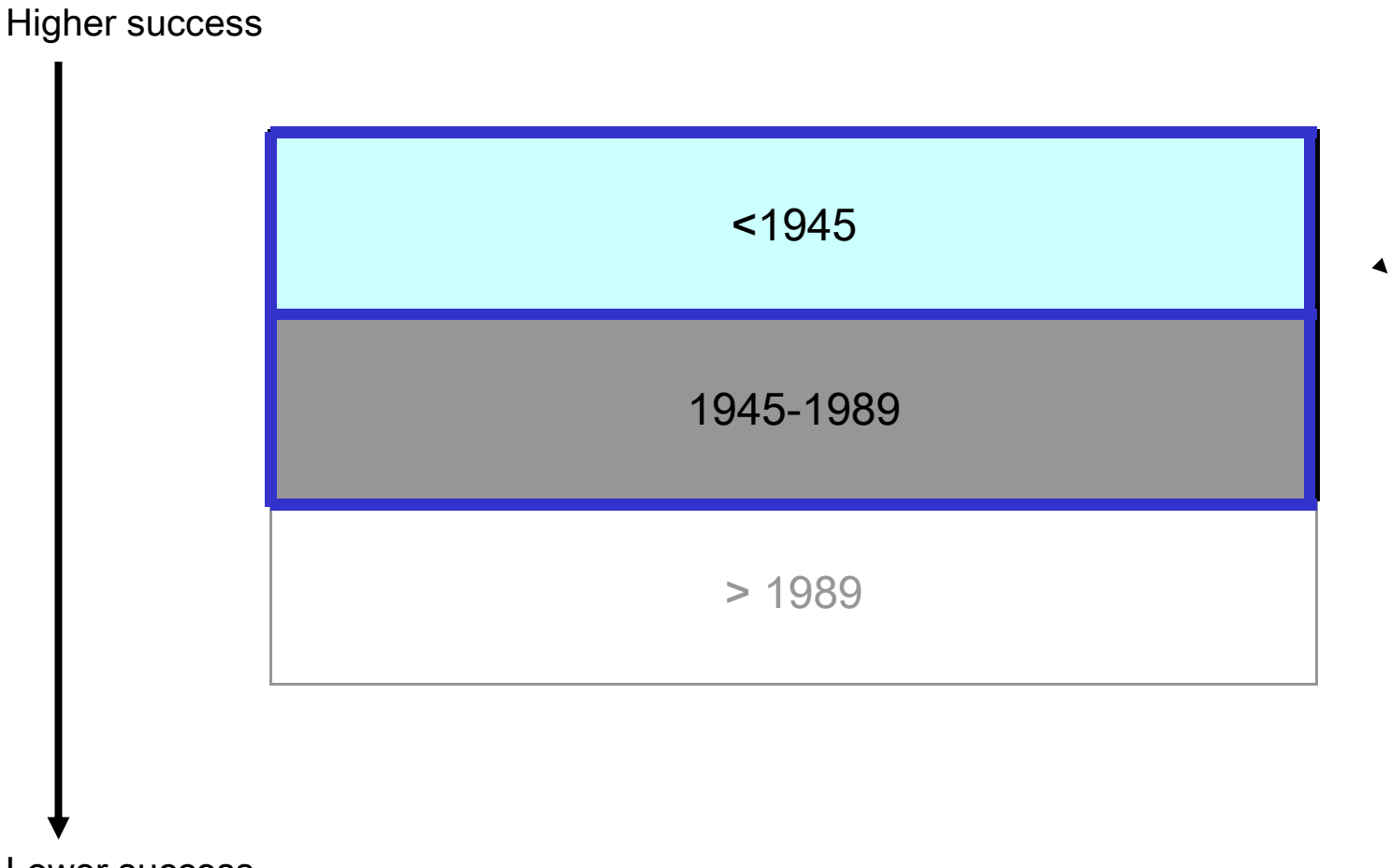
the older a concept at the time the anglicism is introduced, the less successful the anglicism: it has to compete with a well-established alternative lexicalisation

Lower success

Results

	Estim.	St.Dev.	z/t-value	pr(>|z)	Signif.
(Intercept)	5.38	1.20	4.48	0.000	***
necess_lux: cat_21	-0.74	0.30	-2.50	0.012	*
necess_lux : cat_22	-0.37	0.15	-2.52	0.012	*
era_loanword: group_21	-0.46	0.25	-1.82	0.069	.
era_loanword: group_22	-0.32	0.15	-2.17	0.030	*
log(RL_concfreq)	-0.69	0.12	-5.56	0.000	***
lexfield: sports_leisure	-0.68	0.75	-0.90	0.366	
lexfield: moneymaking	-1.04	0.72	-1.45	0.148	
lexfield: social_neutral	-0.42	0.71	-0.59	0.558	
lexfield: social_deviance	-1.22	0.75	-1.63	0.104	.

