



# *From bitches to lovers*

The usage-based onomasiology of anglicisms

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



University of Leuven  
RU Quantitative Lexicology and Variational Linguistics



Identifying & Describing Lexical Borrowing, March 2010

# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case Study: Person Reference Nouns
  - What?
  - Method:
    - Data-collection
    - Variables
    - Analyses
  - Results
- Conclusions & Prospects



# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case Study: Person Reference Nouns
  - What?
  - Method:
    - Data-collection
    - Variables
    - Analyses
  - Results
- Conclusions & Prospects



# The Spread of English

## The aftermath of WW II:

Massive increase in the use of English around the world, both as

- language for communication (*macro-level*)
- a resource for borrowing (*micro-level*)

## Hot topic in linguistics

- *macro-level*: ELF, Business English, World Englishes, TEFL
- to a lesser extent on the *micro-level*: Anglicism Research



# The Spread of English

The aftermath of WW II:

Massive increase in the use of English around the world, both as

- language for communication (*macro-level*)
- a resource for borrowing (*micro-level*)

Hot topic in linguistics

- *macro-level*: ELF, Business English, World Englishes, TEFL
- to a lesser extent on the *micro-level*: **Anglicism Research**



# Anglicism Research: Shortcomings

## 1. Structuralist Paradigm

### Main focus

Classifying anglicisms based on internal linguistic features

Ignored:

Why borrow which English items; importance of:

- Lectal factors
- Conceptual factors

# Anglicism Research: Shortcomings

## 1. Structuralist Paradigm

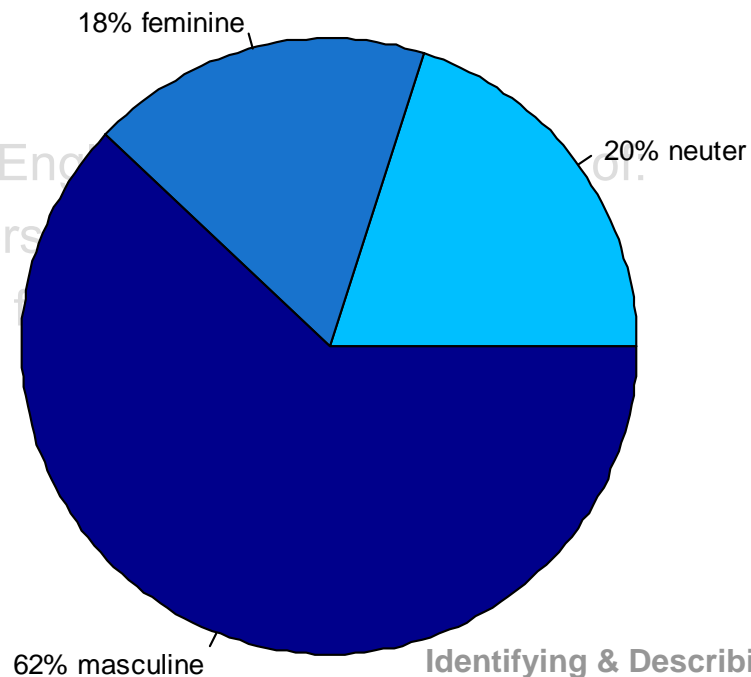
Main focus

Classifying anglicisms based on internal linguistic features (Onysko 2007: 137)

Ignored:

Why borrow which English word?

- Lectal factors
- Conceptual factors



# Anglicism Research: Shortcomings

## 1. Structuralist Paradigm

Main focus

Classifying anglicisms based on internal linguistic features (Onysko 2007: 137)

Ignored:

Why borrow and use which English items; importance of:

- Lectal factors
- Conceptual factors





# Anglicism Research: Shortcomings

## 2. Rudimentary Success Measures

### Main focus

Descriptive type- or token-counts based on small corpora

Simplified:

- topic text
- presence domestic alternative

# Anglicism Research: Shortcomings

## 2. Rudimentary Success Measures

Main focus

Descriptive type- or token-counts based on small corpora (Yang 1990: 27)

| Year | Tokens | Nr. of Pages | Tokens per Page |
|------|--------|--------------|-----------------|
| 1950 | 633    | 235          | 2.7             |
| 1960 | 1114   | 475          | 2.35            |
| 1970 | 3557   | 1175         | 3               |
| 1980 | 4766   | 1465         | 3.25            |

# Anglicism Research: Shortcomings

## 2. Rudimentary Success Measures

### Main focus

Descriptive type- or token-counts based on small corpora

### Simplified:

- topic text
- presence domestic alternative

# Anglicism Research: Shortcomings

## 2. Rudimentary Success Measures

### Main focus

Descriptive type- or token-counts based on small corpora

### Simplified:

- topic text
- presence domestic alternative

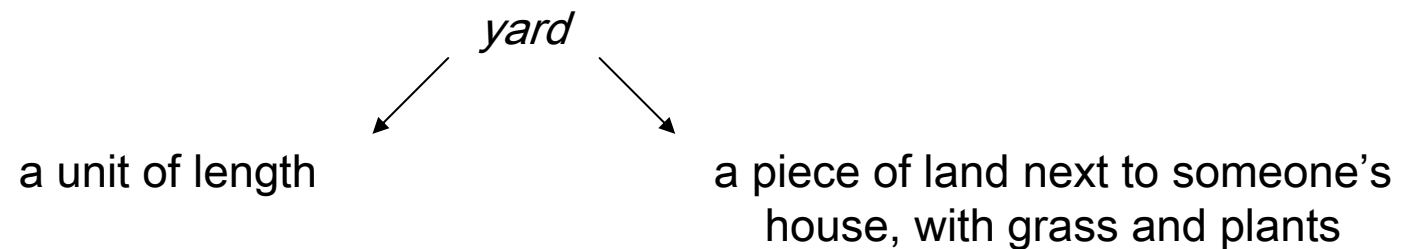
→ profile-based method of onomasiological variation  
(Geeraerts *et al.* 1999)

