Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Freie Universität Berlin

Particles and Priming –

Combining sociolinguistic and psycholinguistic determinants of variation

Martin Schweinberger

Universität Hamburg
martin.schweinberger@uni-hamburg.de

Freie Universität Berlin
martin.schweinberger@fu-berlin.de

Setting the stage

Variationist Sociolinguistics

- Language is not homogeneous: variation is ubiquitous
- Social factors : language use
  - Linguistic variation not random
  - Systematic correlation between certain social factors (age, gender, class, ethnicity, etc.) and language use
- Linguistic differentiation : social stratification

Setting the stage

"[M]ost of the linguistic changes in progress studied in the 2nd half of the 20th century show a high degree of social differentiation" (Labov 2002)

Setting the stage

Variationist Sociolinguistics

- Problem

  "[GoldVarb] is one of the predominant data analysis tools used in sociolinguistics, employed successfully for over three decades. […] [There exists] a more serious problem whereby GoldVarb overestimates the significance of effects." (Johnson 2009:359)

- Stratification = statistical artefact?

Research Question

Q1: Can advanced statistical methods (mixed-effects regression models) enhance our understanding of linguistic variation?

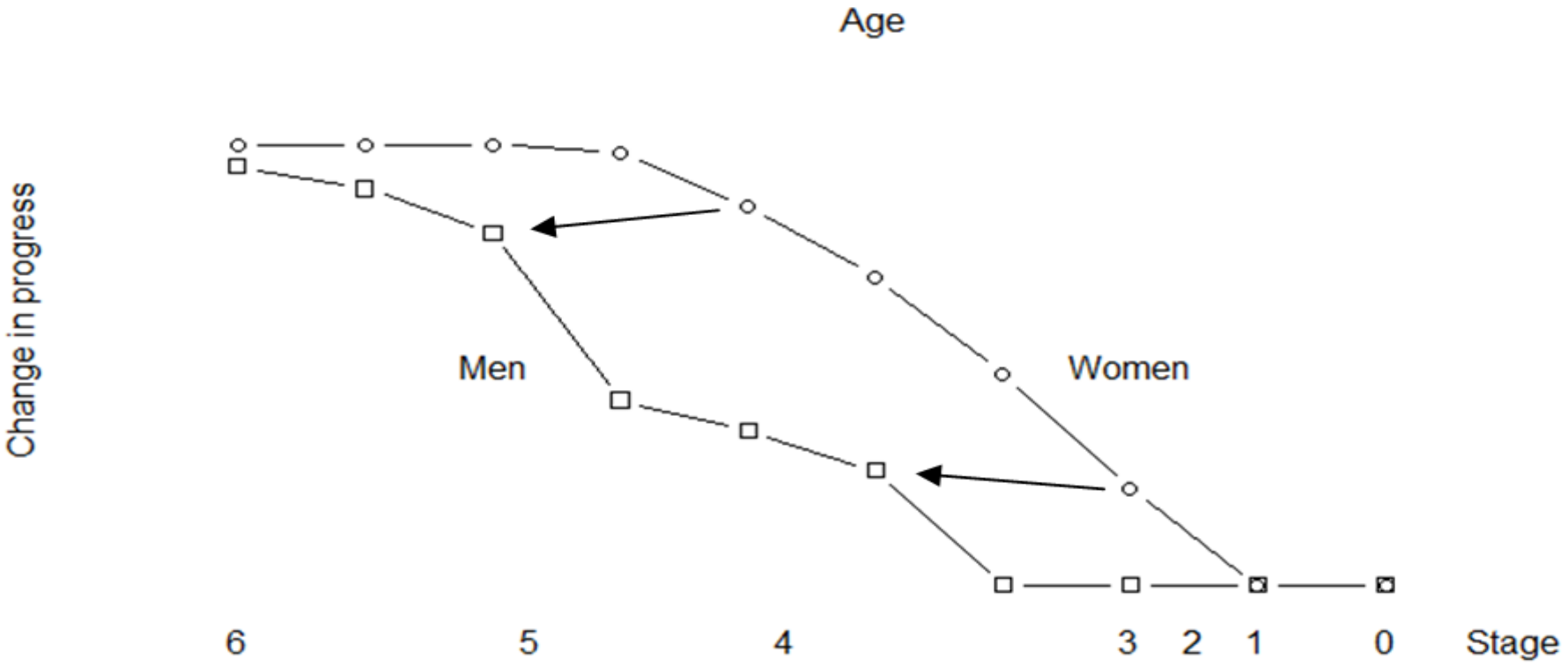
- Research question
- Theoretical framework
- The phenomenon
- Data
  (summary statistics & data plotting,
  statistical designs & variable coding)
- Results
- Discussion, Conclusion & Outlook