Era Loanword



Era Loanword

Higher success

Lower success

the older the loanword, the more success
→ new words need time to settle in

no proof for World War II or Wall effect
→ mainly effect on nr. of types?

Preliminary Summary

Necessary vs. luxury anglicisms

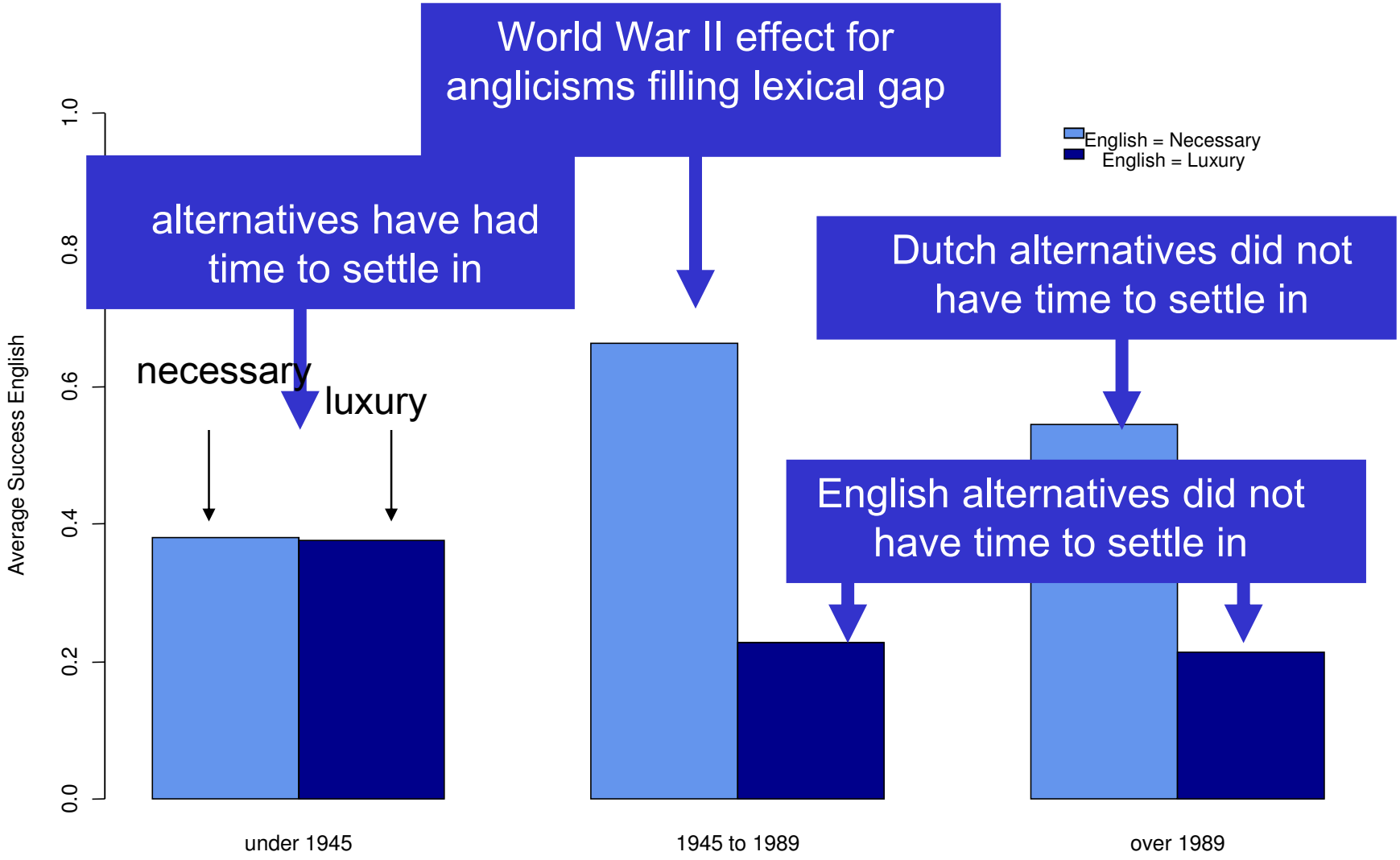
most success for anglicisms filling lexical gaps