# Semasiology vs. Onomasiology

## Semasiological Perspective

lexeme →

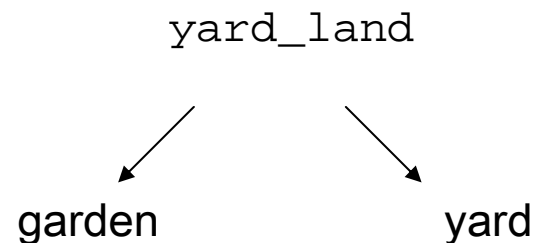
meanings →



## Onomasiological Perspective

concept →

lexicalisations →



# Semasiology vs. Onomasiology

## Semasiological Perspective

lexeme →

*yard*

meanings →

a unit of length

a piece of land next to someone's house, with grass and plants

## Onomasiological Perspective

concept →

yard\_land

lexicalisations →

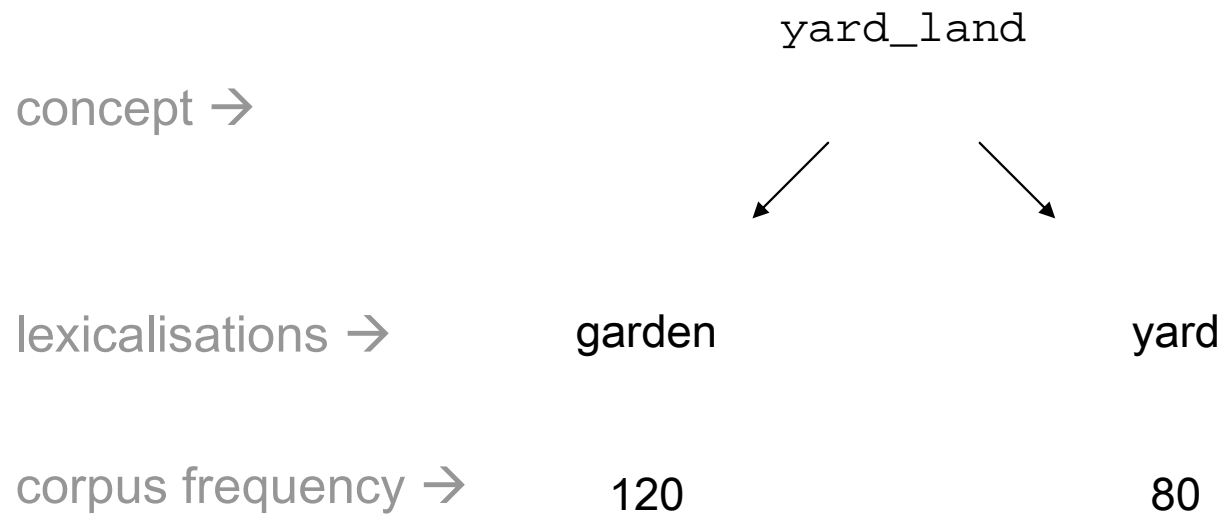
garden

yard



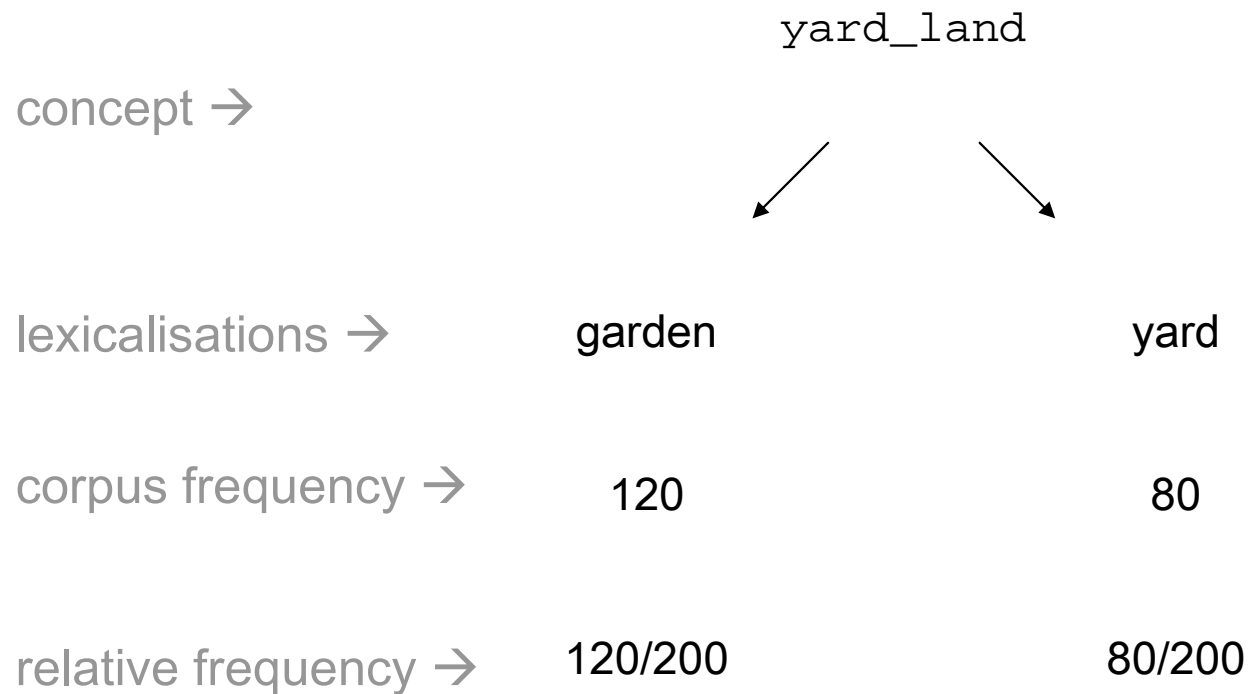
# Profile-based method

Compare success of lexicalizations



# Profile-based method

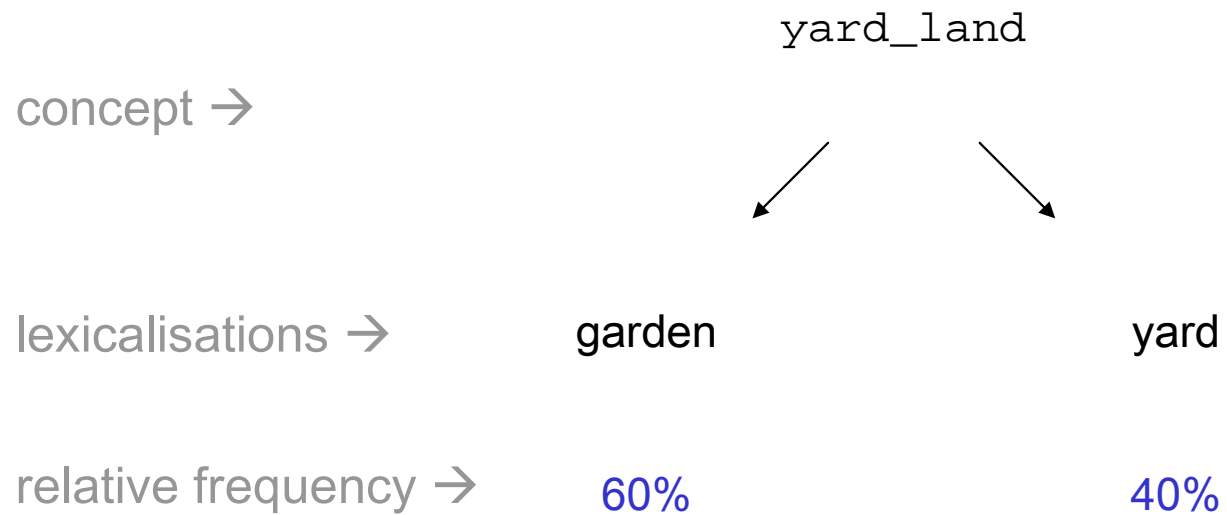
Compare success of lexicalizations





# Profile-based method

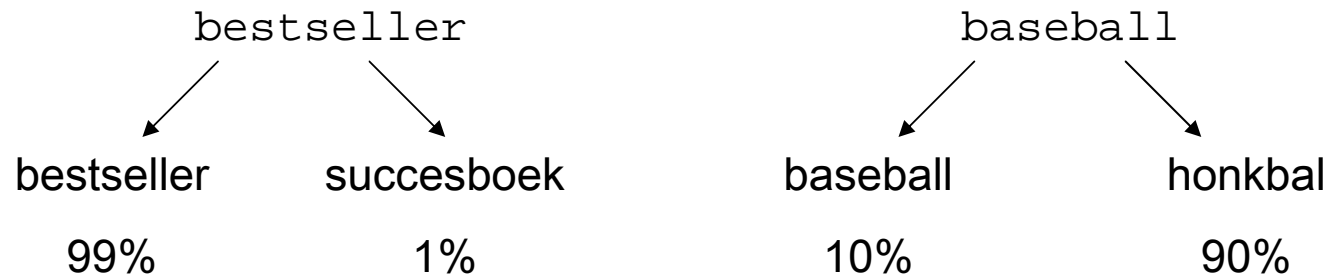
Compare success of lexicalizations



# Profile-based method

Applied to the Success of Anglicisms:

## Variation in Lexicalization Preference



# Profile-based method

Applied to the Success of Anglicisms:

## Variation in Lexicalization Preference



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

# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case-Study: Person Reference Nouns
  - What?
  - Method:
    - Data-collection
    - Variables
    - Analyses
  - Results
- Conclusions & Prospects

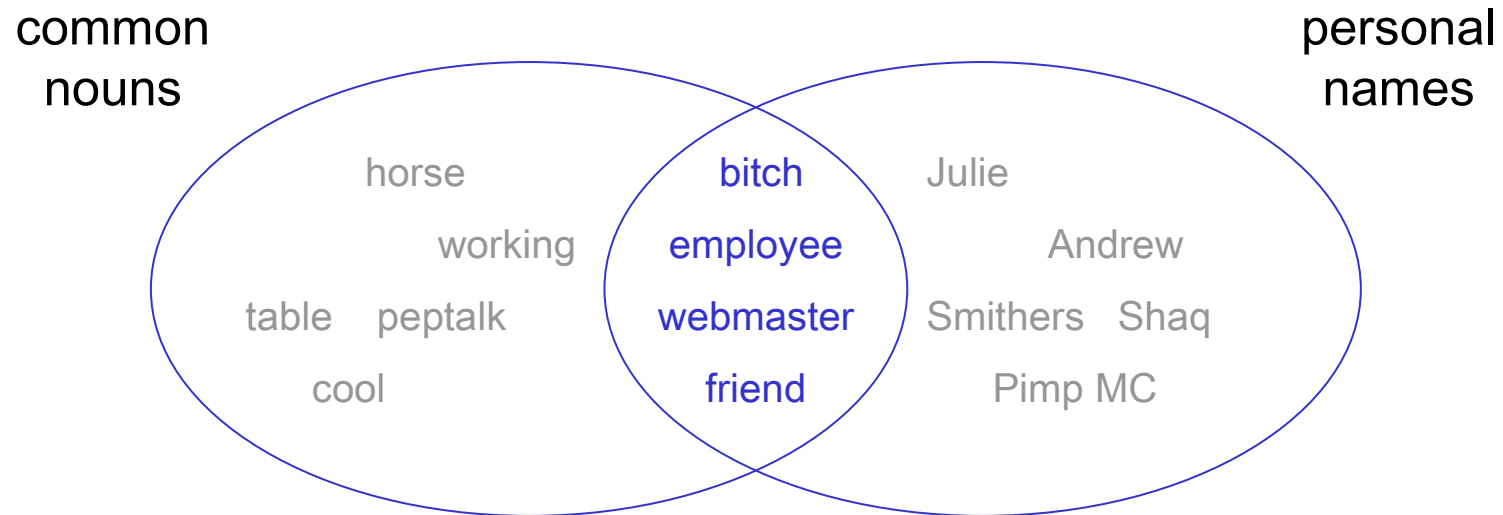


# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case-Study: Person Reference Nouns
  - What?
  - Method:
    - Data-collection
    - Variables
    - Analyses
  - Results
- Conclusions & Prospects



# English Person-Reference Nouns (PRN)



# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case-Study: Person Reference Nouns
  - What?
  - Method:
    - Data-collection
    - Variables
    - Analyses
  - Results
- Conclusions & Prospects



## Method

Which features influence the success-rate of English PRN in Dutch?



Needed:

- set of English PRN with Dutch alternatives & success-rate based on large corpus
- identification possible influential features
- determining impact of features



# Method

Which features influence the success-rate of English PRN in Dutch?



Needed:

- set of English PRN with Dutch alternatives & success-rate based on large corpus
- identification possible influential features
- determining impact of features

# Data Collection: PRN

**Collection:** success-rates for 119 English PRN from 98 concepts

Step 1: corpus of Dutch

Step 2: English PRN occurring in Dutch

Step 3: Dutch alternatives

Step 4. Automatic Retrieval Tokens PRN

Step 5. Calculation success-rate for each of the 114 English PRN



## Data Collection: PRN

Collection: success-rates for 119 English PRN from 98 concepts

### Step 1: corpus of Dutch (newspapers, parsed)

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



# Data Collection: PRN

Collection: success-rates for 119 English PRN from 98 concepts

Step 1: corpus of Dutch

Step 2: English PRN occurring in Dutch

- CD-ROM Dutch Dictionary (“iemand” - *someone* + “Eng.”)
- English WordNet: hyponyms of “person”
- Online List: *2400 x Onnodig Engels*

→ formal definition of anglicism

→ only those items occurring in the Dutch corpus



# Data Collection: PRN

Collection: success-rates for 119 English PRN from 98 concepts

Step 1: corpus of Dutch

Step 2: English PRN occurring in Dutch

Step 3: Dutch alternatives

- dictionaries: synonym dictionaries, translation dictionaries, ...
- automatic detection of synonyms (Peirsman *et al.* 2007)
- the alternative from the online list *2400 x Onnodig Engels*

Excluded Concepts:

- concepts without English-Dutch variation (*rookie* – *eerstejaarssporter*)
- concepts with high-frequent polysemous items (*buddy* – *maatje*)



# Data Collection: PRN

Collection: success-rates for 119 English PRN from 98 concepts

Step 1: corpus of Dutch

Step 2: English PRN occurring in Dutch

Step 3: Dutch alternatives

Step 4. Automatic Retrieval Tokens PRN

Excluded

- Proper names (*Philips Consumer Communications*)
- Lexicalized Compounds (*management consultant*)
- Longer stretches of English (*what kind of fool am I?*)
- For polysemous items: manual disambiguation (*freak*)

→ together > 1.5 million tokens



# Data Collection: PRN

Collection: success-rates for 119 English PRN from 98 concepts

Step 1: corpus of Dutch

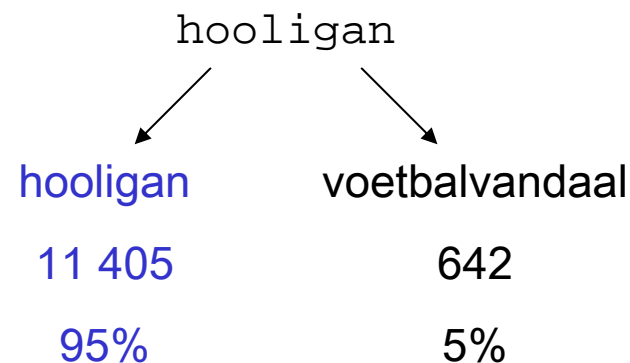
Step 2: English PRN occurring in Dutch

Step 3: Dutch alternatives

Step 4. Automatic Retrieval Tokens PRN

Step 5. Calculation success-rate for each of the 119 English PRN

profile-based method; for each English PRN:



# Method

Which features influence the success-rate of English PRN in Dutch?



