



# Adnominal gender assignment to English loanwords in Belgian Dutch

Karliën Franco, Eline Zenner, Dirk Speelman



RU Quantitative Lexicology and Variational Linguistics

# Overview

1. Background
2. Data and methods
3. Results
4. Conclusion



# Overview

1. Background

2. Data and methods

3. Results

4. Conclusion



# Gender assignment to loanwords

Corbett 1991:

- gender is assigned to loanwords according to the normal gender rules of the target language
- the gender of loanwords is stable

# Gender assignment to loanwords

Corbett 1991:

- gender is assigned to loanwords according to the normal gender rules of the target language
- the gender of loanwords is stable

# Loanwords follow the normal gender system

	<b>masculine</b>	<b>feminine</b>	<b>neuter</b>
definite article	de	de	het
personal pronoun	hij	ze	het
possessive pronoun	zijn	haar	zijn

# Loanwords follow the normal gender system

	<b>masculine</b>	<b>feminine</b>	<b>neuter</b>
definite article	de	de	het
personal pronoun	hij	ze	het
possessive pronoun	zijn	haar	zijn

# Loanwords follow the normal gender system

- normal gender system:
  - **semantic:** sports and games → het  
→ **het** voetbal ('soccer')
  - **formal:** suffix '-ment' → het  
→ **het** instrument ('instrument')



# Loanwords follow the normal gender system

- normal gender system:
  - **semantic:** sports and games → het  
→ **het** voetbal ('soccer')
  - **formal:** suffix '-ment' → het  
→ **het** instrument ('instrument')
- loanwords get 'de' unless there is a reason for using 'het':
  - **de** interface, **de** floppy
  - BUT:**
  - **het** rugby, **het** management

# Loanwords follow the normal gender system

- additional rule for loanwords:
  - analogy with Dutch cognate or translation
    - het billboard ~ het bord ('the board')

# Gender assignment to loanwords

Corbett 1991:

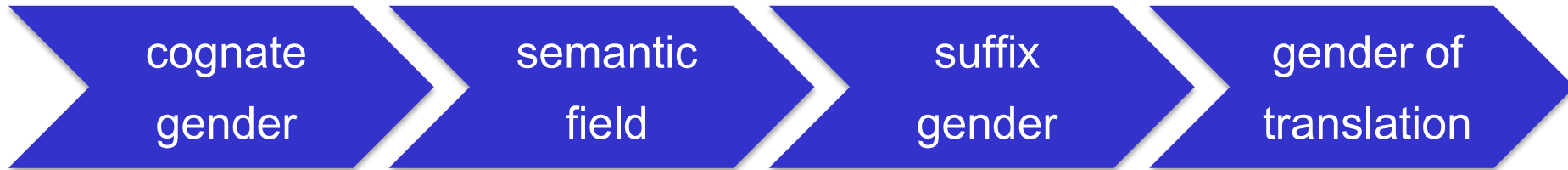
- gender is assigned to loanwords according to the normal gender rules of the target language
- the gender of loanwords is stable

# The gender of loanwords is stable

- previous research:
    - often corpus studies or lists of anglicisms:
      - assume that all language users assign the same gender
        - Pollack et al. (1982): for well-established nouns
- hierarchy of influential factors (with list of exceptions)

# Loanwords follow the normal gender system

- Geerts (1996): hierarchy for anglicisms in Dutch



conflicts are resolved by place in the hierarchy

e.g. 'gingerbeer':      cognate 'het bier'  
                                 semantic field: drinks ('de')

→ **het** gingerbeer

# Case-study on anglicisms in Dutch

- **few systematic studies** on the gender of anglicisms in Dutch
  - exception: Schenck 1985
- **variation** in the pronominal gender of Dutch
  - in our data: only 66% syntactic agreement

# Goal

to determine which factors are important for gender assignment to anglicisms in Belgian Dutch:

- using data from a questionnaire rather than a corpus or a list of words
- using inferential statistics and multifactorial analyses rather than raw frequency or basic proportions
- taking homogeneity between subjects into account

# Goal

to determine which factors are important for gender assignment to anglicisms in Belgian Dutch:

- using data from a questionnaire rather than a corpus or a list of words
- using inferential statistics and multifactorial analyses rather than raw frequency or basic proportions
- taking homogeneity between subjects into account



# Goal

to determine which factors are important for gender assignment to anglicisms in Belgian Dutch:

- using data from a questionnaire rather than a corpus or a list of words
- using inferential statistics and multifactorial analyses rather than raw frequency or basic proportions
- taking homogeneity between subjects into account

# Homogeneity

We will show that:

- respondents agree about the gender of some anglicisms
- for other anglicisms, conflicts between the predictors lead to variation between and within respondents