Era of introduction

most success for old loanwords

no WW II effect

CAUTION: INTERACTION


World War II effect for anglicisms filling lexical gap
alternatives have had time to settle in
Dutch alternatives did not have time to settle in
English alternatives did not have time to settle in
English = Necessary
English = Luxury
necessary
luxury
Average Success English
0.0
0.2
0.4
0.6
0.8
1.0
under 1945
1945 to 1989
over 1989

Preliminary Summary

World War II effect

for anglicisms filling lexical gaps

Luxury vs. Necessary

absent for older words: alternatives have had time to settle in

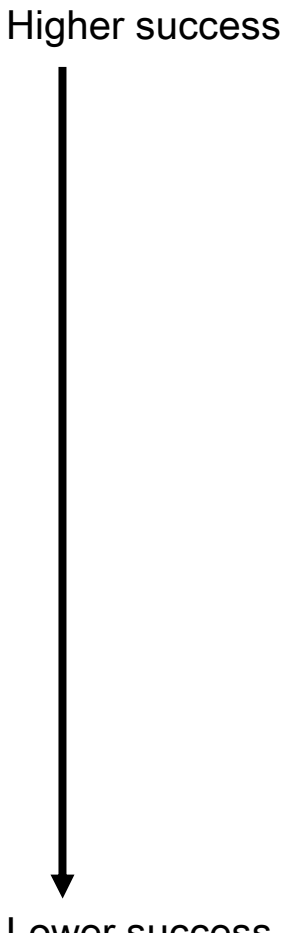
present for younger words:

alternative lexicalisations need more time to settle in

Results

	Estim.	St.Dev.	z/t-value	pr(>|z)	Signif.
(Intercept)	5.38	1.20	4.48	0.000	***
age_conc: cat_21	-0.74	0.30	-2.50	0.012	*
age_conc: cat_22	-0.37	0.15	-2.52	0.012	*
era_loanword: group_21	-0.46	0.25	-1.82	0.069	.
era_loanword: group_22	-0.32	0.15	-2.17	0.030	*
log(RL_concfreq)	-0.69	0.12	-5.56	0.000	***
lexfield: sports_leisure	-0.68	0.75	-0.90	0.366	
lexfield: moneymaking	-1.04	0.72	-1.45	0.148	
lexfield: social_neutral	-0.42	0.71	-0.59	0.558	
lexfield: social_deviance	-1.22	0.75	-1.63	0.104	.