Needed:

- set of English PRN with Dutch alternatives & success-rate based on large corpus
- **identification possible influential features**
- determining impact of features





# Influential Features

Lexical Features

Conceptual Features

Word-Related Features



# Influential Features

## Lectal Features

- Region Belgian Dutch vs. Netherlandic Dutch
- Register Qualitative Newspapers vs. Regional Newspapers
- Year 1999-2000 vs. 2001-2002

## Conceptual Features

## Word-Related Features



# Influential Features

Lexical Features

## Conceptual Features

- Age Concept
- Neutrality of the Concept
- Lexical Field

Word-Related Features



# Influential Features

Lexical Features

## Conceptual Features

- Age Concept (Age of oldest lexicalisation: entries editions Van Dale)

*drugsverslaafde* (1976) > *junkie* = *junk* = *addict* (1984)

- <1900          nigger
- 1900-1950      filmstar
- 1950-1980      junkie
- > 1980          talentscout



# Influential Features

Lexical Features

## Conceptual Features

- Age Concept
- Neutrality of the Concept      whore vs. teenager



# Influential Features

Lexical Features

## Conceptual Features

- Age Concept
- Neutrality of the Concept
- Lexical Field (Yahoo Directories, not typology categories):
  - Business & Economy                      marketeer
  - Media & IT                                      hacker
  - Recreation & Sports                      golfer
  - Sexuality & Appraisal                      lover
  - Other    kidnapper



# Influential Features

Lexical Features

Conceptual Features

## Word-Related Features

- **conceptual:** innovation of the PRN
- **etymological:** age word
- **formal:** length of the word
- **formal:** spelling difficulties



# Influential Features

Lectal Features

Conceptual Features

## Word-Related Features

- **conceptual**: innovation of the English PRN (Van Dale)  
YES: the PRN is the first lexicalisation of a new concept  
*webmaster*  
NO: the PRN is an extra lexicalisation for an old concept  
*designer*





# Influential Features

Lexical Features

Conceptual Features

## Word-Related Features

- **conceptual**: innovation of the PRN
- **etymological**: age word (Entries Editions Van Dale)

|           |                   |
|-----------|-------------------|
| <1950     | <i>dandy</i>      |
| 1950-1980 | <i>womanizer</i>  |
| 1980-2000 | <i>cateraar</i>   |
| >1999     | <i>researcher</i> |



# Influential Features

Lectal Features

Conceptual Features

## Word-Related Features

- **conceptual:** innovation of the PRN
- **etymological:** age word
  
- **formal:** length of the word
  - 1 or 2 syllables *fool*
  - > 2 syllables *royaltywatcher*



# Influential Features

Lexical Features

Conceptual Features

## Word-Related Features

- **conceptual**: innovation of the PRN
- **etymological**: age word
- **formal**: length of the word
  
- **formal**: spelling difficulties (*Woordenlijst Nederlandse Taal*)
  - YES      *accountant*
  - NO        *editor*



## Method

Which features influence the success-rate of English PRN in Dutch?



Needed:

- set of English PRN with Dutch alternatives & success-rate based on large corpus
- identification possible influential features
- **determining impact of features**

## Summary: Success of PRN

→ Response variable

- Success-rate for English lexemes
- Based on lexicalization preference in concept
- Split out for (1) region; (2) register; (3) year

|  |                      |
|--|----------------------|
| <a href="#">hooligan_BelgDutch_QUAL_9900</a> | <a href="#">.974</a> |
| <a href="#">hooligan_BelgDutch_QUAL_0102</a> | <a href="#">.980</a> |
| <a href="#">hooligan_BelgDutch_POP_9900</a>  | <a href="#">.969</a> |
| ...  | ...                  |
| <a href="#">hooligan_NethDutch_POP_0102</a>  | <a href="#">.883</a> |

# Summary: Influential Features

- Lectal features
- Conceptual features
- Word-Related features
  - Conceptual
  - Etymological
  - Formal



# 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 98 concepts under scrutiny



# 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 98 concepts under scrutiny

→ mixed-effect linear regression analysis



# 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 98 concepts under scrutiny

→ mixed-effect linear regression analysis

Main effects model

Interpretation -- extra:

- Interactions (based on more complex model)
- Regional variation (based on comparison of models per region)



# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case Study: Person Reference Nouns
  - What?
  - Method:
    - Data-collection
    - Variables
    - Analyses
  - Results
- Conclusions & Prospects



## Results

|                           | Value  | Std.Error | DF  | t-value | p-val |
|---------------------------|--------|-----------|-----|---------|-------|
| (Intercept)               | 0.426  | 0.064     | 755 | 6.670   | 0.000 |
| conc_lexfield_2Recr_Sport | -0.041 | 0.102     | 95  | -0.404  | 0.687 |
| conc_lexfield_2Media_IT   | -0.010 | 0.096     | 95  | -0.100  | 0.921 |
| conc_lexfield_2Relat_Sex  | -0.111 | 0.074     | 755 | -1.495  | 0.135 |
| conc_lexfield_2Other      | -0.419 | 0.085     | 755 | -4.920  | 0.000 |
| word_age1                 | 0.103  | 0.025     | 755 | 4.214   | 0.000 |
| word_age2                 | 0.021  | 0.010     | 755 | 2.096   | 0.037 |
| word_age3                 | -0.017 | 0.012     | 755 | -1.416  | 0.157 |
| conc_innovYES             | 0.251  | 0.039     | 755 | 6.443   | 0.000 |
| word_syll_2>2syll         | -0.101 | 0.032     | 755 | -3.164  | 0.002 |



# Results →

Only sign.predictors are shown

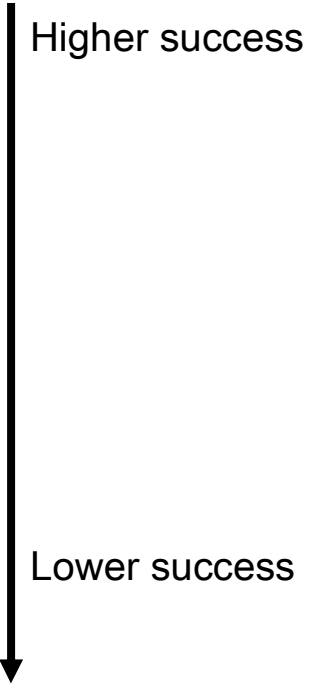
Compared to reference value  
(Business & Economy)

|                           | Value  | Std.Error | DF  | t-value | p-val |
|---------------------------|--------|-----------|-----|---------|-------|
| (Intercept)               | 0.426  | 0.064     | 755 | 6.670   | 0.000 |
| conc_lexfield_2Recr_Sport | -0.041 | 0.102     | 95  | -0.404  | 0.687 |
| conc_lexfield_2Media_IT   | -0.010 | 0.096     | 95  | -0.100  | 0.921 |
| conc_lexfield_2Relat_Sex  | -0.111 | 0.074     | 755 | -1.495  | 0.135 |
| conc_lexfield_2Other      | -0.419 | 0.085     | 755 | -4.920  | 0.000 |
| word_age1                 | 0.103  | 0.025     | 755 | 4.214   | 0.000 |
| word_age2                 | 0.021  | 0.010     | 755 | 2.096   | 0.037 |
| word_age3                 | 0.017  | 0.012     | 755 | -1.416  | 0.157 |
| conc_innovYES             |        |           | 755 | 6.443   | 0.000 |
| word_syll_2>2syll         | -0.101 | 0.032     | 755 | -3.164  | 0.002 |

positive = English less success  
negative = English more success

< 0.05 = significant





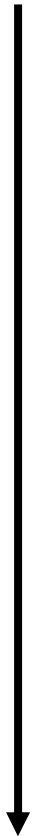
# Results

|                           | Value  | Std.Error | DF  | t-value | p-val |
|---------------------------|--------|-----------|-----|---------|-------|
| (Intercept)               | 0.426  | 0.064     | 755 | 6.670   | 0.000 |
| conc_lexfield_2Recr_Sport | -0.041 | 0.102     | 95  | -0.404  | 0.687 |
| conc_lexfield_2Media_IT   | -0.010 | 0.096     | 95  | -0.100  | 0.921 |
| conc_lexfield_2Relat_Sex  | -0.111 | 0.074     | 755 | -1.495  | 0.135 |
| conc_lexfield_2Other      | -0.419 | 0.085     | 755 | -4.920  | 0.000 |
| word_age1                 | 0.103  | 0.025     | 755 | 4.214   | 0.000 |
| word_age2                 | 0.021  | 0.010     | 755 | 2.096   | 0.037 |
| word_age3                 | -0.017 | 0.012     | 755 | -1.416  | 0.157 |
| conc_innovYES             | 0.251  | 0.039     | 755 | 6.443   | 0.000 |
| word_syll_2>2syll         | -0.101 | 0.032     | 755 | -3.164  | 0.002 |