- Research question

Theoretical framework

- The phenomenon
- Data
  (summary statistics & data plotting,
  statistical designs & variable coding)
- Results
- Discussion, Conclusion & Outlook

Theoretical framework



Six-stage model of gender relations in linguistic change from below (Labov 1994: 65)

Theoretical framework

Modern sociolinguistic theory

- Age
  - Young before old
- Gender
  - Leaders of change : adolescent women
  - Women & prestige/stigmatization
    - Status/acts of identity/symbolic means
  - NORMS
- Class
  - Spread from centre (top : bottom much rarer)

Theoretical framework

Modern sociolinguistic theory

- Age
  - Young before old
- Gender
  - Leaders of change : adolescent women
  - Women & prestige/stigmatization
    - Status/acts of identity/symbolic means
  - NORMS
- Class
  - Spread from centre (top : bottom much rarer)

Theoretical framework

Modern sociolinguistic theory

- Age
  - Young before old
- Gender
  - Leaders of change : adolescent women
  - Women & prestige/stigmatization
    - Status/acts of identity/symbolic means
  - NORMS
- Class
  - Spread from centre (top : bottom much rarer)

Theoretical framework

Modern sociolinguistic theory

- Age
  - Young before old
- Gender
  - Leaders of change : adolescent women
  - Women & prestige/stigmatization
    - Status/acts of identity/symbolic means
  - NORMS
- Class
  - Spread from centre (top : bottom much rarer)

Theoretical framework

Modern sociolinguistic theory

- Our current models of language variation and change have the advantage that they are…
  - Based on many studies (highly stable)
  - High predictive and explanatory power
- But these models also have shortcomings…
  - Neglect of language contact and multilingualism
  - Do not take psycholinguistic factors into account
  - Rely on overcome statistical methods (GoldVarb, fixed-effects models)

Theoretical framework

Modern sociolinguistic theory

- Sociolinguists are aware of these shortcomings
  - priming (persistence) (cf. Szmrecsanyi 2006).
  - Critique of classic fixed-effects (GoldVarb) regressions (cf. Johnson 2009)
    - Unreliable results
    - High α-error rate