Concept Frequency



anglicisms naming low-frequent concepts

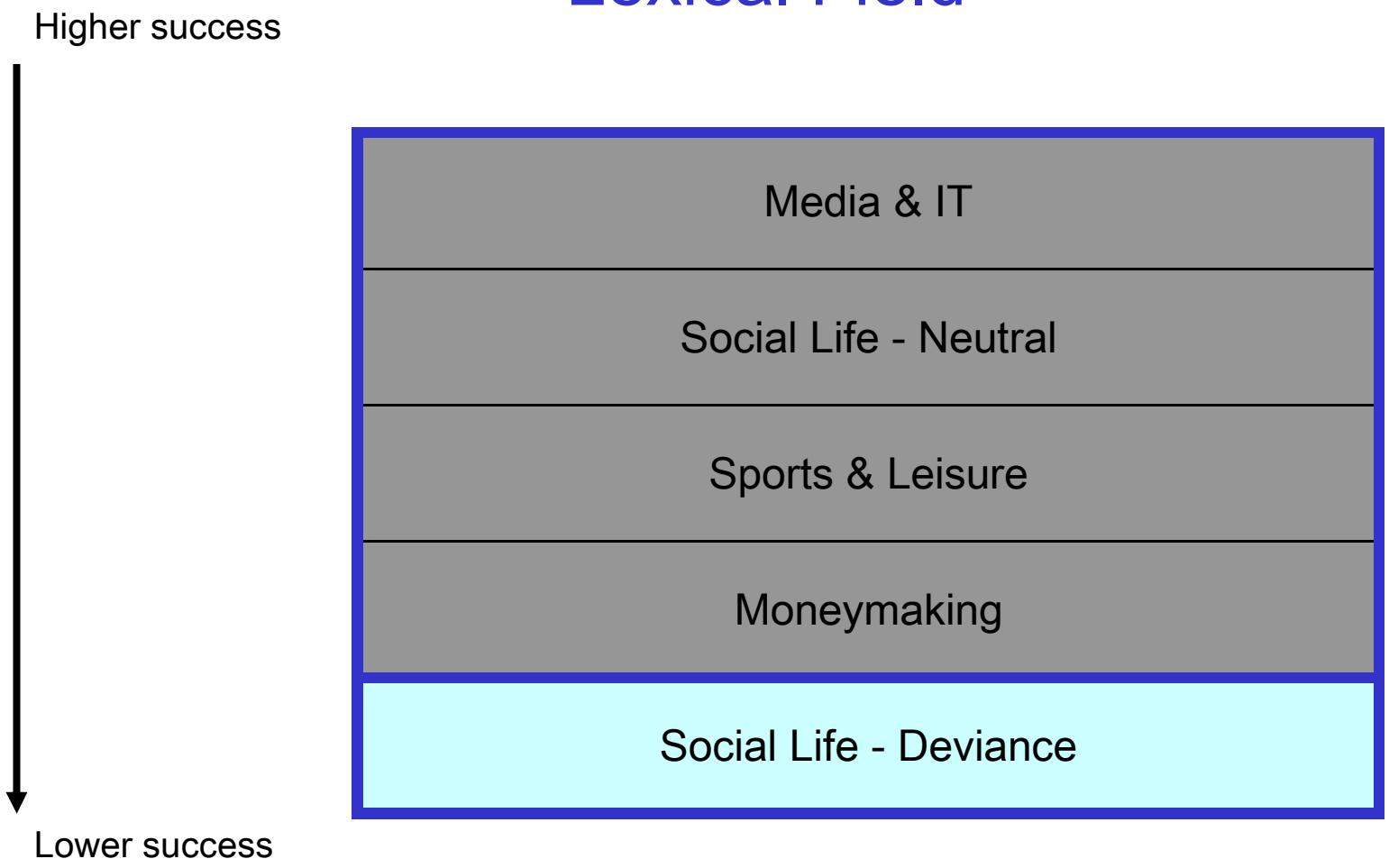