# Overview

1. Background

2. Data and methods

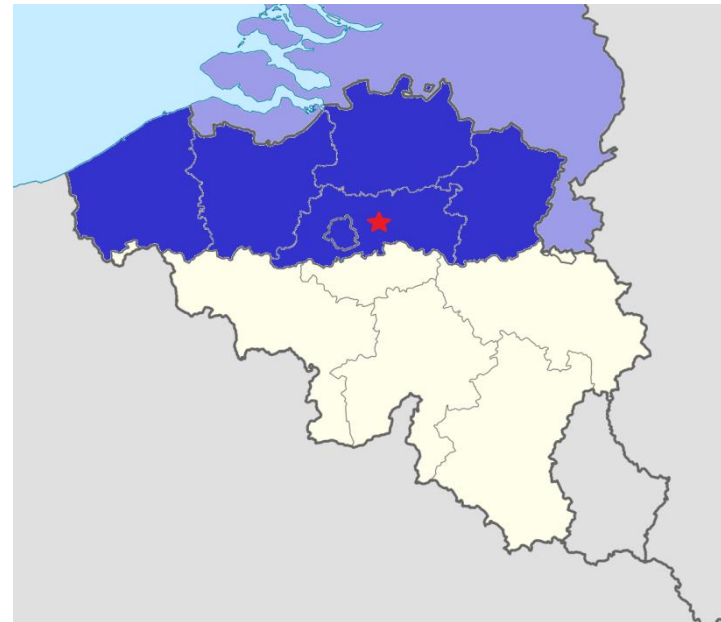
3. Results

4. Conclusion



# Questionnaire

180 sentences, each containing one anglicism



# Questionnaire

two tasks per sentence:

1. article: de / het
2. pronoun: hij / ze / het

5. (De / Het) **tenniscoach** is maandag op 70-jarige leeftijd overleden (Hij / Ze / Het) ontdekte tijdens zijn carrière Steffi Graf en Boris Becker.
- Ik ken dit woord niet.
6. Een stel avonturiers brengen er per ongeluk (de / het) **mummy** tot leven en (hij / ze / het) duikt daarna op in Londen.
- Ik ken dit woord niet.

# Questionnaire

two tasks per sentence:

1. article: de / het
2. pronoun: hij / ze / het

5. (De / Het) **tenniscoach** is maandag op 70-jarige leeftijd overleden (Hij / Ze / Het) ontdekte tijdens zijn carrière Steffi Graf en Boris Becker.
- Ik ken dit woord niet.
6. Een stel avonturiers brengen er per ongeluk (de / het) **mummy** tot leven en (hij / ze / het) duikt daarna op in Londen.
- Ik ken dit woord niet.

# Questionnaire

180 sentences, each containing one anglicism:

new anglicisms:

not yet in Van Dale<sup>14</sup>

based on two large news corpora (TwNC and LeNC)

established anglicisms:

included in Van Dale<sup>14</sup>

borrowed in 1950s or later

recognizable as English loans

# Questionnaire: anglicisms

development	stream	researcher	comedian	vibe	cooler	scorecard
facelift	contest	majorette	popart	freebee	tracker	showbiz
dress	lovestory	quizzer	party	empathie	queen	surround
reality	baseballbat	network	diskette	rate	career day	battle
tenniscoach	witch	showcase	briefing	fairness	steeple	image
mummy	sit-in	spot	shoppingcenter	marketing	combo	jogger
entertainment	believer	computersoftware	callgirl	sticker	soundsystem	channel
suspense	bodypainting	noise	youngster	package	mall	clash
copy	trench	suburb	franchising	inspector	tuning	resource
release	wall	escalatie	voucher	click	patchwork	supercomputer
statement	printing	tennismatch	demand	opportunity	babyfoon	creditcard
hippie	leisure	concept car	squat	braindrain	input	hotelmanager
carpooling	sensor	supermarkt	ace	black-out	publicity	it-girl
aquaplaning	error	technology	roadbook	safety	underdog	wash
busticket	pay	snowboarder	recycling	member	key	cover
exit-poll	spray	disc	boogie	panty	award	nanny
catsuit	establishment	countdown	engine	draft	check-in	cashflow
chick	camping ground	tumbling	hometown	view	gangster	nickname
sneak	assist	sound	guesthouse	freelancester	rock'n'roll	container ship
achievement	skateboard	shooting	supply	beach	jingle	gossip
extra-time	intercom	chill-out	hospitality	shareholder	smog	screen
blooper	search	stationwagon	track	supermodel	column	
knowhow	gallery	indian summer	e-learning	showtime	theory	
teamspirit	jack	slam	hovercraft	town	sense of urgency	
dinner	dancefloor	character	football	security	publishing	
American dream	basketbal	bulldog	talkshow	loafer	director	





# Questionnaire

180 sentences, each containing one anglicism:

new anglicisms:

not yet in Van Dale<sup>14</sup>

based on two large news corpora (TwNC and LeNC)

established anglicisms:

included in Van Dale<sup>14</sup>

borrowed in 1950s or later

recognizable as English loans

2 randomizations

# Respondents

## 45 university students

- mean age 21.49
- gender 35 female  
10 male
- 3 Flemish regions: 24 Antwerp or Flemish Brabant  
13 Limburg  
8 East or West Flanders
- profession level of parents: 27 high  
18 low
- 26 know a local dialect
- 25 have a good knowledge of English

# Methodology

- two types of research questions:
  1. Which factors influence the gender of anglicisms?
  2. Do subjects always agree about the gender of anglicisms?

# Methodology

- two types of research questions:
  1. Which factors influence the gender of anglicisms?
  2. Do subjects always agree about the gender of anglicisms?