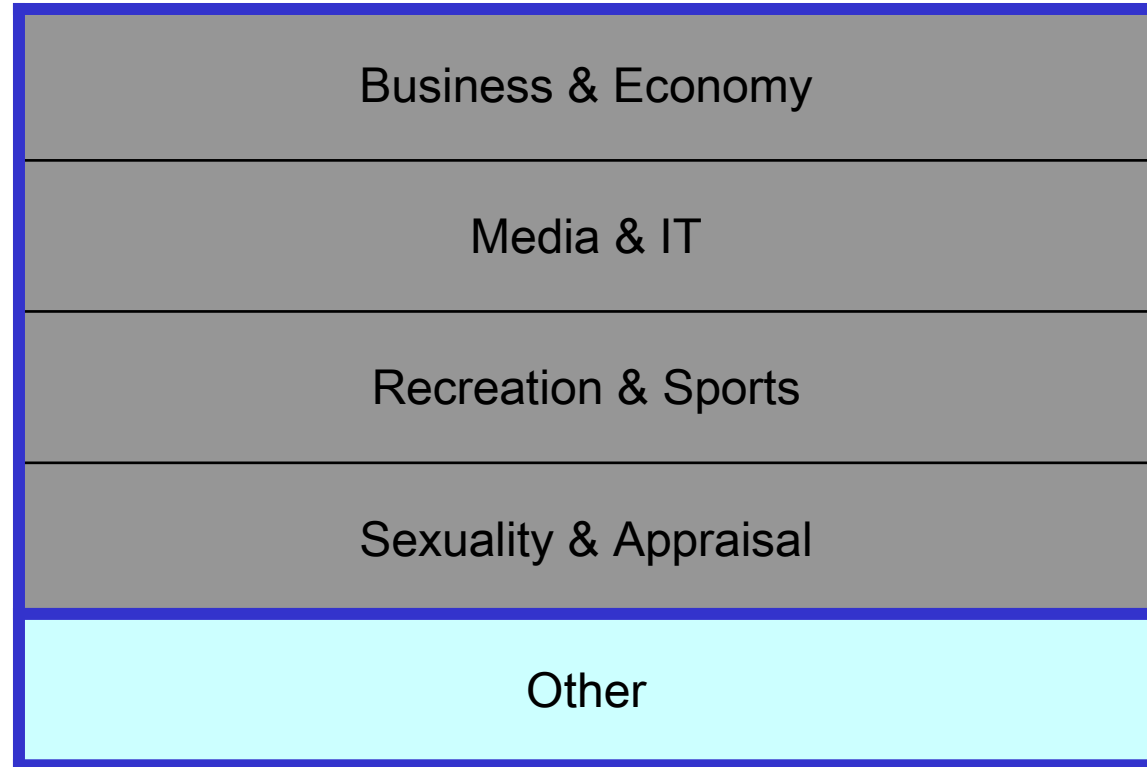


# Lexical Field

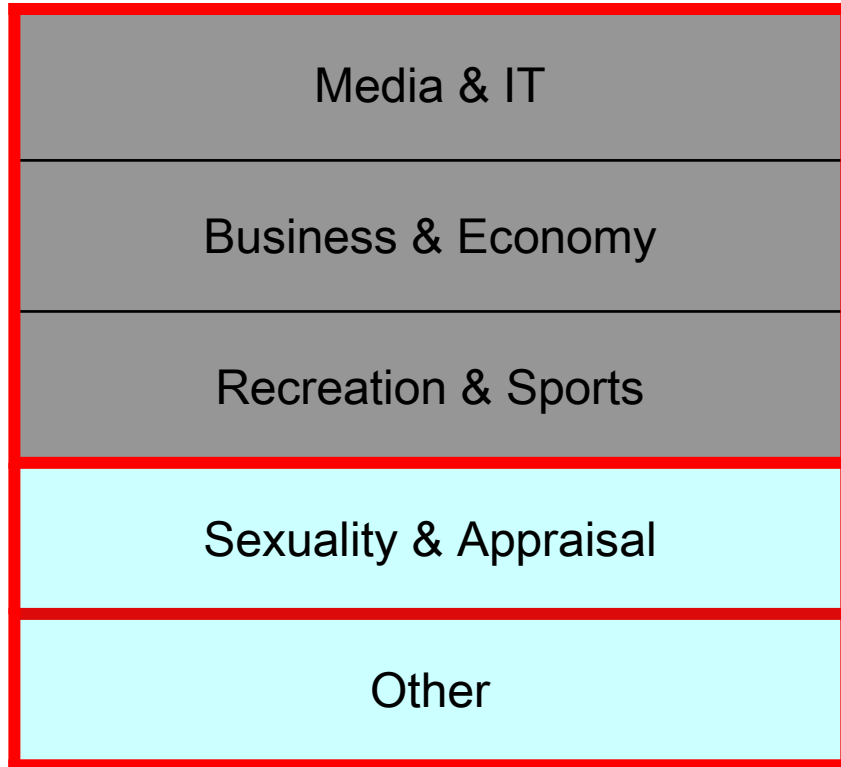
Higher success



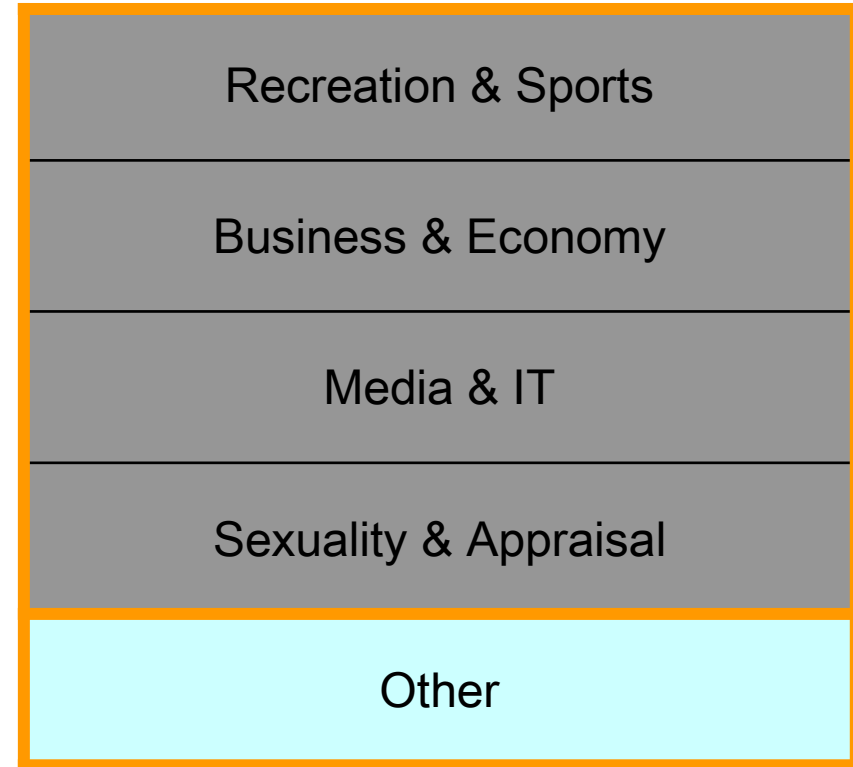
Lower success



# Lexical Field



Belgian Dutch



Netherlandic Dutch





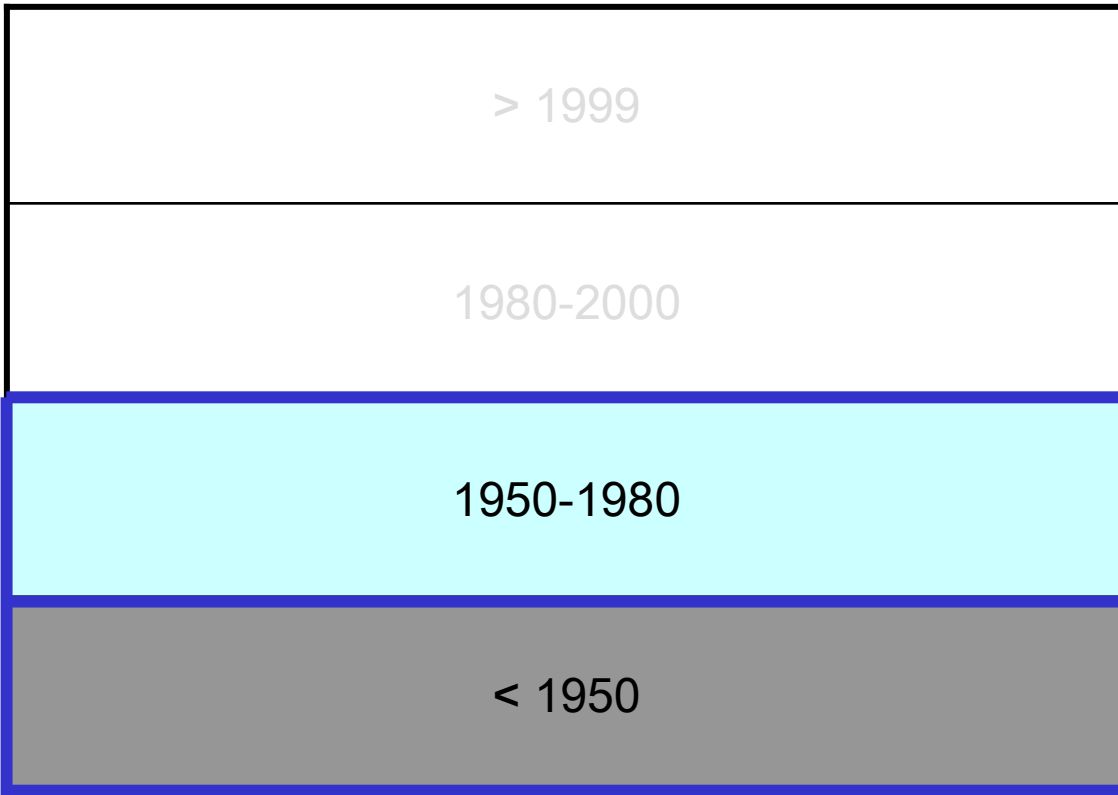
## Results

|                           | Value  | Std.Error | DF  | t-value | p-val |
|---------------------------|--------|-----------|-----|---------|-------|
| (Intercept)               | 0.426  | 0.064     | 755 | 6.670   | 0.000 |
| conc_lexfield_2Recr_Sport | -0.041 | 0.102     | 95  | -0.404  | 0.687 |
| conc_lexfield_2Media_IT   | -0.010 | 0.096     | 95  | -0.100  | 0.921 |
| conc_lexfield_2Relat_Sex  | -0.111 | 0.074     | 755 | -1.495  | 0.135 |
| conc_lexfield_2Other      | -0.419 | 0.085     | 755 | -4.920  | 0.000 |
| word_age1                 | 0.103  | 0.025     | 755 | 4.214   | 0.000 |
| word_age2                 | 0.021  | 0.010     | 755 | 2.096   | 0.037 |
| word_age3                 | -0.017 | 0.012     | 755 | -1.416  | 0.157 |
| conc_innovYES             | 0.251  | 0.039     | 755 | 6.443   | 0.000 |
| word_syll_2>2syll         | -0.101 | 0.032     | 755 | -3.164  | 0.002 |