>>>

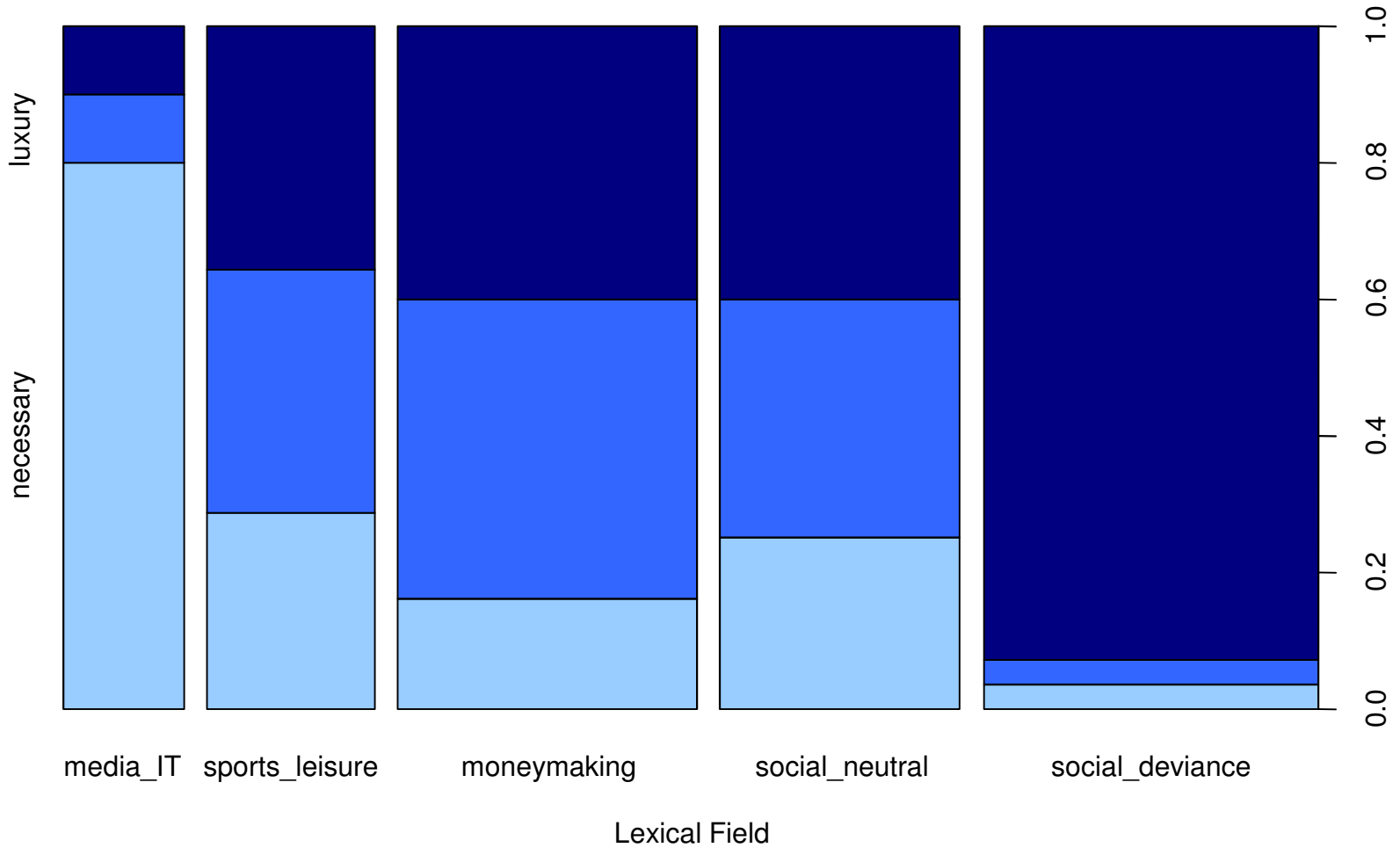
anglicisms naming high-frequent concepts