# Determining which factors influence gender

- response variable:
  - **article:** de / het
- analyses
  - **exploratory descriptive** analyses:
    - associations between variables
    - association between predictor and response
  - **heatmaps:** gender assignment per type per respondent
  - **conditional inference trees and random forests**
  - **mixed effects logistic regression**

# Predictors

- language-external variables: **respondent information**
- gender **cognate or translation**: 133 de  
42 het
- **semantic field**: 27 animate  
148 other
- **suffix gender** (if present): 40 de  
5 het

# Predictors

- degree of establishment: 

66	established & 'de'
12	established & 'het'
49	new frequent
48	new infrequent

→ modelling two questions:

- Which factors are important for gender assignment?
- Do language users agree with the established gender?

# Methodology

- two types of research questions:
  1. Which factors influence the gender of anglicisms?
  2. Do subjects always agree about the gender of anglicisms?



# Methodology

- two types of research questions:
  1. Which factors influence the gender of anglicisms?
  2. Do subjects always agree about the gender of anglicisms?
    - How can we measure homogeneity?

# Measuring homogeneity

- measure of global internal uniformity (Geeraerts et al. 1999):

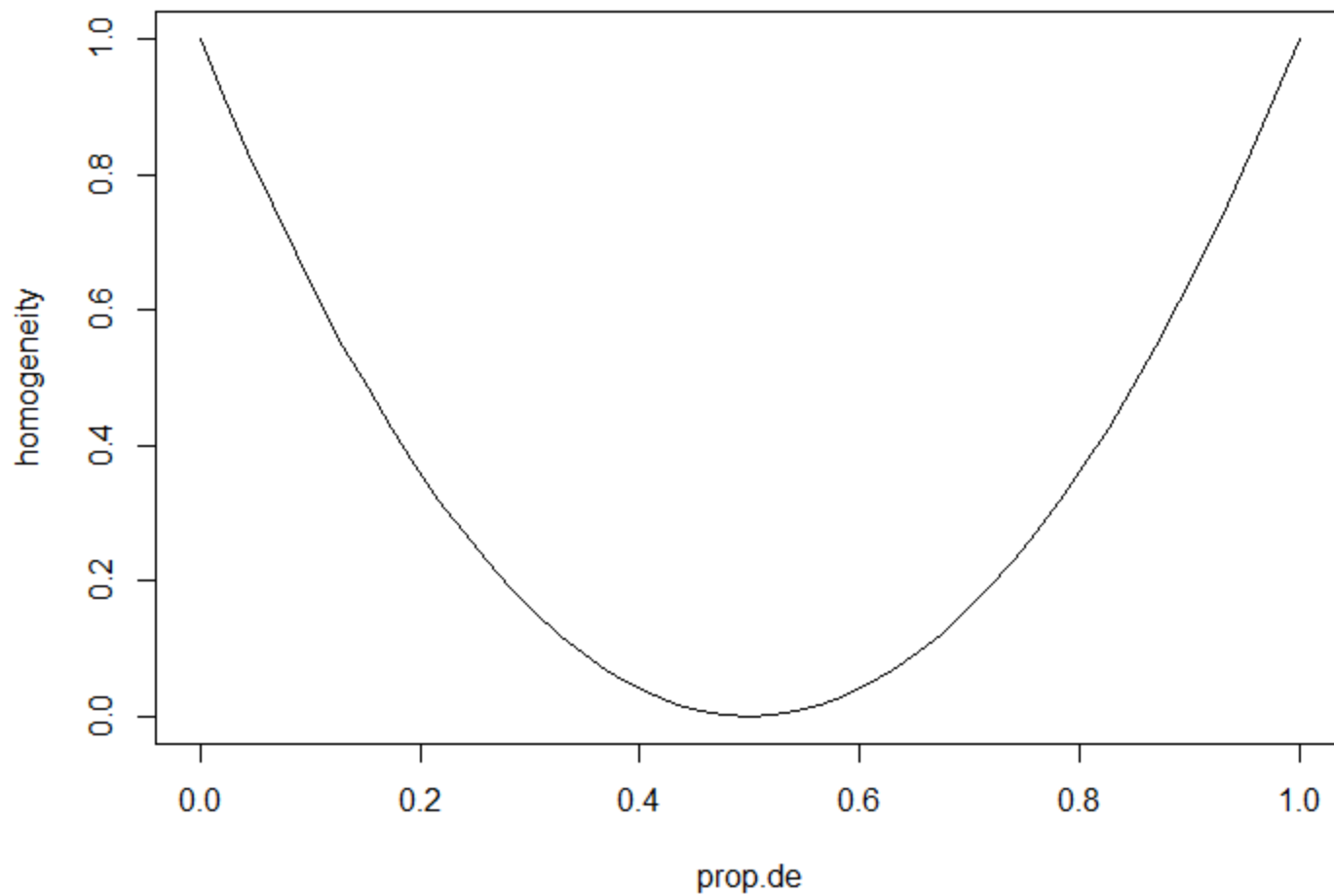
$$\begin{aligned} \text{homogeneity per type} &= I_Z(Y) = \sum_{i=1}^n F_{Z,Y}(x_i)^2 \\ &= \text{proportion.de}^2 + \text{proportion.het}^2 \end{aligned}$$