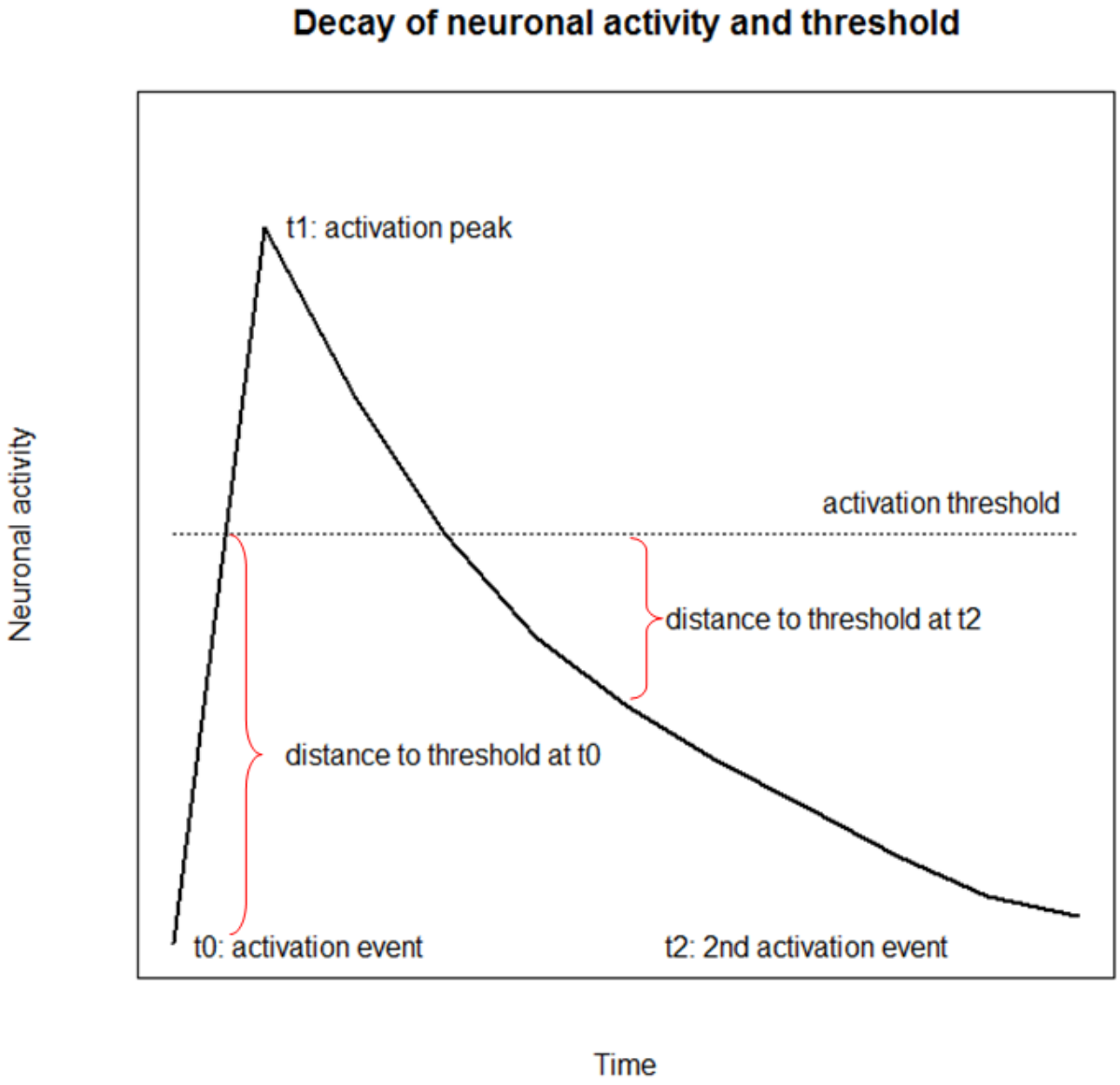
Theoretical framework

Traditional sociolinguistics	*Modern sociolinguistics*	*Tomorrow's sociolinguistics*
	Methodology	
Frequency analysis & bivariate statistics	Multivariate statistics	Sophisticated statistical modelling
(non-)parametric tests	GoldVarb analysis	R
χ2-test, t-test, Wilcoxon Sign-Rank test	Logistic Regression	Generalized linear (mixed-effects) models, PCA, MDS
	Proponents	
Trudgill	Labov	Szmrecsanyi
Chambers	Rickford	?
Preston	Tagliamonte	?

Theoretical framework

Q2: To which degree is the high degree of social stratification of language use an artefact of inappropriate statistics and ignoring psycholinguistic factors?

Priming


Decay of neuronal activity and threshold
t1: activation peak
activation threshold
distance to threshold at t2
distance to threshold at t0
t0: activation event
t2: 2nd activation event
Neuronal activity
Time

- Research question
- Theoretical framework

The phenomenon

- Data
  (summary statistics & data plotting,
  statistical designs & variable coding)
- Results
- Discussion, Conclusion & Outlook

The phenomenon

eh in New Zealand English: Examples

(1) ICE-NZ:S1A-002#B: remind me to ring up early in the morning not early about er … what time does the …
ICE-NZ:S1A-002#B: oh be up there at ten *eh*
ICE-NZ:S1A-002#Q: i'm out of here at nine

(2) ICE-NZ:S1A-004#M: they're frightened that they might stop them sitting states
ICE-NZ:S1A-004#M: it's a really bad buzz *eh*
ICE-NZ:S1A-004#G: oh yeah yeah

(3) ICE-NZ:S1A-004#M: er she's working with piki teaching
ICE-NZ:S1A-004#G: oh is she
ICE-NZ:S1A-004#M: cos you know her *eh*
ICE-NZ:S1A-004#G: marie yeah i know her

The phenomenon

eh in New Zealand English (NZE): Properties

- Typically in turn-final position (non-polar tag).
- Stereotyped in NZE (vernacular NZE).
- Maori > Pakeha (British/European New Zealanders)
  - In-group marker of ethnic identity (Meyerhoff 1994:371)
- Men > women
  - Maori men > (young) Pakeha women > Pakeha men > Maori women
- Working class> middle class
  - Marker of class membership (Stubbe & Holmes 1995:84)

- Research question
- Theoretical framework
- The phenomenon

Data
(bivariate statistics & data plotting,
statistical designs & variable coding)

- Results
- Discussion, Conclusion & Outlook

Data summary

International Corpus of English (ICE)