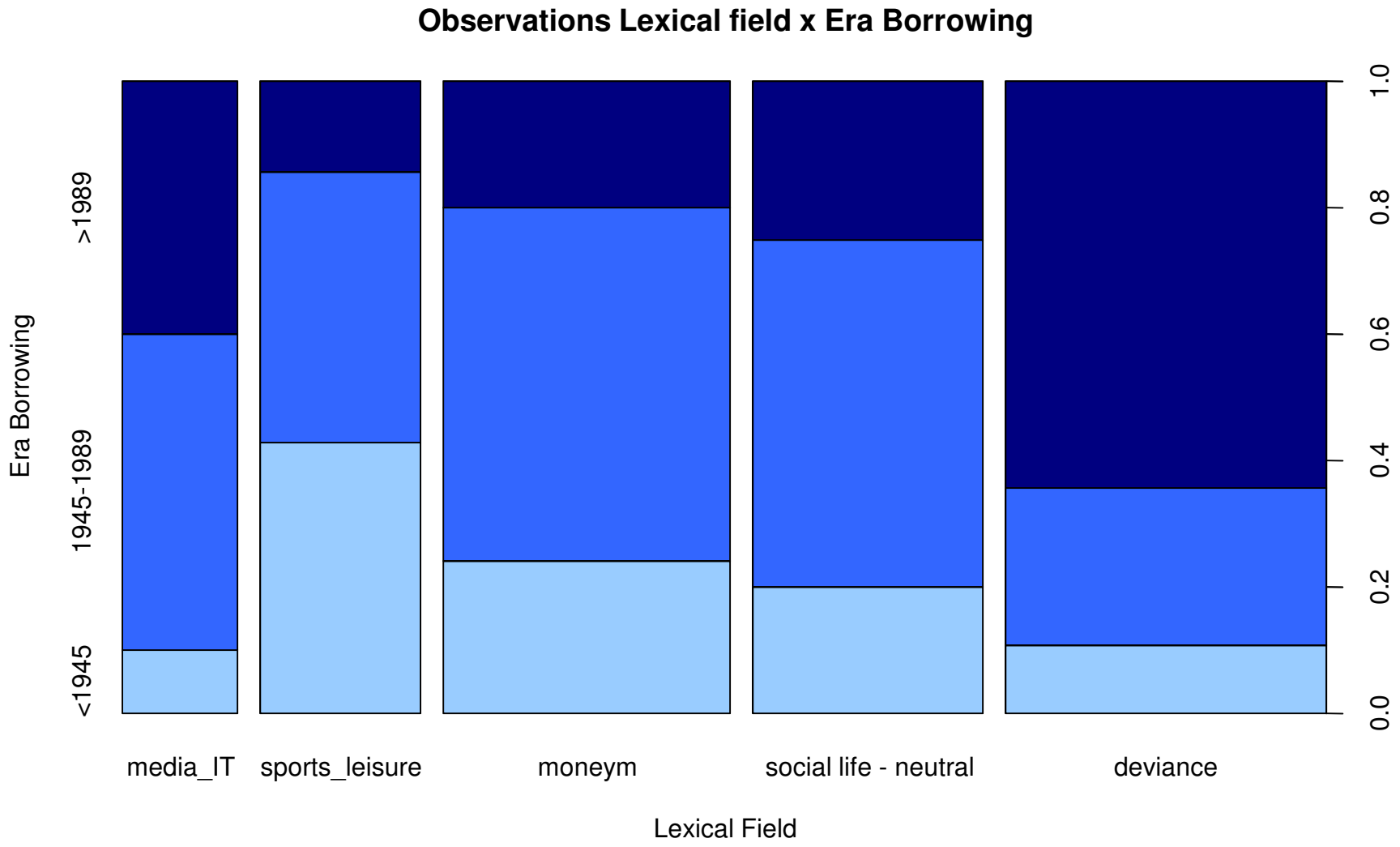
Results

	Estim.	St.Dev.	z/t-value	pr(>|z)	Signif.
(Intercept)	5.38	1.20	4.48	0.000	***
age_conc: cat_21	-0.74	0.30	-2.50	0.012	*
age_conc: cat_22	-0.37	0.15	-2.52	0.012	*
era_loanword: group_21	-0.46	0.25	-1.82	0.069	.
era_loanword: group_22	-0.32	0.15	-2.17	0.030	*
log(RL_concfreq)	-0.69	0.12	-5.56	0.000	***
lexfield: sports_leisure	-0.68	0.75	-0.90	0.366	
lexfield: moneymaking	-1.04	0.72	-1.45	0.148	
lexfield: social_neutral	-0.42	0.71	-0.59	0.558	
lexfield: social_deviance	-1.22	0.75	-1.63	0.104	.

Lexical Field




Observations Lexical field x Luxury/Necessary Loans
luxury
necessary
1.0
0.8
0.6
0.4
0.2
0.0
media_IT
sports_leisure
moneymaking
social_neutral
social_deviance
Lexical Field


Observations Lexical field x Era Borrowing
Era Borrowing
>1989
1945-1989
<1945
0.0
0.2
0.4
0.6
0.8
1.0
media_IT
sports_leisure
moneym
social life - neutral
deviance
Lexical Field

Conclusions

Methodologically:

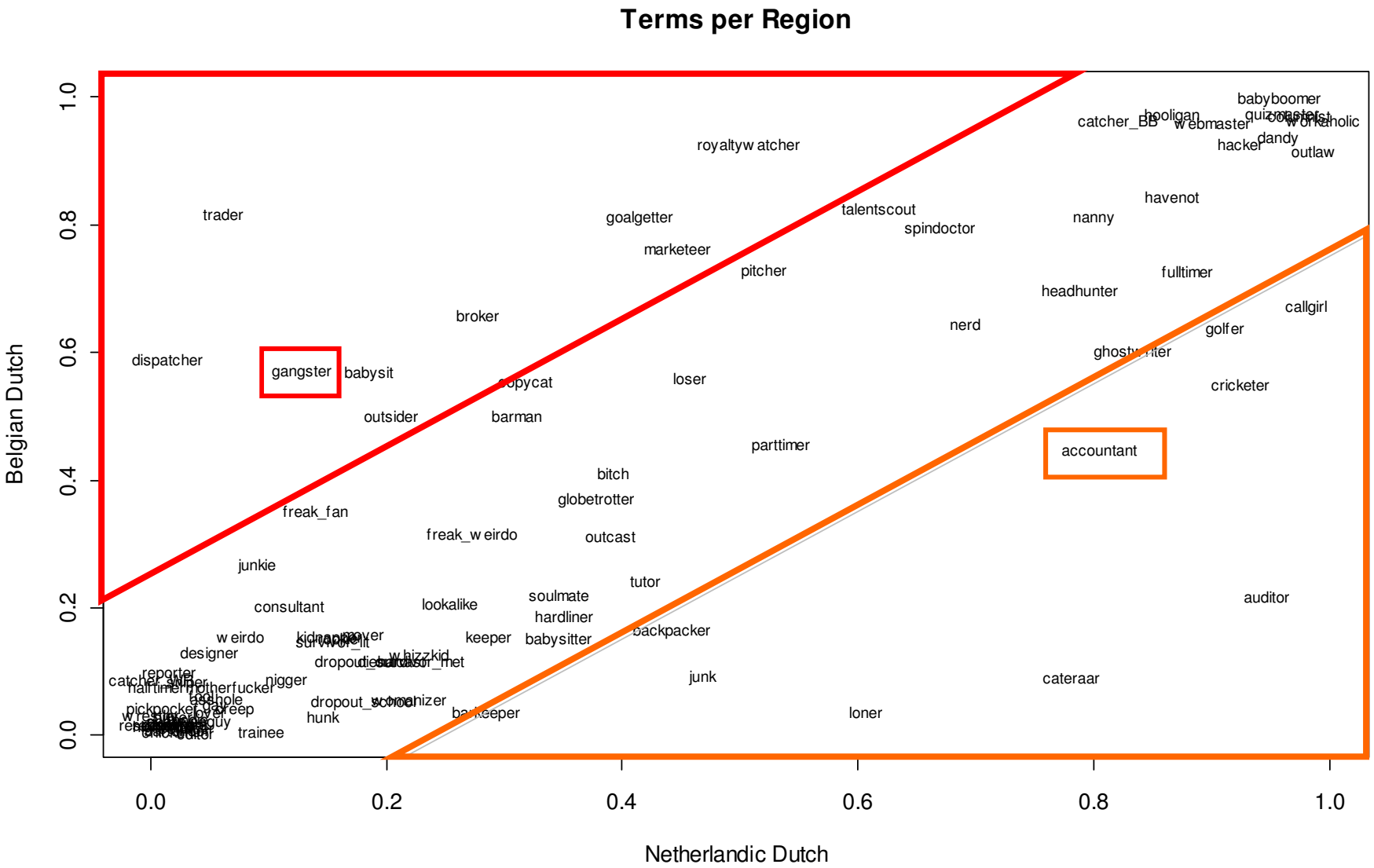
synonymy in contact linguistics: the success of loanwords

Results → Significant:

- necessary vs. luxury loans
- era of borrowing
- concept frequency
- lexical field

Results → No Significant Effect:

- length of the word
- lectal parameters: more effect when scrutinizing items


Terms per Region
Belgian Dutch
Netherlandic Dutch
0.0
0.2
0.4
0.6
0.8
1.0
royaltyw atcher
trader
goalgetter
talentscout
spindoctor
nanny
havenot
marketeer
pitcher
fulltimer
headhunter
broker
nerd
callgirl
golfer
dispatcher
gangster
babysit
copycat
loser
cricketer
outsider
barman
parttimer
accountant
bitch
globetrotter
freak_fan
freak_w eirdo
outcast
junkie
tutor
consultant
lookalike
soulmate
hardliner
auditor
w eirdo
keeper
babysitter
backpacker
designer
junk
cateraar
nigger
hunk
loner
trainee
babyboomer
catcher_BB
hooligan
w ebmaster
hacker
dandy
outlaw
w orkaholic

Conclusions

Attenuating the results

– “synonymy exists”

→ prospects:

- Closer semantic analyses needed: JUNKIE (comprehensive approach)
- Regional variation

For more information:

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

eline.zenner@arts.kuleuven.be