# Measuring homogeneity

- measure of global internal uniformity (Geeraerts et al. 1999):

$$\begin{aligned} \text{homogeneity per type} &= I_Z(Y) = \sum_{i=1}^n F_{Z,Y}(x_i)^2 \\ &= \text{proportion.de}^2 + \text{proportion.het}^2 \end{aligned}$$

transformation:    0 = no homogeneity  
                          1 = complete homogeneity



# Measuring homogeneity

- measure of global internal uniformity (Geeraerts et al. 1999):

$$\begin{aligned}\text{homogeneity per type} &= I_Z(Y) = \sum_{i=1}^n F_{Z,Y}(x_i)^2 \\ &= \text{proportion.de}^2 + \text{proportion.het}^2\end{aligned}$$

transformation:    0 = no homogeneity  
                          1 = complete homogeneity

- analyses:
  - exploratory descriptive analyses
  - linear regression

# Overview

1. Background

2. Data

3. Results

A. Factors that influence gender

B. Homogeneity

4. Conclusion



# Factors that influence gender

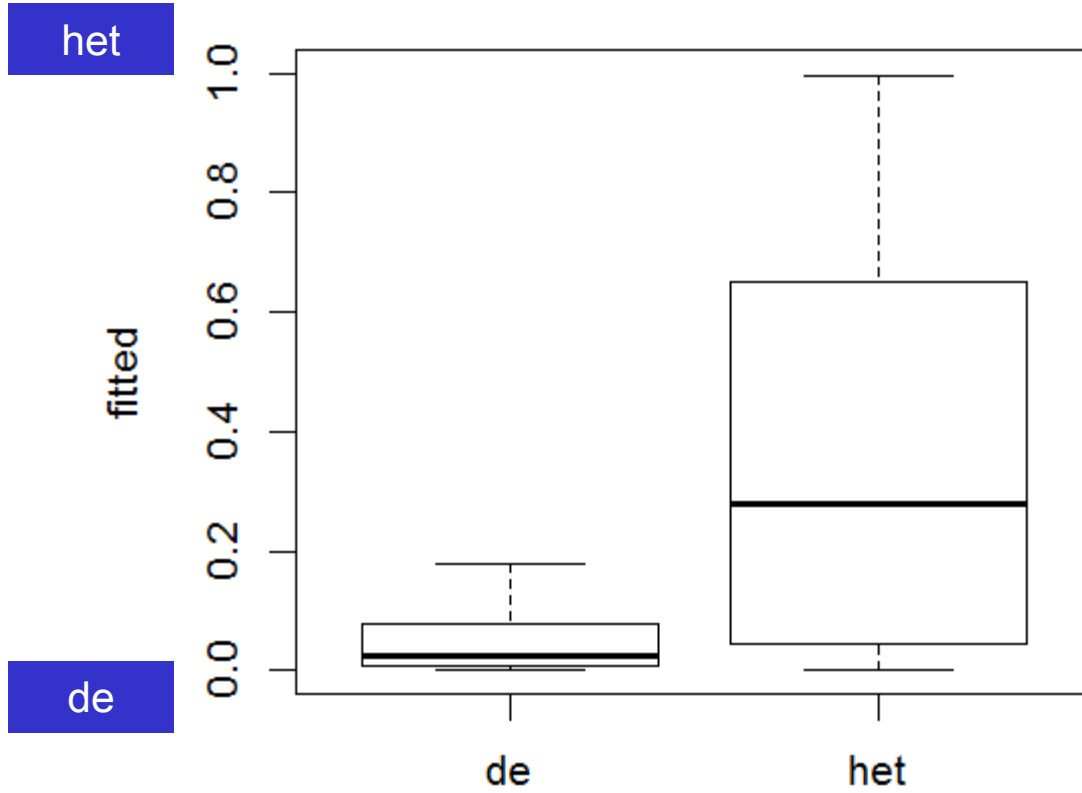
- preference for 'de' (15.02% 'het')

# Factors that influence gender

- preference for 'de' (15.02% 'het')
- mixed effects logistic regression:
  - random effects: type, respondent
  - $C = 0.93$
  - significant predictors:
    - gender cognate or translation
    - degree of establishment
    - semantic field