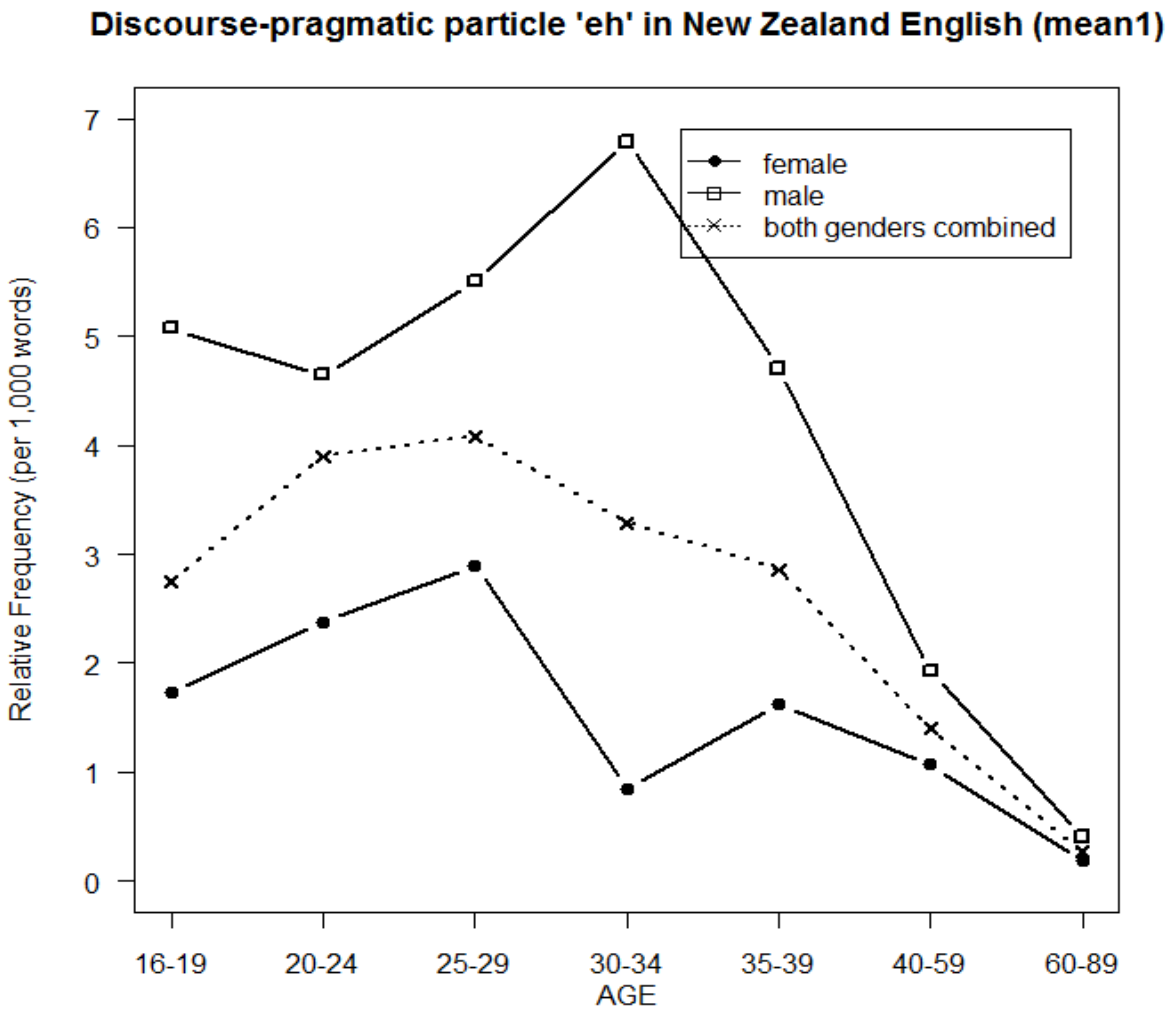
- ICE New Zealand
  - Most informal register (S1A): face-to-face conversation, telephone calls (highest frequency of non-standard and discourse features).

	Sex		
	male	*female*	*Total*
speakers	86	144	230
files	---	---	100
turns	11,821	19,394	31,215
words	89,935	164,695	254,630
eh	185	257	442