# Age Word

Higher success



More success:  
WW II-effect

Lower success

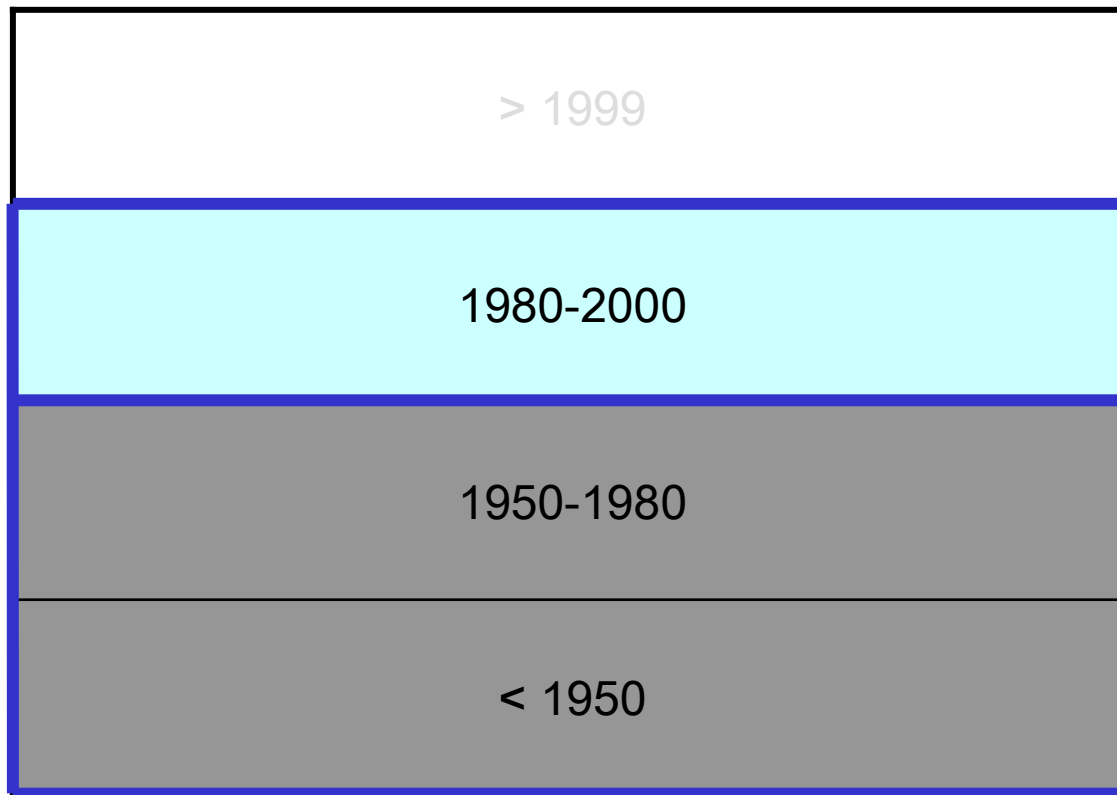


# Age Word

Higher success



Lower success



More success:  
IT-effect

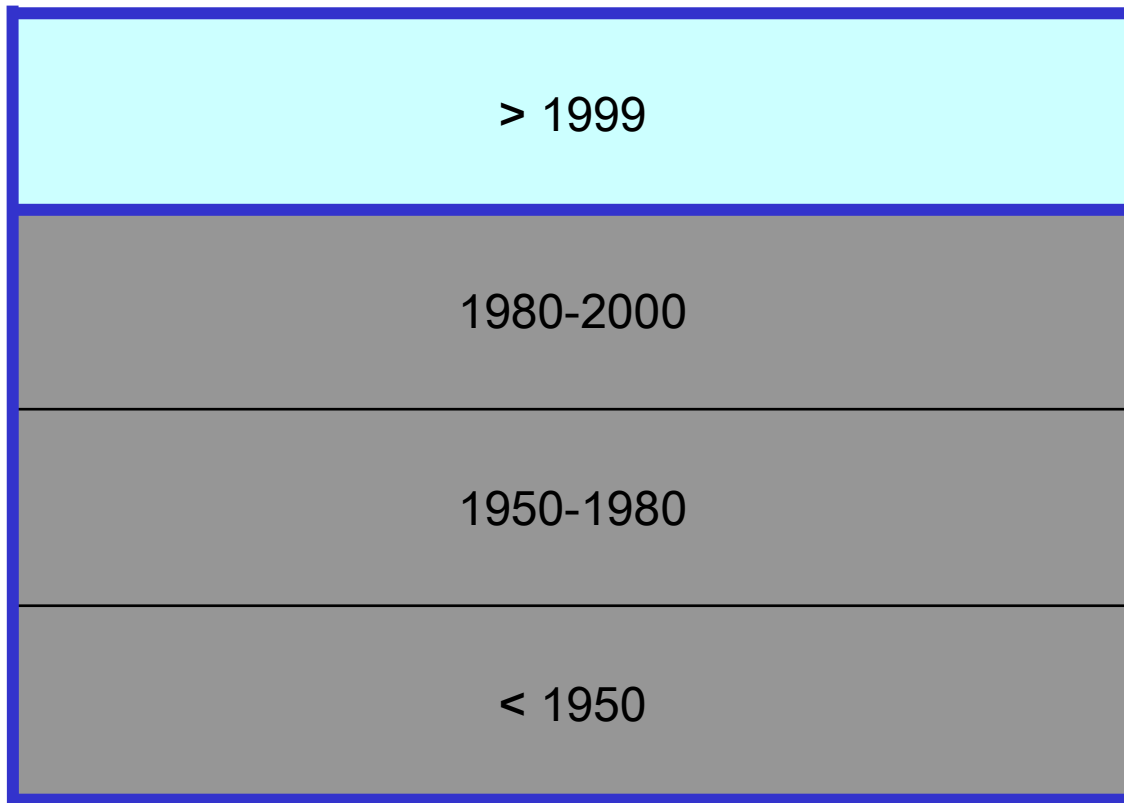


# Age Word

Higher success



Lower success



not significant:  
too young?



## Results

|                           | Value  | Std.Error | DF  | t-value | p-val |
|---------------------------|--------|-----------|-----|---------|-------|
| (Intercept)               | 0.426  | 0.064     | 755 | 6.670   | 0.000 |
| conc_lexfield_2Recr_Sport | -0.041 | 0.102     | 95  | -0.404  | 0.687 |
| conc_lexfield_2Media_IT   | -0.010 | 0.096     | 95  | -0.100  | 0.921 |
| conc_lexfield_2Relat_Sex  | -0.111 | 0.074     | 755 | -1.495  | 0.135 |
| conc_lexfield_2Other      | -0.419 | 0.085     | 755 | -4.920  | 0.000 |
| word_age1                 | 0.103  | 0.025     | 755 | 4.214   | 0.000 |
| word_age2                 | 0.021  | 0.010     | 755 | 2.096   | 0.037 |
| word_age3                 | -0.017 | 0.012     | 755 | -1.416  | 0.157 |
| conc_innovYES             | 0.251  | 0.039     | 755 | 6.443   | 0.000 |
| word_syll_2>2syll         | -0.101 | 0.032     | 755 | -3.164  | 0.002 |



# Innovation

English words introducing a new concept (*webmaster*)

>>>

English word that are borrowed as extra lexicalization (*designer*)

# Innovation

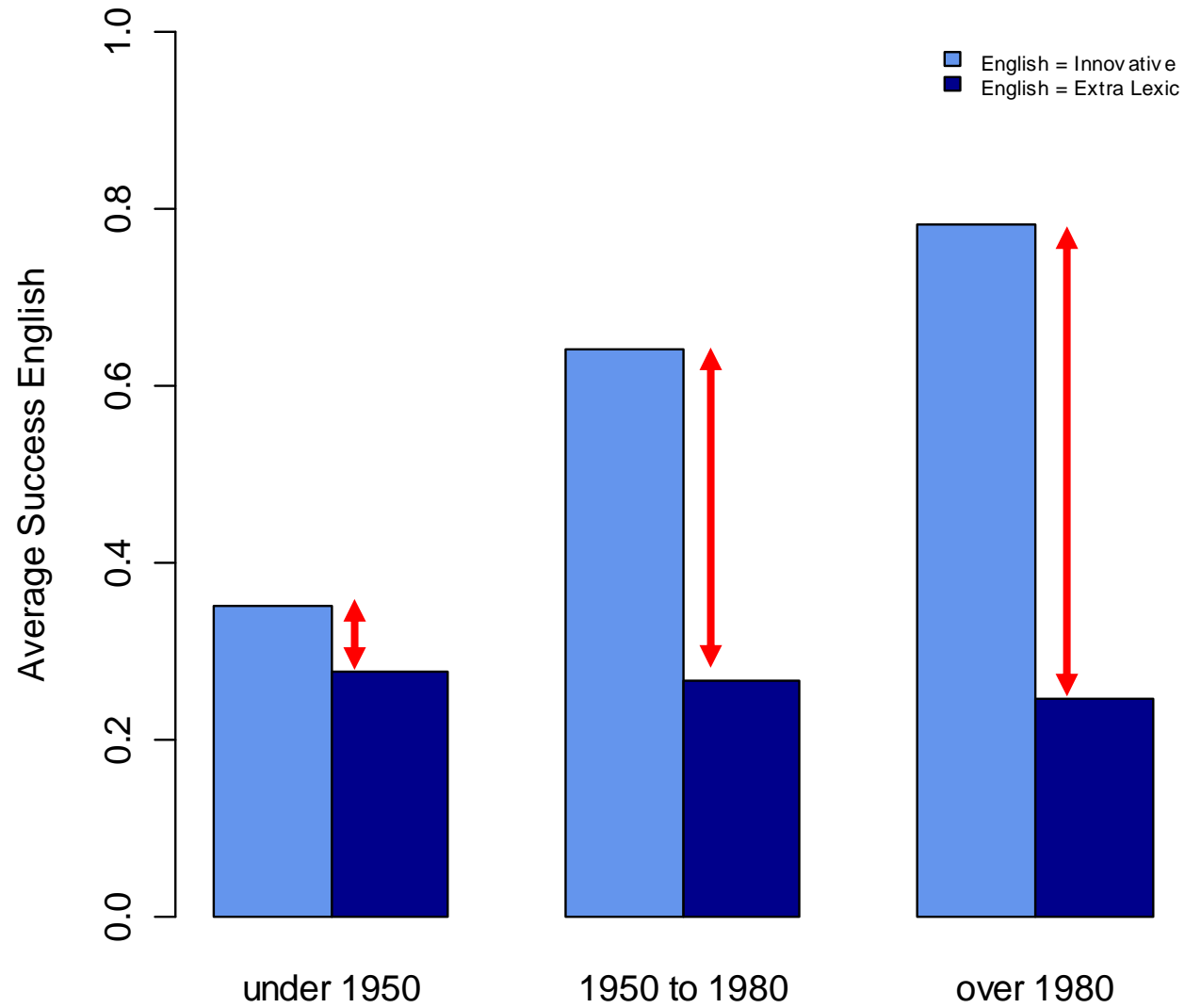
English words introducing a new concept (*webmaster*)