## gender of cognate or translation



het

de

fitted

de

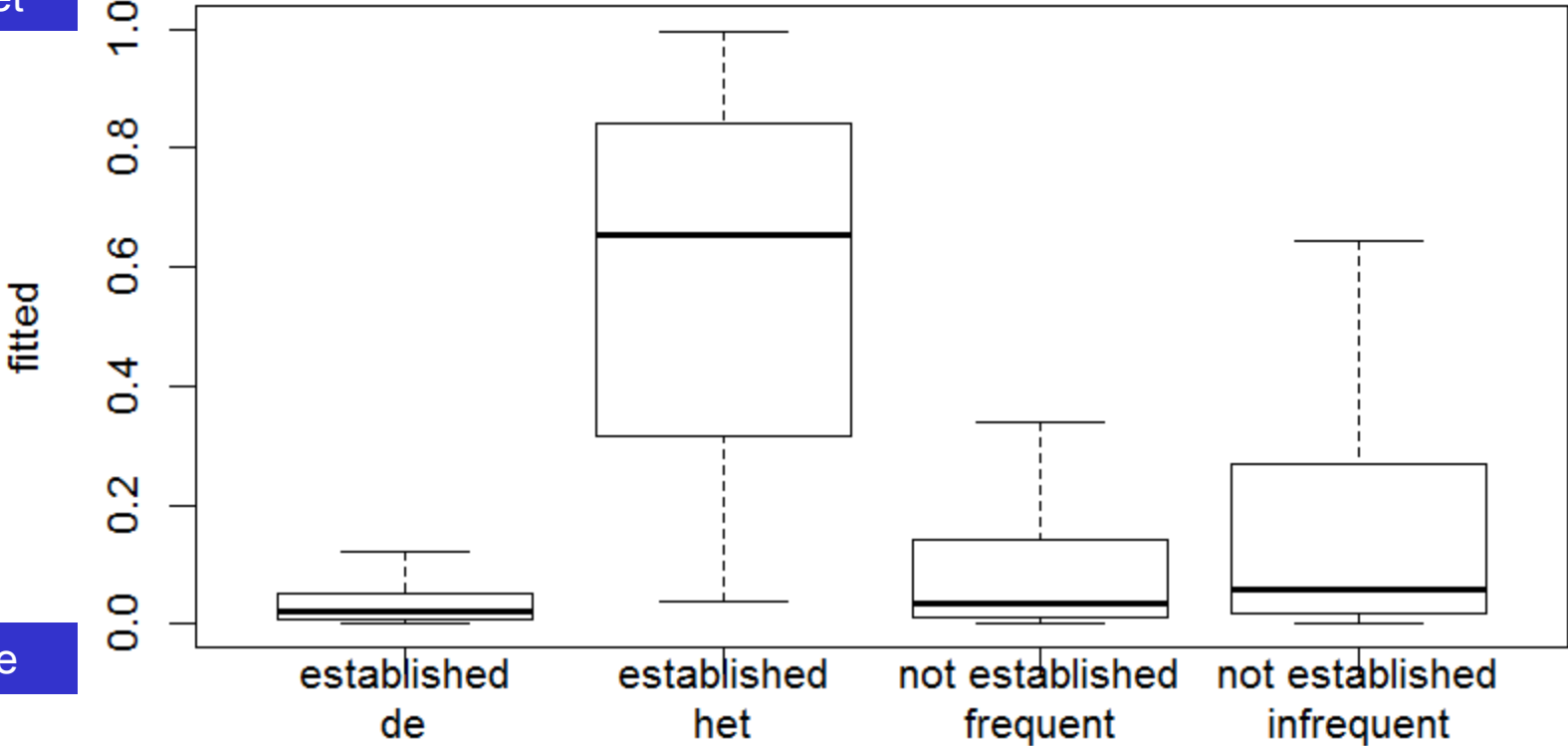
het



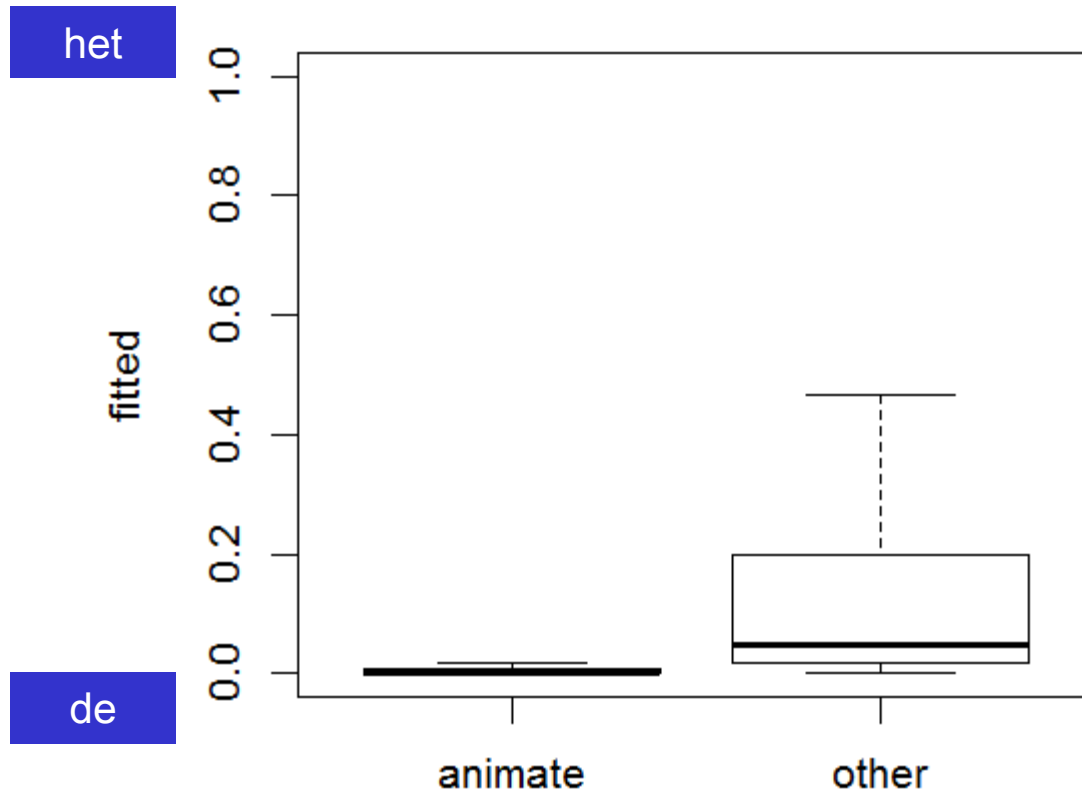
# degree of establishment

het

de



## semantic field



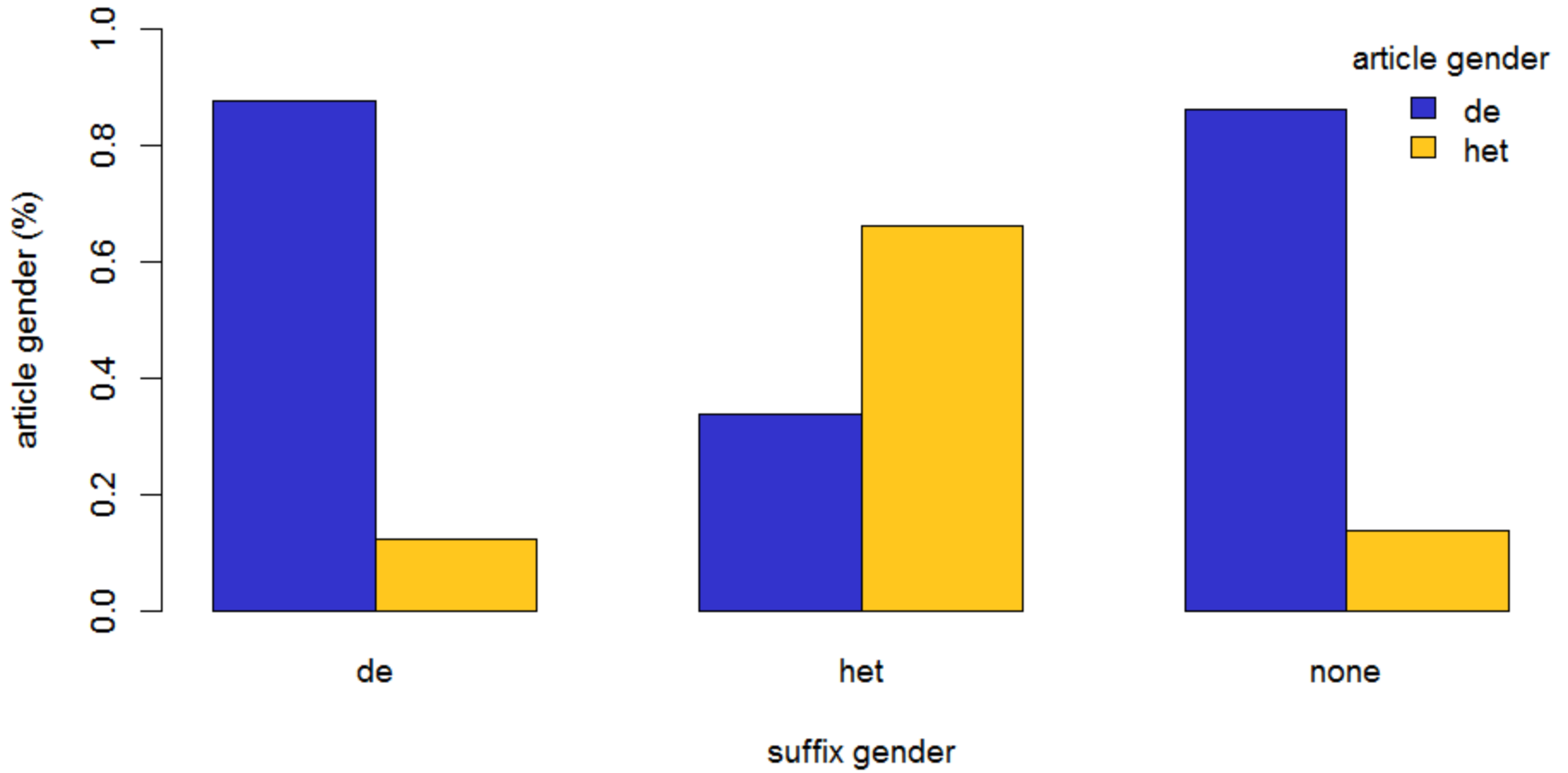
het

de



# Factors that influence gender

- suffix gender
  - significant if included in mixed model
- data sparsity neuter suffixes:
  - only 5 types: achievement, development, entertainment  
establishment, statement
  - only suffix –ment



# Overview

1. Background

2. Data

3. Results

A. Factors that influence gender

B. Homogeneity

4. Conclusion



# Homogeneity?

- complete homogeneity: everyone always uses the same variant for each concept
- Do subjects agree about gender of the anglicisms?





# Homogeneity?

- complete homogeneity: everyone always uses the same variant for each concept
- Do subjects agree about gender of the anglicisms?