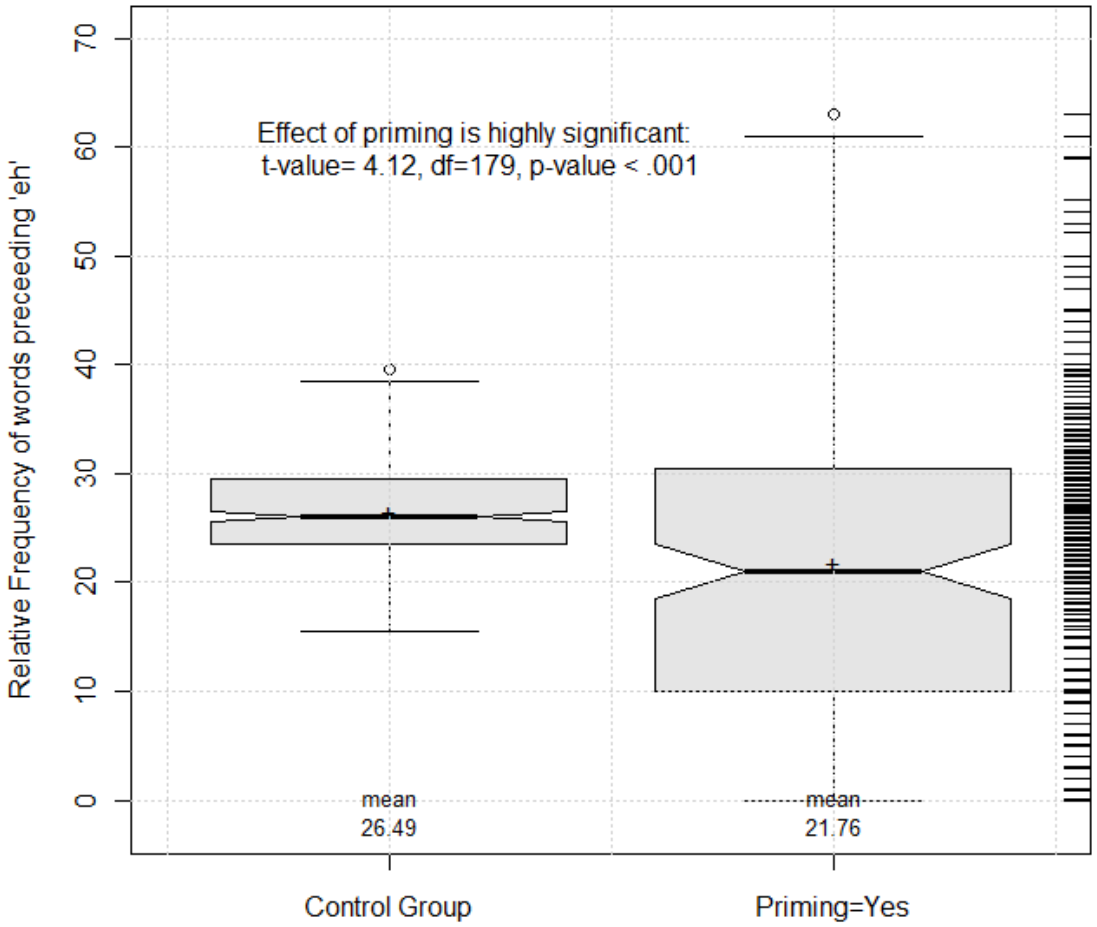
Variable Coding

Dependent variable		
HIT	nominal	yes/no occurrence of turn-final eh
Independent variables		
FILE	categorical	File header
SPEAKER	categorical	Speaker ID
AGE	numeric	Age groups in ascending order
SEX	nominal	Male/female
ETHNICITY	categorical	Pakeha/Maori/other
PRIMING	nominal	Yes/no occurrence of *eh* within last 30 words before turn
OCCUPATION	categorical	Unskilled manual labor (UML), skilled manual labor (SML), clerical (CLE), managerial (MAN), professional (PRO)