>>>

English word that are borrowed as extra lexicalization (*designer*)

MIND:  
interaction word-age

## Innovation x Age Word





# Innovation \* Word Age

## 1. “Fair Battle”

older words: Dutch alternative has had a fair chance to establish itself

## 2. Change in Language Policies

older words: Dutch alternative has been introduced more “ardently”

## Results

|                           | Value  | Std.Error | DF  | t-value | p-val |
|---------------------------|--------|-----------|-----|---------|-------|
| (Intercept)               | 0.426  | 0.064     | 755 | 6.670   | 0.000 |
| conc_lexfield_2Recr_Sport | -0.041 | 0.102     | 95  | -0.404  | 0.687 |
| conc_lexfield_2Media_IT   | -0.010 | 0.096     | 95  | -0.100  | 0.921 |
| conc_lexfield_2Relat_Sex  | -0.111 | 0.074     | 755 | -1.495  | 0.135 |
| conc_lexfield_2Other      | -0.419 | 0.085     | 755 | -4.920  | 0.000 |
| word_age1                 | 0.103  | 0.025     | 755 | 4.214   | 0.000 |
| word_age2                 | 0.021  | 0.010     | 755 | 2.096   | 0.037 |
| word_age3                 | -0.017 | 0.012     | 755 | -1.416  | 0.157 |
| conc_innovYES             | 0.251  | 0.039     | 755 | 6.443   | 0.000 |
| word_syll_2>2syll         | -0.101 | 0.032     | 755 | -3.164  | 0.002 |



# Innovation

Short Words

>>>

Long Words

# Innovation

Short Words

>>>

Long Words

Region:

Only significant in Flanders

Different proficiency level?

# Innovation

Short Words

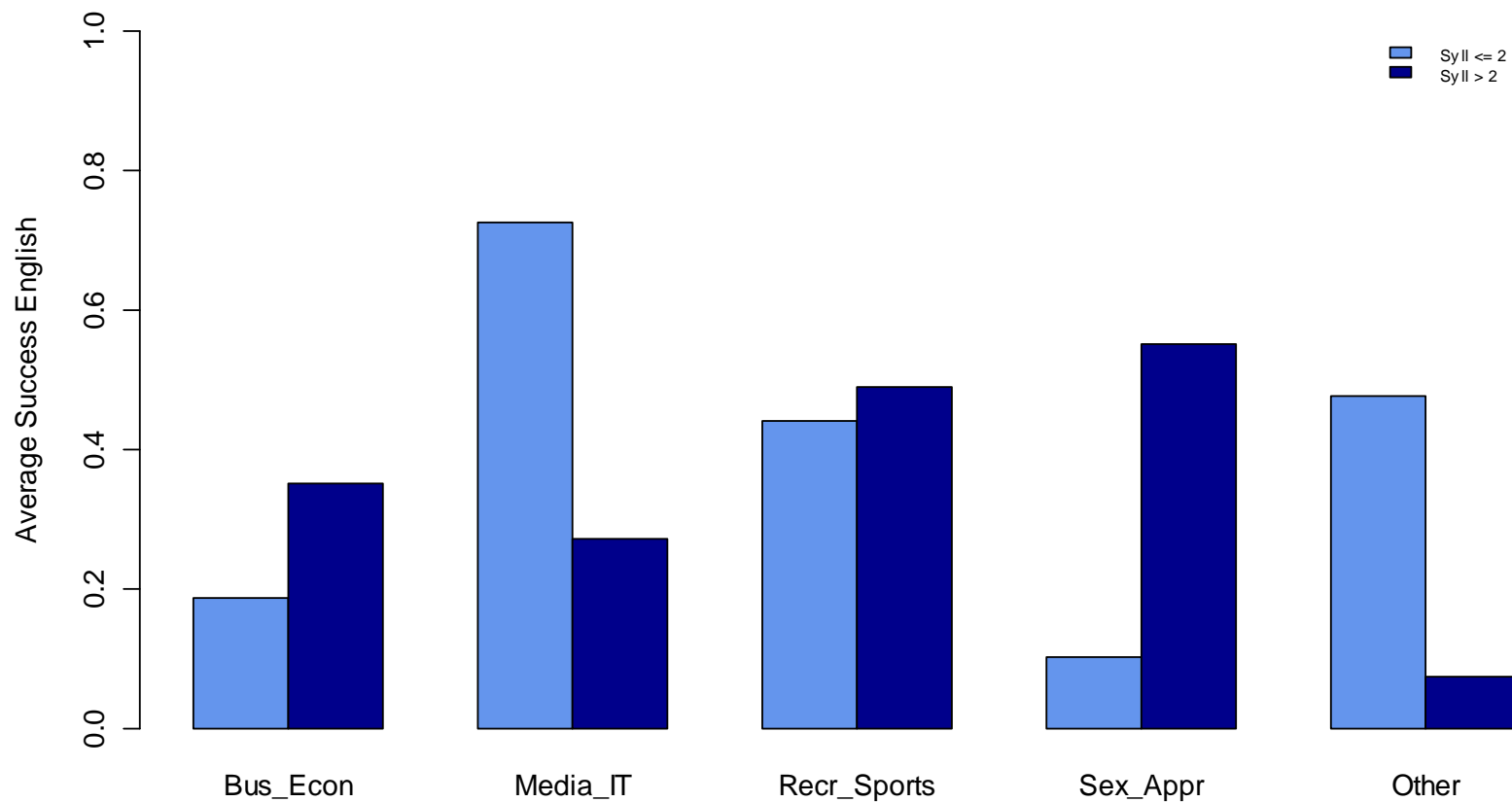
>>>

Long Words

**CAUTION:**  
artefact of other predictors?

# Syllables Word \* LexField

Lex Field x Syllables



# Success of English Person Reference nouns in Dutch

- Theoretical Background
- Case Study: Person Reference Nouns
  - What?
  - Method
  - Data-collection
  - Variables
  - Results
- Conclusions & Prospects



# Conclusion

- concept-related features
  - lexical field: smaller effect than has been presumed in descriptive analyses
- word-related features
  - age words, in interaction with conceptual innovation
  - caution with formal features!
- lectal features
  - register & year: no effect
  - region: some effects, but rather limited





# Prospects

## 1. lectal variation:

- (1) other statistical techniques that help us explain the behaviour of all 98 concepts (vs. presented technique, which tries to generalize over the concepts)
- (2) expanding to other registers and bigger time-span

## 2. semantic and stylistic specialization:

- not possible for all 98 concepts, but well-chosen case-studies
- in-depth analysis of e.g. JUNKIE, by means of logistic regression analysis



For more information:

<http://www.ling.arts.kuleuven.be/qlvl>

[eline.zenner@arts.kuleuven.be](mailto:eline.zenner@arts.kuleuven.be)