→NO

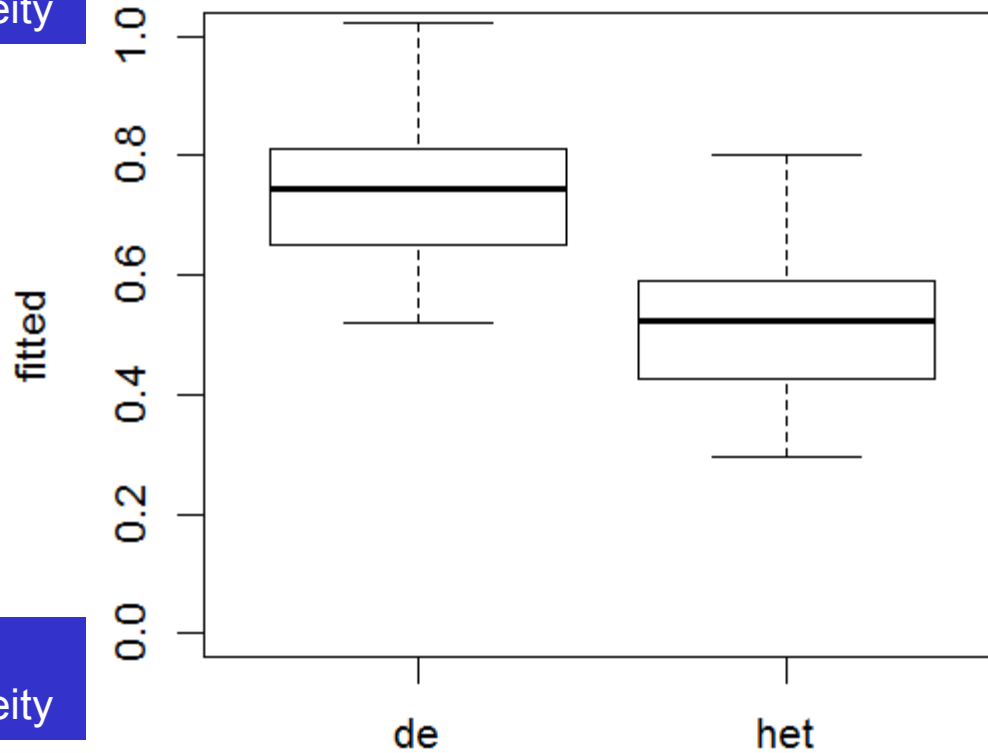
- Which factors promote homogeneity between subjects?

## Article: homogeneity

- mean homogeneity = 70.85%
- linear regression model
  - **response** variable: homogeneity
  - $R^2$ : 0.2654
- significant **predictors**:
  - gender of Dutch cognate or translation
  - degree of establishment
  - semantic field