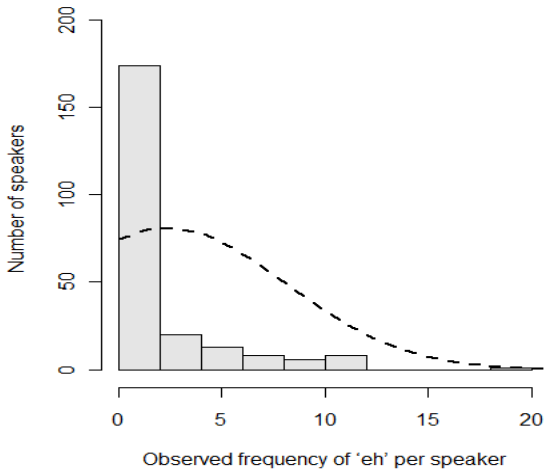
Bivariate statistics

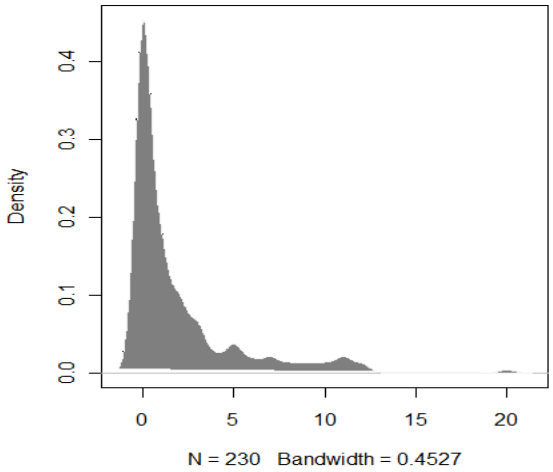


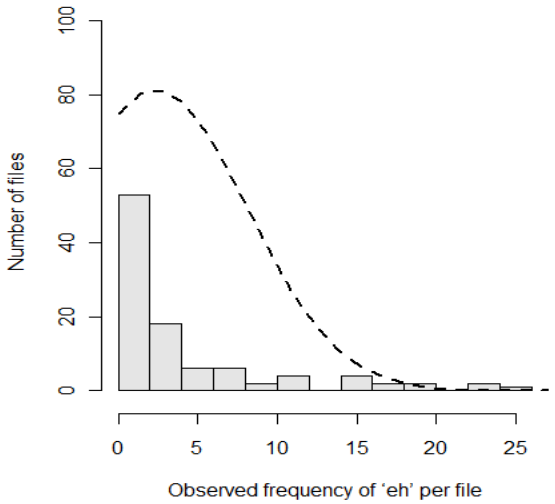
Bivariate statistics

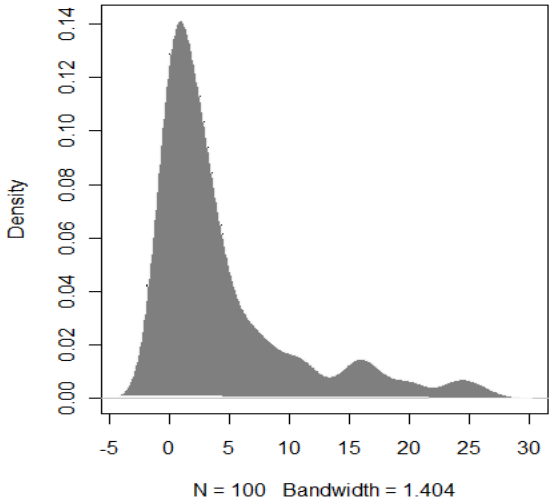


Data Plotting









Statistical Designs

What is a Regression?

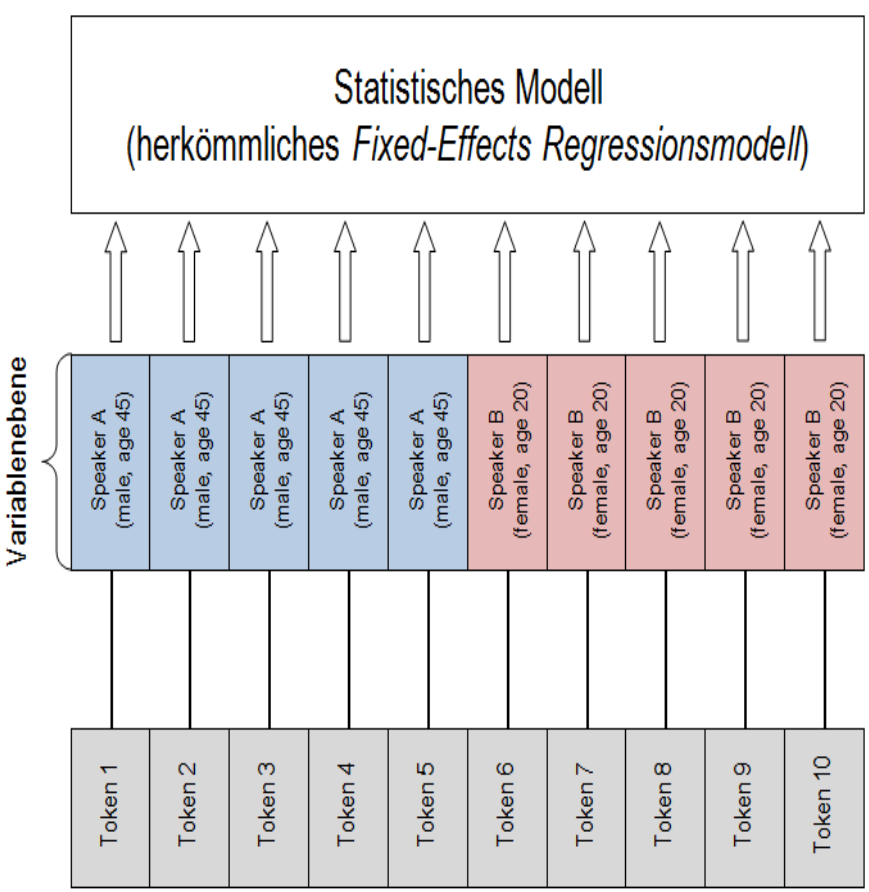
- Statistical method
- Measures the relationship between some phenomenon (dependent variable) and various predictors(independent variables/factors).
- Example
  - Monthly income (dep.) ~ degree of education (indep.),
  - Do people that have attended university earn, on average, more than people who have not attended university?