complete  
homogeneity

## gender of Dutch cognate or translation



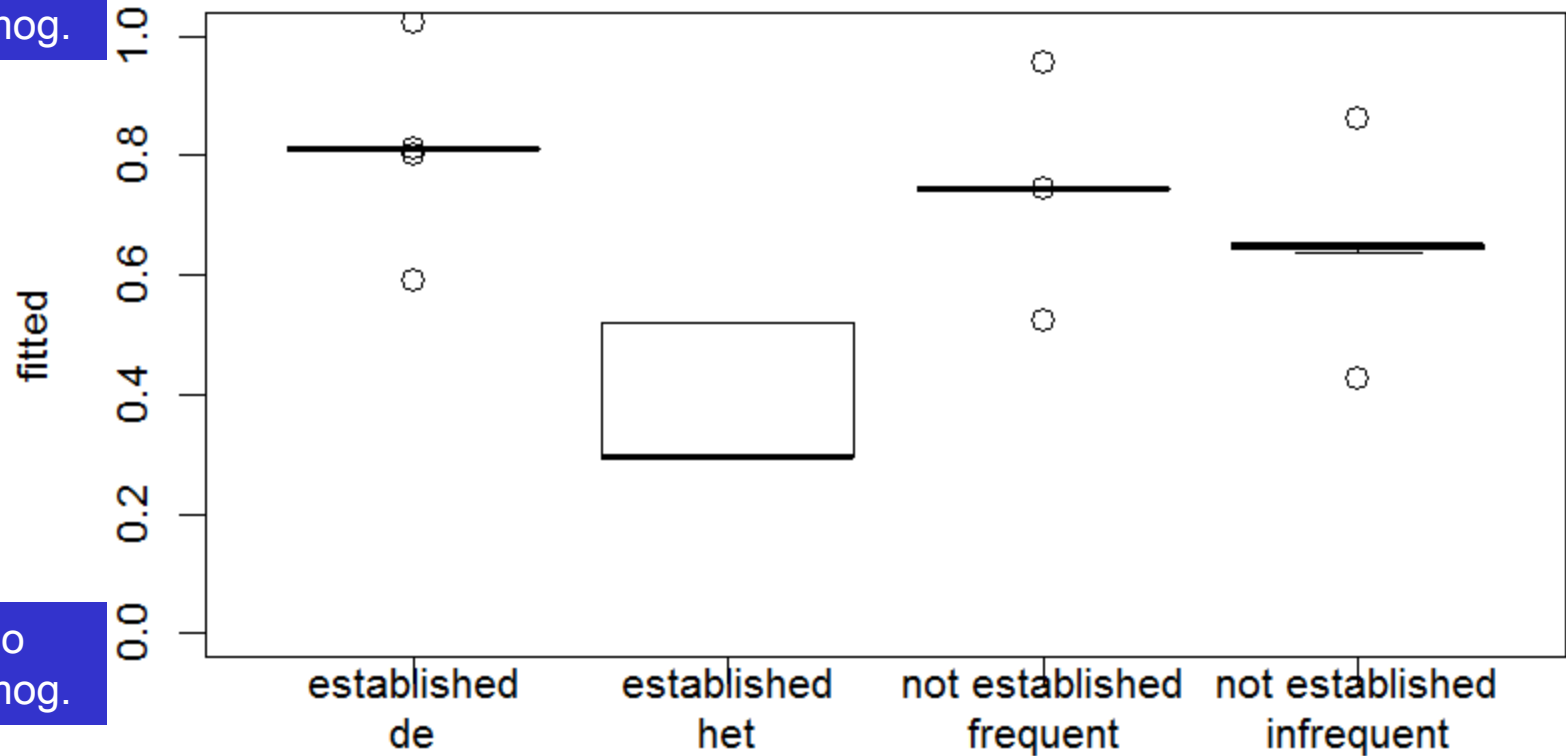
no  
homogeneity



# degree of establishment

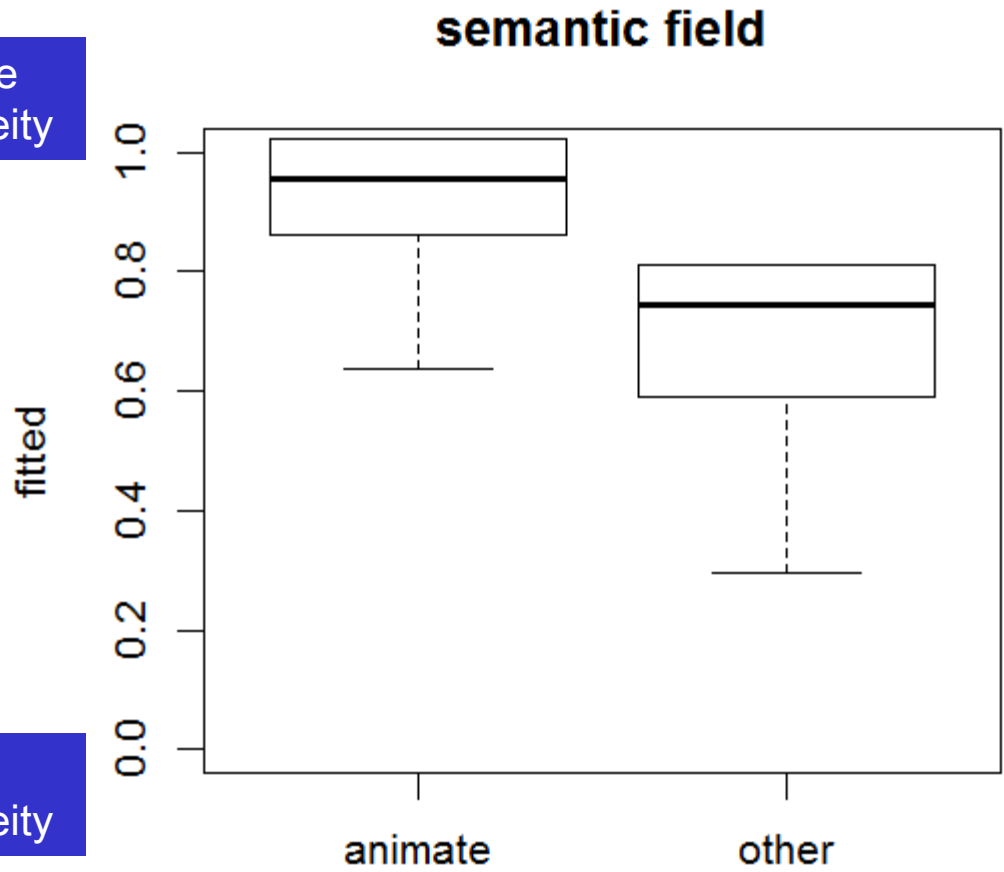
complete  
homog.

no  
homog.



complete  
homogeneity

no  
homogeneity



# Overview

1. Background

2. Data

3. Results

A. Factors that influence gender

B. Homogeneity

4. Conclusion



# Conclusion

de as default article

1. Which factors influence the gender of anglicisms?
2. Do subjects always agree about the gender of anglicisms?

# Conclusion

de as default article

## 1. Which factors influence the gender of anglicisms?

- gender of **cognate or translation** in source language
- **semantic** properties
- **suffix gender**
- **established gender** can serve as predictor, but no unanimity

## 2. Do subjects always agree about the gender of anglicisms?



# Conclusion

de as default article

1. Which factors influence the gender of anglicisms?

2. Do subjects always agree about the gender of anglicisms?

gender assignment to anglicisms is:

- not homogeneous
- not rule-based

conflicting factors: hierachy → less homogeneity



Thank you!

for further information:

[karlien.franco@kuleuven.be](mailto:karlien.franco@kuleuven.be)

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