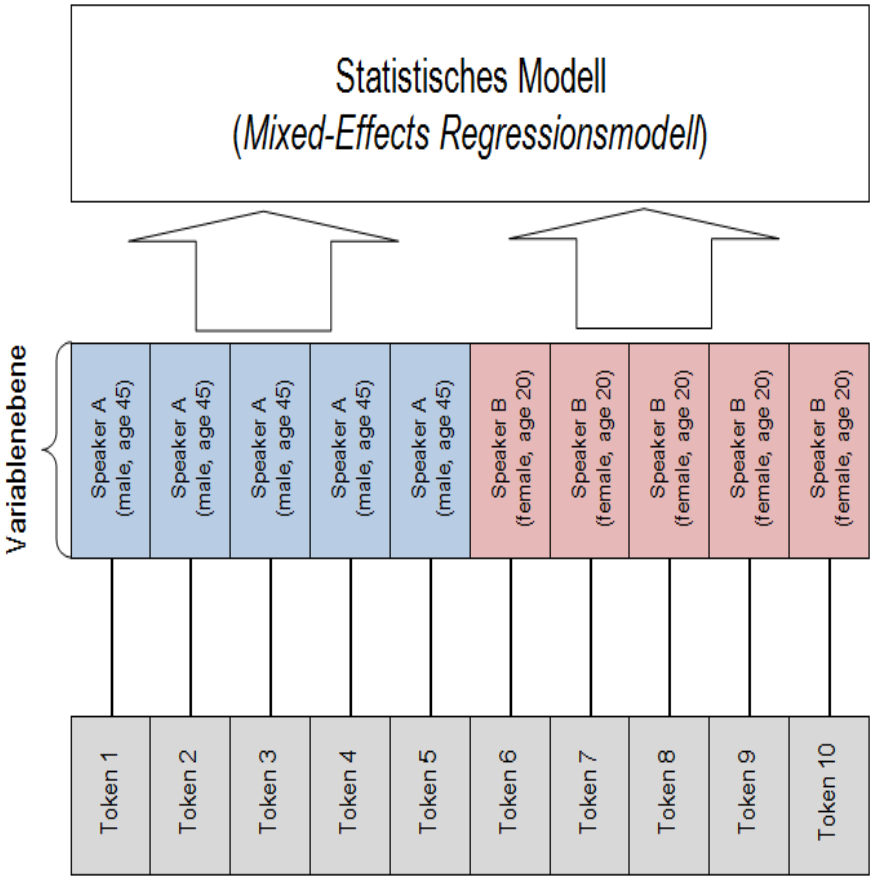
Statistical Designs

What is a Regression?

- Meaningful relationship : „significant“
- „Effect size“ : How strong is this relationship?
- Example
  - A-levels (Abitur) correlates significantly with attending a university
  - “Predictor“ (having A-levels) has a substantial effect size, i.e. it tells us that if you have your A-levels, then it is very likely that you will also have attended university.

Statistical Designs





- Research question
- Theoretical framework
- The phenomenon
- Data
  (summary statistics & data plotting,
  statistical designs & variable coding)

Results

- Discussion, Conclusion & Outlook

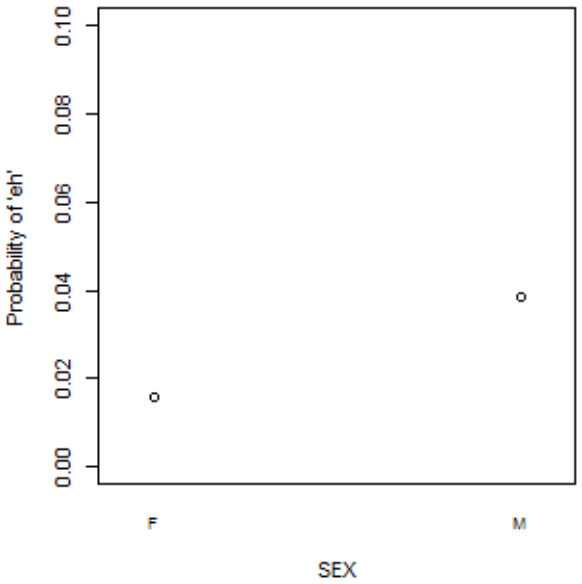
Results

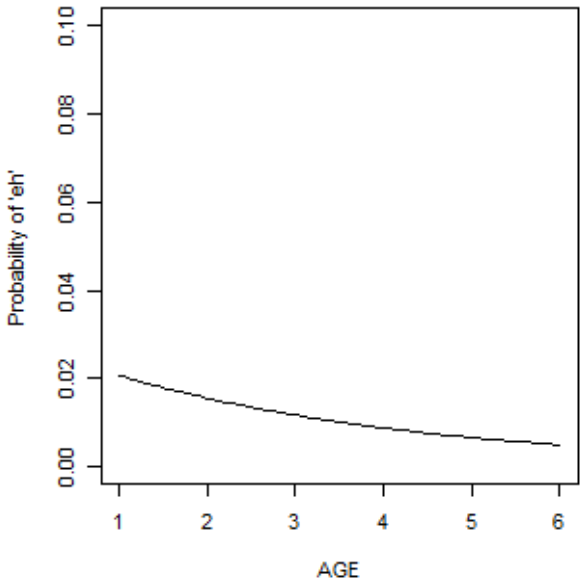
	Traditional statistics (Logistic Regression Model)			*Modern statistics (Generalized Linear Mixed Model)*			*Tomorrow's statistics (Gerneralized Linear Mixed Model)*		
	odds ratio	*increase/ decrease (%)*	*p-value*	*odds ratio*	*increase/ decrease (%)*	*p-value*	*odds ratio*	*increase/ decrease (%)*	*p-value*
Intercept	0.0442	-95.5	<.001	0.0338	-96.6	<.001	0.0282	-97.1	<.001
SEX=M	15.514	55.1	<.001	16.162	61.6	.134	25.320	153.2	<.001
AGE	0.7715	-22.8	<.001	0.7464	-25.3	<.001	0.7478	-25.2	<.001
PAK	0.2343	-76.5	<.001	0.2091	-79.0	<.001	0.2765	-72.3	<.001
PRIMING=Y	19.516	95.1	<.001	12.581	-25.8	.072	---	---	---
SEX=M * PAK	19.657	96.5	<.001	19.867	98.6	.093	---	---	---
	model statistics								
D_{xy}	0.449			0.670			0.674		
R^2	0.065			**AIC & BIC** 4295, 4362			4296, 4347		
C	0.721			0.835			0.837		

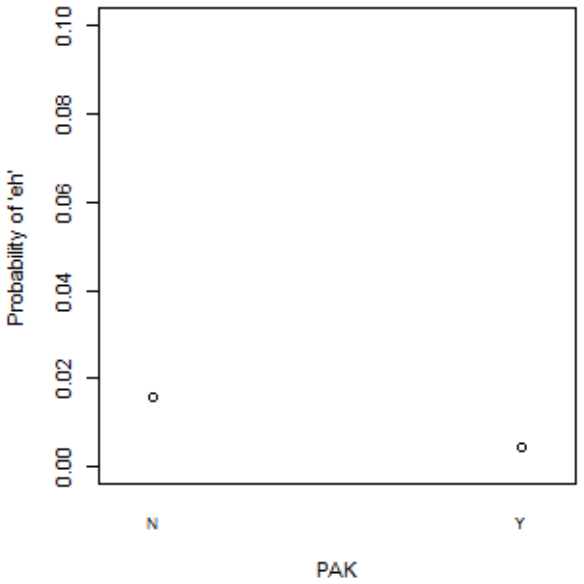
- Research question
- Theoretical framework
- The phenomenon
- Data
  (summary statistics & data plotting,
  statistical designs & variable coding)
- Results

Discussion, Conclusion & Outlook

Discussion


Probability of 'eh'
0.00
0.02
0.04
0.06
0.08
0.10
F
M
SEX


Probability of 'eh'
0.00
0.02
0.04
0.06
0.08
0.10
1
2
3
4
5
6
AGE


Probability of 'eh'
0.00
0.02
0.04
0.06
0.08
0.10
N
Y
PAK

Discussion

Sociolinguistic interpretation

- No significant correlation between *eh* use and occupation
  - Challenges that *eh* is typically a feature of working class speech.
- No significant correlation between *eh* use and priming
  - Effect of priming not relevant for ongoing change?

Discussion

A statisticians interpretation

- Bivariate statistics
  - Misleading: do not control for confounding factors.
- Traditional statistical models
  - Tend to overestimate significance and the impact of extra-linguistic variables
  - May well be fatally flawed and may have led to flawed generalizations about the workings of language.

Conclusion

Converging psycholinguistics & sociolinguistics

- This case study of *eh*
  - Distinct sociolinguistic profile
  - Example: improving our understanding of determinants of linguistic variability by using up-to-date statistics
- Advanced statistical models outperform traditional fixed-effects models
  - More robust/reliable results
  - Allow including more predictors (converging socio- and psycholinguistics)
  - Impact on theorizing!

Outlook

What to do next and what to do better

- Qualitative analysis
- Broaden scope (regional varieties, more phenomena)
- Converge psycholinguistic and sociolinguistic predictors

Outlook

The whole field of variationist socio-linguistics is open for re-analysis and discussion again…

Thank you very much for your attention, eh

You can find this presentation, the data, an annotated R-code and the sources on my homepage:

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Freie Universität Berlin

Particles and Priming –

Combining sociolinguistic and psycholinguistic determinants of variation

Martin Schweinberger

Universität Hamburg
martin.schweinberger@uni-hamburg.de

Freie Universität Berlin
martin.schweinberger@fu-berlin.de